The Annals of Statistics
2016, Vol. 44, No. 4, 1467–1494
DOI: 10.1214/15-AOS1419

SIEVE-BASED INFERENCE FOR INFINITE-VARIANCE LINEAR PROCESSES[1]

BY GIUSEPPE CAVALIERE[1,4,*], ILIYAN GEORGIEV[2,*,†] AND A. M. ROBERT TAYLOR[3,4,‡]

University of Bologna, Universidade Nova de Lisboa† and University of Essex‡*

We extend the available asymptotic theory for autoregressive sieve estimators to cover the case of stationary and invertible linear processes driven by independent identically distributed (i.i.d.) infinite variance (IV) innovations. We show that the ordinary least squares sieve estimates, together with estimates of the impulse responses derived from these, obtained from an autoregression whose order is an increasing function of the sample size, are consistent and exhibit asymptotic properties analogous to those which obtain for a finite-order autoregressive process driven by i.i.d. IV errors. As these limit distributions cannot be directly employed for inference because they either may not exist or, where they do, depend on unknown parameters, a second contribution of the paper is to investigate the usefulness of bootstrap methods in this setting. Focusing on three sieve bootstraps: the wild and permutation bootstraps, and a hybrid of the two, we show that, in contrast to the case of finite variance innovations, the wild bootstrap requires an infeasible correction to be consistent, whereas the other two bootstrap schemes are shown to be consistent (the hybrid for symmetrically distributed innovations) under general conditions.

1. Introduction. A large body of statistical literature exists around the related inference problems of consistent parameter estimation and hypothesis testing within autoregressive and moving average models of (potentially) infinite orders. Key applications include: (i) estimation of the (scale free) spectral density, (ii) inference on impulse response functions, (iii) lag length selection in autoregressive specifications and (sieve) approximations, (iv) point and interval forecasts. Following the pioneering work of Berk (1974), the majority of this literature has been

Received January 2015; revised November 2015.

[1]Supported by the Italian Ministry of Education, University and Research (MIUR), PRIN project "Multivariate statistical models for risk assessment" (MISURA).

[2]Supported by the Fundação para a Ciência e a Tecnologia, Portugal (grants PTDC/EGE-ECO/108620/2008).

[3]Supported by the Economic and Social Research Council of the United Kingdom under research Grant ES/M01147X/1.

[4]Supported by the Danish Council for Independent Research, Sapere Aude | DFF Advanced Grant (Grant nr: 12-124980).

MSC2010 subject classifications. Primary 62M10, 62M15; secondary 62G09.

Key words and phrases. Bootstrap, sieve autoregression, infinite variance, time series.

articulated in the familiar L_2 norm and is therefore not applicable in the case of time series driven by innovations which display infinite variance (IV). Such heavy tailed data are widely encountered in many areas of application including financial, insurance, macroeconomic, actuarial, telecommunication network traffic and meteorological time series; see, *inter alia*, Embrechts, Klüppelberg and Mikosch (1997), Resnick (1997), Finkenstädt and Rootzén (2003) and Davis (2010).

The extension of these time series methods to the case of IV innovations is particularly challenging for at least two distinct reasons. First, under IV the asymptotic distributions of estimators and statistics obtained from autoregressive and moving average time series models are in general non-standard [in particular, they depend on unknown nuisance parameters, such as the so-called "tail index"; see, for example, Davis and Resnick (1985a)]. Second, the bootstrap techniques which are frequently used to approximate the asymptotic distributions of these quantities in the finite-variance case, tend not to be robust to infinite second-order moments and require some modification. This is due to the fact that the bootstrap distributions are dominated by sample extremes [Athreya (1987), Knight (1989)].

In the finite-variance case, sieve-based inference on linear stationary processes is well understood and is based on fitting an autoregressive approximation whose order increases with the sample size. Berk (1974) and Lewis and Reinsel (1985) study the asymptotic distributions of the resulting sieve OLS estimators for univariate and multivariate processes, respectively, while Kreiss (1997) and Paparoditis (1996) demonstrate the asymptotic validity of the associated standard i.i.d. and wild bootstrap sieve inference procedures. In this paper, we explore asymptotic and bootstrap sieve-based methods of inference for stationary linear processes driven by IV innovations, restricting our attention to ordinary least squares (OLS) estimators. While other estimators, including M estimators [see, *inter alia*, Knight (1987), and Davis, Knight and Liu (1992)], and estimators based on trimmed data [Hill (2013), and references therein] can be more efficient than OLS [most notably where the tail index is considerably below 2; see Calder and Davis (1998)], these estimators are dominated by OLS in the finite variance case; see Maronna, Martin and Yohai [(2006), page 269] for a comparison of M and OLS estimators. Importantly, OLS remains widely used by applied workers, especially in economics and finance, and part of our contribution is to show how the small sample efficacy of OLS-based methods can be considerably improved when the innovations are IV using bootstrap methods.

For finite-order autoregression driven by i.i.d. IV errors, it has been established that the OLS estimators of the autoregressive parameters are consistent but that three possible types of asymptotic behavior can occur; see, *inter alia*, Hannan and Kanter (1977) and Davis and Resnick (1985b, 1986). To illustrate this via a simple example, consider the AR(1) process

$$X_t = \beta X_{t-1} + \varepsilon_t, \qquad t \in \mathbb{Z},$$

where $|\beta| < 1$ and ε_t are symmetric i.i.d. in the domain of attraction of an α-stable distribution (defined formally in Section 2) with tail index $\alpha \in (0, 2)$. The large sample behavior of the OLS estimate of β, denoted $\hat{\beta}$, depends on the *unknown* distribution of $\{\varepsilon_t\}$. In particular, three possible cases arise:

CASE (i): If $\mathrm{E}|\varepsilon_1|^\alpha = \infty$, then there exists a sequence l_T, slowly varying at infinity and depending on the distribution of ε_t, such that

$$l_T T^{1/\alpha}(\hat{\beta} - \beta) \xrightarrow{w} \frac{1 - \beta^2}{(1 - \beta^\alpha)^{1/\alpha}} \frac{S_1}{S_0}, \tag{1.1}$$

where S_1 and S_0 are independent α and $\alpha/2$-stable random variables (r.v.'s), respectively; see Davis and Resnick [(1986), page 557];

CASE (ii): If $\mathrm{E}|\varepsilon_1|^\alpha < \infty$ and $\lim_{t\to\infty} P(|\varepsilon_1\varepsilon_2| > t)/P(|\varepsilon_1| > t) = 2\mathrm{E}|\varepsilon_1|^\alpha$, then there exists a slowly varying sequence l_T such that (1.1) holds but where S_1 and S_0 are now dependent α and $\alpha/2$-stable r.v.'s; see Davis and Resnick [(1985b), page 279];

CASE (iii): If $\mathrm{E}|\varepsilon_1|^\alpha < \infty$ and $\lim_{t\to\infty} P(|\varepsilon_1\varepsilon_2| > t)/P(|\varepsilon_1| > t) = \infty$ (note that there are no other possible values for this limit than $2\mathrm{E}|\varepsilon_1|^\alpha$ and ∞), then $\hat{\beta} - \beta$ cannot be normalised such that a non-degenerate limiting distribution obtains; see Davis and Resnick (1985b).

Our first contribution is to show that this asymptotic trichotomy carries over to the general IV linear process case, thereby extending the range of available asymptotic theory for OLS sieve estimators to cover the case of i.i.d. IV innovations. In doing so, we establish the consistency of the OLS sieve estimators and the rates at which the order of the autoregressive approximation must increase with the sample size for these results to hold. We also use these results to demonstrate the consistency of two important estimators derived from the OLS sieve-based estimates, namely estimates of the impulse responses and of the scale-free spectral density function.

As the example above demonstrates, even with knowledge of the tail index, α, asymptotic inference based on the OLS sieve estimator may not be possible and, if it is, it will not be known which form of the asymptotic distribution should be used. Our second contribution is then to investigate the usefulness of bootstrap approximations to the distribution of the OLS sieve estimators in the IV case, complementing the recent work of Kreiss, Paparoditis and Politis (2011) who highlight the wide range of validity of autoregressive sieve bootstrap methods for the case of finite-variance data.

Whilst standard i.i.d. bootstrap methods are inconsistent in the IV case, other bootstrap methods can yield consistent inference for the case of the location parameter; these include the "m out of n" bootstrap [Arcones and Giné (1989)], a parametric bootstrap [Cornea-Madeira and Davidson (2015)], the permutation bootstrap [LePage and Podgórski (1996)] and the wild bootstrap [Cavaliere, Georgiev and Taylor (2013)]. Of these, the latter two preserve the sample extremes (even asymptotically) and are therefore anticipated to lead to more concentrated reference distributions than the unconditional distribution estimated by the "m out of n" and parametric bootstraps [see the numerical evidence in LePage (1992)], and hence to deliver more powerful bootstrap tests. Moreover, issues surrounding sample length selection with the "m out of n" bootstrap and preliminary estimation of the tail index and the asymmetry parameter with the parametric bootstrap are avoided. For these reasons, our focus will be on the permutation bootstrap, the wild bootstrap and a hybrid combination of the two.

In the context of the present problem, with the existence of asymptotic distributions not guaranteed in case (iii) above, we discuss consistency in terms of the proximity between some conditional finite-sample distributions of the OLS sieve estimate and their bootstrap counterparts. We show that the permutation and hybrid bootstraps are consistent under general conditions (the latter provided the innovations are symmetric), but that, in contrast to the case of finite variance innovations, the wild bootstrap is inconsistent unless an *infeasible* correction term is added to the difference between the original and bootstrap sieve estimates. Monte Carlo simulation results are presented which suggest that the permutation and hybrid bootstraps outperform the uncorrected wild bootstrap, "m out of n" bootstrap and i.i.d. bootstrap procedures in terms of finite sample size properties (and the latter two, also in terms of power). These results, consistent with the findings of LePage (1992) and Cavaliere, Georgiev and Taylor (2013) for the location testing problem, also show that the permutation, wild and hybrid bootstrap methods can lead to considerable gains in the finite-sample precision of OLS-based inference under IV, especially when the tail index is small, yet retain the superior properties of OLS-based inference when the innovations have finite variance.

The plan of the paper is as follows. In Section 2, we detail our reference data generating process (DGP) and introduce the autoregressive sieve approximation, and associated OLS sieve estimators. In Section 3, we establish the large sample properties of these estimators. Section 4 investigates the use of sieve bootstrap methods. Results from a Monte Carlo study are reported in Section 5. An application to impulse response functions is offered in Section 6. Main proofs are contained in Section 7; additional theory and proofs are reported in the accompanying supplement [Cavaliere, Georgiev and Taylor (2016)].

2. The DGP and sieve approximation.

Suppose that

$$X_t = \sum_{i=0}^{\infty} \gamma_i \varepsilon_{t-i}, \qquad t \in \mathbb{Z}, \tag{2.1}$$

is a stationary and invertible linear process with IV innovations. Specifically, the following set of conditions is taken to hold.

ASSUMPTION 1. (a) The random variables ε_t ($t \in \mathbb{Z}$) form an i.i.d. sequence which is in the domain of attraction of an α-stable law, $\alpha \in (0, 2)$; that is, the tails of the distribution of ε_t exhibit the power law decay, $P(|\varepsilon_t| > x) = x^{-\alpha} L(x)$, for $x > 0$, with $L(\cdot)$ a slowly varying function at infinity, and $\lim_{x\to\infty} P(\varepsilon_t > x)/P(|\varepsilon_t| > x) =: p \in [0, 1]$, $\lim_{x\to\infty} P(\varepsilon_t < -x)/P(|\varepsilon_t| > x) = 1 - p$. If $\mathrm{E}|\varepsilon_1| < \infty$, it is assumed that $\mathrm{E}\varepsilon_1 = 0$.

(b) There exists a $\delta \in (0, \alpha) \cap [0, 1]$ such that $\sum_{i=0}^{\infty} i|\gamma_i|^{\delta/2} < \infty$.

(c) The power series $\gamma(z) := \sum_{i=0}^{\infty} \gamma_i z^i$, where we set $\gamma_0 = 1$ with no loss of generality, has no roots on the closed complex unit disk.

(d) Its reciprocal $1 - \sum_{i=1}^{\infty} \beta_i z^i := (\sum_{i=0}^{\infty} \gamma_i z^i)^{-1}$ satisfies $\sum_{i=0}^{\infty} |\beta_i|^{\delta} < \infty$, where δ is as defined in (b).

REMARK 2.1. (i) The parameter α in part (a) of Assumption 1, which will be treated as unknown in this paper, controls the thickness of the tails of the distribution of ε_t, and, as such, is often referred to as the tail index, index of stability or characteristic exponent; see, for example, Chapter XVII of Feller (1971). Moments $\mathrm{E}|\varepsilon_t|^r$ are finite for $r < \alpha$ and infinite for $r > \alpha$; the moment $\mathrm{E}|\varepsilon_t|^\alpha$ can be either finite or infinite, discriminating between some results in Section 3. The tail index is inherited by the limiting distribution of the appropriately normalised (and for $\alpha = 1$, also centred) sums of ε_t, belonging to the class of so-called stable distributions. Heavy tailed data are widely encountered in applied research; reported estimates of α include 1.85 for stock returns [McCulloch (1997)], above 1.5 for income, about 1.5 for wealth and trading volumes, about 1 for firm and city sizes [all in Gabaix (2009), and references therein] and even below 1 for returns from technological innovations [Silverberg and Verspagen (2007)].

(ii) Part (b) of Assumption 1 imposes strict stationarity on X_t, guarantees almost sure convergence of $\sum_{i=0}^{\infty} \gamma_i \varepsilon_{t-i}$ (as well as some series in ε_t^2) and underlies the asymptotics for sample correlations [Davis and Resnick (1985b), page 270, and (1986), page 547]. This condition also implies that $\sum_{i=1}^{\infty} i^{2/\delta}|\gamma_i| < \infty$. Therefore, part (b) of Assumption 1 would also impose weak stationarity on X_t in the case where the mean and variance of ε_t were both finite and constant. Part (c) ensures that the MA polynomial, $\gamma(z)$, is invertible, while part (d) implies, among other things, that the infinite autoregressive series in (2.2) below converges absolutely with probability one.

Under Assumption 1, X_t in (2.1) is strictly stationary and invertible and, equivalently, solves the (potentially) infinite-order difference equation

$$X_t = \sum_{i=1}^{\infty} \beta_i X_{t-i} + \varepsilon_t, \qquad t \in \mathbb{Z}. \tag{2.2}$$

The coefficients in (2.2) satisfy $\sum_{i=1}^{\infty} i^{2/\delta}|\beta_i| < \infty$ due to the analogous property imposed on the $\{\gamma_i\}$ in part (b) of Assumption 1; see Brillinger [(2001), pages 76–77].

In this paper, we study inference based on a sieve approximation to (2.2); this is obtained using the truncated autoregression

$$X_t = \sum_{i=1}^{k} \beta_i X_{t-i} + \varepsilon_{t,k}, \tag{2.3}$$

where the lag truncation parameter, k, is an increasing function of the sample size. In (2.3), $\varepsilon_{t,k} := \varepsilon_t + \rho_{t,k}$ where $\rho_{t,k} := \sum_{i=k+1}^{\infty} \beta_i X_{t-i}$ represents the sieve approximation error. The OLS estimates of the sieve parameters $\boldsymbol{\beta}_k := (\beta_1, \dots, \beta_k)'$ in (2.3), given the sample observations $(X_1, \dots, X_T)$, are given by $\hat{\boldsymbol{\beta}}_k := (S_{00}^k)^{-1} \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^k X_t =: (\hat{\beta}_1, \dots, \hat{\beta}_k)'$, where $S_{00}^k := \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^k (\mathbf{X}_{t-1}^k)'$ with $\mathbf{X}_{t-1}^k = (X_{t-1}, \dots, X_{t-k})'$.

3. Convergence results for OLS sieve estimators. Here, we establish the large sample properties of the OLS estimators from the sieve regression (2.3) when the DGP is a linear process driven by IV innovations, as in (2.1). We initially show consistency of the OLS sieve estimators from (2.3). The usual Euclidean vector norm is denoted by $\|\cdot\|$.

THEOREM 1. *Let $\{X_t\}$ be generated according to (2.1) under the conditions of Assumption 1. Then, provided $1/k + k^2/T \to 0$ as $T \to \infty$, it follows that $\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = o_P(1)$.*

Having established the consistency properties of the OLS sieve estimators, we now turn to studying the asymptotic distributions (where they exist) of the OLS sieve estimators, demonstrating how the assumptions used so far need to be strengthened to achieve finer results. We begin by stating a lemma which shows how the asymptotic argument can be reduced to an analysis of the sample autocorrelations. This lemma employs some additional notation that we introduce and discuss next.

First, define $a_T := \inf\{x : P(|\varepsilon_1| > x) \le T^{-1}\}$. By part (a) of Assumption 1, there exists a sequence l_T, slowly varying at infinity, such that $a_T = T^{1/\alpha} l_T$. For the case where $\mathrm{E}|\varepsilon_1|^\alpha < \infty$ and $\lim_{t\to\infty} P(|\varepsilon_1 \varepsilon_2| > t)/P(|\varepsilon_1| > t) = 2\mathrm{E}|\varepsilon_1|^\alpha$, define $\tilde{a}_T = a_T$; otherwise, define $\tilde{a}_T := \inf\{x : P(|\varepsilon_1 \varepsilon_2| > x) \le T^{-1}\}$. In the latter case $\tilde{a}_T = a_T \tilde{l}_T$ for some $\tilde{l}_T$, slowly varying at infinity, such that $\tilde{l}_T \to \infty$ as $T \to \infty$; see Davis and Resnick [(1985b), page 263, and (1986), page 542].

Second, define the infinite Toeplitz matrix $\Sigma := (r_{|i-j|})_{i,j=0}^{\infty}$ formed from the *scale-free autocovariances*, $r_{|i-j|} := \sum_{s=0}^{\infty} \gamma_s \gamma_{s+|i-j|}$. It is a standard fact that Σ generates a bounded operator on the space ℓ_2 of square summable sequences endowed with the Euclidean metric; see Theorem 1.9 of Böttcher and Silbermann [(1999), page 10]. Moreover, under Assumption 1, the operator generated by Σ is invertible; see Theorem 1.15 of Böttcher and Silbermann [(1999), page 18]. We denote the matrix of the inverse operator with respect to the canonical base of ℓ_2 by Σ^{-1}.

Finally, we denote by L a generic $m \times \infty$ selection matrix of constants, with (i, j)th element l_{ij}, and let $L_k := (L_{\cdot 1}, \dots, L_{\cdot k})$ denote the matrix formed from the first k columns of L. The matrix L, and hence L_k, will determine the linear combination(s) of the coefficients, β_j, $j = 1, 2, \dots$ from (2.2) we are interested in conducting inference on, via constructing confidence intervals or, equivalently, testing hypotheses of the generic form $L\boldsymbol{\beta} = l$, where $\boldsymbol{\beta} := (\beta_1, \beta_2, \ldots)'$ is the $\infty \times 1$ vector of AR coefficients from (2.2) and l is a $m \times 1$ vector of constants. For example: inference on β_1 would require $L = (1, 0, 0, \ldots)$; inference involving the sum of β_1 and β_2 would require $L = (1, 1, 0, \ldots)$; a joint simple hypothesis on β_1 and β_2 would require the first row of L to be as in the first example above and the second row to be $(0, 1, 0, \ldots)$. We are now in a position to state our preparatory lemma.

LEMMA 1. *Let the conditions of Theorem 1 hold, and assume further that* $a_T \sum_{i=k+1}^{\infty} |\beta_i| \to 0$ *as* $T \to \infty$. *Also assume that there exists some* $\delta' \in (\delta, \frac{2\alpha}{2+\alpha})$, *where* δ *is as defined in part* (a) *of Assumption 1, such that the selection matrix* L *has* δ'*-summable rows under linear weighting* (*i.e., such that* $\sum_{j=1}^{\infty} j|l_{ij}|^{\delta'} < \infty$, $i = 1, \ldots, m$). *Then, provided* $1/k + k^3/T \to 0$ *as* $T \to \infty$, *with the additional condition that* k *is not a slowly varying function of* T *for the particular value* $\alpha = 1$, *it holds that*

$$\left\| L_k\{(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) - d_T\} - \sigma_T^{-2} \sum_{j=1}^{\infty} A_j \sum_{t=k+1}^{T} (\varepsilon_{t-j}\varepsilon_t - \mu_T) \right\| = o_P(a_T^{-2}\tilde{a}_T),$$

where: $d_T := (T-k)\gamma(1)\mu_T(S_{00}^k)^{-1}\mathbf{u}_k$ *with* $\mu_T := \mathrm{E}(\varepsilon_1\varepsilon_2\mathbb{I}_{\{|\varepsilon_1\varepsilon_2| \le \tilde{a}_T\}})$ *and* $\mathbf{u}_k$ *a* k*-dimensional vector of ones*; $\sigma_T^2 := \sum_{t=k+1}^{T} \varepsilon_t^2$; *finally*, $A_j \in \mathbb{R}^m$ ($j \in \mathbb{N}$) *are given by* $A_j := \sum_{i=1}^{j} L(\Sigma^{-1})_{\cdot i}\gamma_{j-i}$.

REMARK 3.1. (i) The analogue of our condition $a_T \sum_{i=k+1}^{\infty} |\beta_i| \to 0$ in the finite-variance case is $T^{1/2} \sum_{i=k+1}^{\infty} |\beta_i| \to 0$; see Berk (1974) and Lewis and Reinsel (1985). Both conditions involve the order of magnitude of the (possibly centred) error sums $\sum_{t=1}^{T} \varepsilon_t$, respectively, a_T and $T^{1/2}$ for infinite and finite variance. Our condition entails that k is, in general, required to grow at a faster rate the smaller is α. However, in the important special case of a finite-order autoregression, k is only required to be at least as large as the true autoregressive order, while in the case where the β_i, $i = 1, 2, \ldots$ exhibit exponential decay (as happens for finite-order ARMA processes), any power rate of the form $k = T^r$ $[r \in (0, 1)]$ is sufficient uniformly in α. As regards the summability condition on the rows of L, again a similar condition is imposed on L in the finite-variance case; see Theorem 2(iv) of Lewis and Reinsel (1985).

(ii) An important implication of the approximation given in Lemma 1 is that the large sample behavior of the OLS sieve estimator is determined by the same three cases for $\tilde{a}_T^{-1} \sum_{t=k+1}^{T} (\varepsilon_{t-j}\varepsilon_t - \mu_T)$ studied in Davis and Resnick (1985b, 1986) as in the finite-order autoregressive setup discussed in the Introduction. Cases (i) and (ii), where an asymptotic distribution exists, will be detailed in Theorem 2 below. Under case (iii), and as in Davis and Resnick (1985b), the OLS sieve estimators cannot be normalised such that a non-degenerate limiting distribution is obtained.

(iii) Given part (a) of Assumption 1 and the assumption of δ'-row-summability of L under linear weighting, the A_j ($j \in \mathbb{N}$) are also row-wise δ'-summable under linear weighting; that is, $\sum_{j=1}^{\infty} j|A_{ij}|^{\delta'} < \infty$, $i = 1, \dots, m$ (see Section 7.2). This property is sufficient for the series in Lemma 1 and Theorem 2 to be a.s. convergent, and hence, for David and Resnick's [(1985b) and (1986)] asymptotic theory for sample autocovariances to be applied in Theorem 2 below. Notice that the upper bound on δ' is used to control the convergence rate of the quantity $L_k \Sigma_k^{-1}$ to $L\Sigma$.

We now provide in Theorem 2 the asymptotic distribution of the OLS sieve estimator from (2.3) under cases (i) and (ii) of the three possible cases outlined in Section 1.

THEOREM 2. *Let the conditions of Lemma 1 hold, including the rate conditions on k imposed therein. Then we have the following*:

CASE (i): *If* $\mathrm{E}|\varepsilon_t|^\alpha = \infty$, *then*

$$a_T^2 \tilde{a}_T^{-1} L_k\{(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) - d_T\} \xrightarrow{w} S^{-1} \sum_{j=1}^{\infty} A_j S_j, \tag{3.1}$$

where $\{S_j\}_{j=1}^{\infty}$ is an i.i.d. sequence of α-stable r.v.'s and S is an a.s. positive $\alpha/2$-stable r.v. independent of $\{S_j\}_{j=1}^{\infty}$, with remaining notation as in Lemma 1;

CASE (ii): *If* $\mathrm{E}|\varepsilon_t|^\alpha < \infty$ *and* $\liminf_{t\to\infty} P(|\varepsilon_1\varepsilon_2| > t)/P(|\varepsilon_1| > t) = 2\mathrm{E}|\varepsilon_1|^\alpha$, *then* (3.1) *holds with $a_T^2 \tilde{a}_T^{-1} = a_T$, and where $\{S_j\}_{j=1}^{\infty}$ and S are as described in case* (i) *except that they are now dependent with joint distribution as given in Theorem* 3.5 *of Davis and Resnick* (1985b).

REMARK 3.2. (i) The requirement on the lag truncation parameter that $1/k + k^3/T \to 0$, as $T \to \infty$, is standard in analogous theorems in the finite-variance case; see, *inter alia*, Berk [(1974), Theorem 6] and Lewis and Reinsel [(1985), Theorem 2]. However, this rate condition can be weakened in our setting to $1/k + k^{\max\{2,1+\alpha\}+\zeta}/T \to 0$, as $T \to \infty$, for some $\zeta > 0$. Clearly, this condition becomes weaker the further α is from 2, while approaching the usual k^3/T rate as α approaches 2. This weaker rate entails that k is allowed to grow at a faster rate the smaller is α.

(ii) If the distribution of $\{\varepsilon_1\}$ is symmetric (about zero), then so is the distribution of $\varepsilon_1\varepsilon_2$, and the centering term d_T in (3.1) will be zero. If the distribution of $\{\varepsilon_1\}$ is asymmetric and $\alpha \in (0,1) \cup (1,2)$, then the centering of $(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$ can be omitted but at the cost of a location shift in S_j. In the case where $\alpha \in (0,1)$, we have that $T\tilde{a}_T^{-1}\mu_T \to (2\tilde{p} - 1)\frac{\alpha}{1-\alpha}$ as $T \to \infty$, by Karamata's theorem [see

Feller (1971), page 283], where $\tilde{p} := p^2 + (1-p)^2$ [see Assumption 1(a)], and so omitting d_T requires S_j to be replaced by $\tilde{S}_j := S_j + (2\tilde{p}-1)\frac{\alpha}{1-\alpha}$ in (3.1). For $\alpha \in (1,2)$, omitting d_T requires S_j to be replaced by $S_j - \mathrm{E}S_j = S_j + (2\tilde{p}-1)\frac{\alpha}{1-\alpha}$ again. The centering cannot in general (other than in the symmetric case) be omitted when $\alpha = 1$.

(iii) If the distribution of $\{\varepsilon_1\}$ is symmetric (about zero), then so is the distribution of S_j. In this case the ith component of the limit distribution in (3.1) is equal in distribution to $(\sum_{j=1}^{\infty} |A_{ji}|^{\alpha})^{1/\alpha} S_1/S$. This is analogous to the finite-variance case, where the same holds with $\alpha = 2$, S_1 standard Gaussian and $S = 1$. If a consistent estimator $\hat{\alpha}$ of α were available, then $(\sum_{j=1}^{\infty} |A_{ji}|^{\alpha})^{1/\alpha}$ could be consistently estimated. If a Studentising statistic growing at the rate of $a_T^{-2}\tilde{a}_T$ were available (which is in itself an open question, as pointed out by a referee), then an asymptotic test for one-dimensional restrictions could be constructed by reference to the quantity S_1/S. However, and in contrast to the finite-variance case, it does not seem possible to find a full-rank linear transformation of the limit in (3.1) which depends on α alone, precluding a similar simplification of the joint asymptotic test of several restrictions.

The asymptotic results given in this section highlight the infeasibility of classical asymptotic inference for testing linear hypotheses (or constructing confidence intervals) concerning the elements of $\boldsymbol{\beta}$. In particular, as Remark 3.2(ii) makes clear, even under the special case discussed there inference would still not be feasible without knowledge of which of cases (i) and (ii) held [and indeed, that it was one of these cases, rather than case (iii) which held]. An obvious alternative therefore, which we consider in the next section, is to explore bootstrap methods of inference, which may be thought of as a device for approximating the finite sample distributions of the test statistics involved. As in practice it is rarely clear if the data exhibit IV, it will be desirable to have available bootstrap procedures that are valid for testing hypotheses concerning the parameters of linear processes driven either by finite variance or IV innovations.

4. Bootstrap methods. In this section, we propose and discuss three bootstrap methods of inference for IV linear processes. First, we consider the wild bootstrap (based on random sign changes in the residuals), which for the benchmark problem of inference on the location has been shown to be robust to errors with symmetric IV distributions; see Cavaliere, Georgiev and Taylor (2013). LePage (1992) also shows that a wild bootstrap based on random signs can yield very significant improvements in precision since it approximates a conditional version of the test statistic's distribution. Importantly for precision, in the IV case the randomness due to conditioning remains in the limiting distribution of the bootstrap statistic, in contrast to what happens in the finite-variance case. An alternative to the wild bootstrap, which approximates a different conditional distribution of the test statistic with asymptotic randomness, is the permutation bootstrap proposed by LePage and Podgorski (1996) in the context of regressions with fixed regressors and IV errors. Unlike the wild bootstrap, the permutation bootstrap does not require the assumption of distributional symmetry. Finally, we also consider a hybrid of these two, which we term the permuted wild bootstrap, where the residuals are permuted and their signs drawn randomly.

In the problem of inference on the location [Cavaliere, Georgiev and Taylor (2013)], a particularity of the wild bootstrap statistic used is permutation-invariance. Only two of the three bootstrap schemes outlined above will deliver statistics which have permutation-invariant distributions in the present setting and it will turn out to matter for the asymptotic properties of the bootstrap approximation. Specifically, unlike the location case, here bootstrap statistics computed by randomly changing the signs of the residuals (as is done with the wild bootstrap) are not permutation invariant; they are used to approximate the distribution of the test statistics conditional on (essentially) $\{|\varepsilon_t|\}_{t=k+1}^T$ and this distribution changes when the elements in this sequence are reshuffled. To obtain a permutation-invariant reference distribution, the residuals need to be permuted explicitly, resulting in an approximation to the distribution of test statistics conditional on (essentially) the order statistics of $\{\varepsilon_t\}_{t=k+1}^T$ and $\{|\varepsilon_t|\}_{t=k+1}^T$ for the permutation bootstrap and the permuted wild bootstrap, respectively. Moreover, because random permuting effectively enlarges the reference population, the reference distributions for the permutation bootstrap and the permuted wild bootstrap can be expected to be more dispersed than that of the pure wild bootstrap, illustrating a cost of achieving permutation invariance.

4.1. *Bootstrap implementations*. In Algorithm 1 below, we formalise the three bootstrap schemes that we will analyse in this section. To simplify notation and ease exposition, we shall assume that L and L_k are $1 \times \infty$ and $1 \times k$, respectively, corresponding to the case of a single linear restriction of the form $L\boldsymbol{\beta} = l$. Moreover, we shall not Studentise the test statistics. Corresponding results for Wald-type tests of multiple restrictions will be discussed in Remark 4.2(ix).

ALGORITHM 1. *Step* (i): Estimate (2.3) by OLS to yield the sieve estimates, $\hat{\beta}_i$, $i = 1, \dots, k$, and the corresponding residuals, $\hat{\varepsilon}_t := X_t - \hat{\boldsymbol{\beta}}_k' \mathbf{X}_{t-1}^k$, $t = k + 1, \dots, T$.

Step (ii): Generate the bootstrap errors $\varepsilon_t^* := \hat{\varepsilon}_{\pi(t)} w_t$, $t = k+1, \dots, T$, where two options are considered for each of π and $\{w_t\}_{t=k+1}^T$, namely: $(\boldsymbol{\pi}_{\text{id}})$ $\pi(t) = t$, that is, π is the identity function on $\{k+1, \dots, T\}$, or $(\boldsymbol{\pi}_R)$ π is a uniformly distributed random permutation of $\{k+1, \dots, T\}$, and $(\mathbf{w}_1)$ $w_t = 1$ $(t = k+1, \dots, T)$, or $(\mathbf{w}_R)$ w_t are i.i.d. Rademacher r.v.'s ($w_t = \pm 1$ each occurring with probability $\frac{1}{2}$). In all options, π and $\{w_t\}_{t=k+1}^T$ are independent of each other and the data. The combinations $(\boldsymbol{\pi}_{\text{id}}, \mathbf{w}_R)$, $(\boldsymbol{\pi}_R, \mathbf{w}_1)$ and $(\boldsymbol{\pi}_R, \mathbf{w}_R)$ correspond respectively to the wild bootstrap, permutations bootstrap and permuted wild bootstrap.

Step (iii): Construct the bootstrap sample using the recursion

$$X_t^* := \begin{cases} X_t, & t = 1, \dots, k, \\ \sum_{i=1}^{k} \hat{\beta}_i X_{t-i}^* + \varepsilon_t^*, & t = k+1, \dots, T \end{cases} \tag{4.1}$$

and define the bootstrap product moment matrices $S_{00}^{*k} := \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^{*k}(\mathbf{X}_{t-1}^{*k})'$ and $S_{0\varepsilon}^{*k} := \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^{*k}\varepsilon_t^*$ with $\mathbf{X}_{t-1}^{*k} := (X_{t-1}^*, \dots, X_{t-k}^*)'$. The bootstrap analogue of the OLS sieve estimator, $\hat{\boldsymbol{\beta}}_k$, is $\hat{\boldsymbol{\beta}}_k^* := (S_{00}^{*k})^{-1} \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^{*k} X_t^*$.

Step (iv): Define the bootstrap statistic $L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k) = L_k(S_{00}^{*k})^{-1} S_{0\varepsilon}^{*k}$ and use its distribution conditional on the data to approximate an appropriate conditional distribution of $L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$.

REMARK 4.1. (i) As is standard, the distribution of the bootstrap statistic $L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k)$ conditional on the data is approximated by numerical simulation. This is achieved by generating B (conditionally) independent bootstrap statistics, $L_k(\hat{\boldsymbol{\beta}}_k^{*(b)} - \hat{\boldsymbol{\beta}}_k)$, $b = 1, \dots, B$, computed as in Algorithm 1 above, with B large. The respective B simulated quantiles are then used as approximations for the quantiles of $L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$. For instance, in the case where inference is on the null hypothesis $\mathsf{H}_0 : L\boldsymbol{\beta} = l$ against the (one sided) alternative $\mathsf{H}_1 : L\boldsymbol{\beta} > l$, the bootstrap p-value associated to the original test statistic $L_k\hat{\boldsymbol{\beta}}_k - l$ is computed as $\tilde{p}_T^* := B^{-1} \sum_{b=1}^{B} \mathbb{I}(L_k(\hat{\boldsymbol{\beta}}_k^{*(b)} - \hat{\boldsymbol{\beta}}_k) > L_k\hat{\boldsymbol{\beta}}_k - l)$.

(ii) Notice that in the implementation of the bootstrap procedures proposed in Algorithm 1, deterministic normalising sequences (such as $T^{1/2}$ or $a_T^2 \tilde{a}_T^{-1}$ as in Theorem 2), are not required when applied simultaneously to the original and bootstrap statistics, as bootstrap test outcomes are invariant to scaling. This exempts one from the need to decide on an appropriate normalising sequence in applications and, in particular, is important for the robustness of bootstrap tests based on the finite-variance normalisation $T^{1/2}$ to the presence of IV. Nevertheless, normalisation is necessary in the asymptotic analysis of the bootstrap to prevent the statistics at hand from vanishing as T diverges.

4.2. *Asymptotic theory for the bootstrap.* The next theorem, in the style of LePage and Podgorski (1996), characterizes the large sample properties of the three bootstrap methods introduced in Algorithm 1. It concerns the proximity (in the Lévy metric ρ_L) of finite-sample distribution functions as $T \to \infty$. Specifically, for a given $\eta > 0$, η-proximity of two cdfs F and F^* at a point x is evaluated by means of the indicator

$$I_\eta^{F,F^*}(x) := \mathbb{I}\big(F^*(x - \eta) - \eta \le F(x) \le F^*(x + \eta) + \eta\big).$$

Then, for F a (conditional) distribution function of $a_T^2\tilde{a}_T^{-1}L_k(\hat{\boldsymbol{\beta}}_k^* - \boldsymbol{\beta}_k)$ and F^* a (conditional) distribution function of the bootstrap statistic $a_T^2\tilde{a}_T^{-1}L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k)$, we will provide sufficient conditions such that the Lévy distance between F and F^* vanishes in probability as T diverges:

$$\rho_L(F, F^*) \xrightarrow{P} 0, \qquad \rho_L(F, F^*) := \inf\{\eta > 0 : \forall x \in \mathbb{R}, I_\eta^{F,F^*}(x) = 1\}.$$

In the theorem, we will discuss two forms of the bootstrap statistic. Along with $L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k)$, which is the usual bootstrap analogue of $L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$, we will consider an infeasible bootstrap statistic of the form $L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k + \Delta\hat{\boldsymbol{\beta}}_k^*)$, where $\Delta\hat{\boldsymbol{\beta}}_k^* := (S_{00}^{*k})^{-1}\{\sum_{t=k+1}^T w_t \mathbf{X}_{t-1}^{*k}(\mathbf{X}_{t-1}^k)'\}(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$ is a correction term. Although this statistic cannot be computed in practice, it allows us to shed some light on the properties of the wild bootstrap approximation in the present framework. The statistics are normalised as in Lemma 1 and Theorem 2; see also Remark 4.2(viii) below.

THEOREM 3. *Let $\{X_t\}$ be generated according to (2.1) under Assumption 1. Let L be such that $\sum_{i=1}^\infty |L_{\cdot i}|^\delta < \infty$, where δ is as given in Assumption 1. Moreover, let F^* and $F^{*,\Delta}$ denote the bootstrap distribution functions conditional on the data of, respectively, $a_T^2\tilde{a}_T^{-1}L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k)$ and $a_T^2\tilde{a}_T^{-1}L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k + \Delta\hat{\boldsymbol{\beta}}_k^*)$. For distribution functions viewed as stochastic processes on the probability space where $\{\varepsilon_t\}_{t=-\infty}^\infty$ are defined, and provided k is such that $1/k + k^5/T \to 0$ and $a_T\sum_{i=k+1}^\infty |\beta_i| \to 0$ as $T \to \infty$, it holds that*:

(a) *If $\{\varepsilon_t\}$ is symmetrically distributed, then $\rho_L(F^{*,\Delta}, F^{|\boldsymbol{\varepsilon}|}) \to 0$ for the wild bootstrap, $(\boldsymbol{\pi}_{\mathrm{id}}, \mathbf{w}_R)$, where $F^{|\boldsymbol{\varepsilon}|}$ denotes the distribution function of $a_T^2\tilde{a}_T^{-1} \times L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$ conditional on $\{|\varepsilon_t|\}_{t=-\infty}^T$.*

(b) *Provided $k^{1+2/\alpha+\zeta}/T \to 0$ for some $\zeta > 0$ if $\alpha \le \frac{1}{2}$, then $\rho_L(F^*, F^{\mathbf{e}}) \to 0$ for the permutations bootstrap, $(\boldsymbol{\pi}_R, \mathbf{w}_1)$, where $F^{\mathbf{e}}$ denotes the distribution function of $a_T^2\tilde{a}_T^{-1}L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$ conditional on $\{\varepsilon_t\}_{t=-\infty}^k$ and the order statistics of $\{\varepsilon_t\}_{t=k+1}^T$.*

(c) *If $\{\varepsilon_t\}$ is symmetrically distributed, and provided $k^{1+2/\alpha+\zeta}/T \to 0$ for some $\zeta > 0$ if $\alpha \le \frac{1}{2}$, then $\rho_L(F^*, F^{|\mathbf{e}|}) \to 0$ for the permuted wild bootstrap, $(\boldsymbol{\pi}_R, \mathbf{w}_R)$, where $F^{|\mathbf{e}|}$ denotes the distribution function of $a_T^2\tilde{a}_T^{-1}L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$ conditional on $\{|\varepsilon_t|\}_{t=-\infty}^k$ and the order statistics of $\{|\varepsilon_t|\}_{t=k+1}^T$.*

REMARK 4.2. (i) The result in part (a) of Theorem 3 shows that an asymptotically exact (in the Lévy metric) approximation of $F^{|\boldsymbol{\varepsilon}|}$ by the wild bootstrap requires the addition of the correction term, $\Delta\hat{\boldsymbol{\beta}}_k^*$, to $\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k$. In contrast, parts (b) and (c) establish that the permutation bootstrap and the permuted wild bootstrap approximations of, respectively, $F^{\mathbf{e}}$ and $F^{|\mathbf{e}|}$ are consistent (in the Lévy metric) with no need for a correction. Some further clarifications in regard to this are given in Remark 4.2(iv)–(vii) below.

(ii) Because $L\boldsymbol{\beta} = L_k\boldsymbol{\beta}_k + o(a_T^{-2}\tilde{a}_T)$ under the conditions of Theorem 3, the results there hold with $L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$ replaced by $L_k\hat{\boldsymbol{\beta}}_k - L\boldsymbol{\beta}$. Hence, the corrected wild, permutation and permuted wild bootstraps always approximate a (conditional) distribution of $\tau := a_T^2\tilde{a}_T^{-1}(L_k\hat{\boldsymbol{\beta}}_k - L\boldsymbol{\beta})$ for the true $L\boldsymbol{\beta}$. Under the null hypothesis $\mathsf{H}_0 : L\boldsymbol{\beta} = l$, the bootstraps approximate a distribution of the test statistic $\tau^0 := a_T^2\tilde{a}_T^{-1}(L_k\hat{\boldsymbol{\beta}}_k - l)$ since $\tau^0 = \tau$. On the other hand, if H_0 does not hold, then $\tau^0 = \tau + a_T^2\tilde{a}_T^{-1}(L\boldsymbol{\beta} - l)$ diverges at rate $a_T^2\tilde{a}_T^{-1}$, while it can be seen that (under the conditional probability measures of Theorem 3) τ, and hence, its consistent distributional approximations by the bootstrap, have lower orders of magnitude. This implies consistency of the bootstrap tests of H_0. The test based on the uncorrected wild bootstrap is consistent for similar reasons.

(iii) Although in the case where $\alpha \le \frac{1}{2}$ additional rate conditions have been placed on k in parts (b)–(c) of Theorem 3 in order to obtain the stated results in a reasonably tractable way, we conjecture that these extra conditions could be weakened. However, given the very limited empirical relevance of distributions with small α, we have not attempted to do so.

(iv) The use of the correction term $\Delta\hat{\boldsymbol{\beta}}_k^*$ is asymptotically equivalent to using the true $\varepsilon_{t,k}w_t$ instead of $\varepsilon_t^* = \hat{\varepsilon}_t w_t$ in the generation of the bootstrap data X_t^*. The correction term can only be calculated if β is completely specified under the null hypothesis [specifying $L\boldsymbol{\beta}$ alone is not enough as $(S_{00}^{*k})^{-1}\{\sum_{t=k+1}^T w_t\mathbf{X}_{t-1}^{*k}(\mathbf{X}_{t-1}^k)'\}$ is not asymptotically equivalent to a scalar matrix]. Therefore, it is of limited practical interest. Nevertheless, it can be calculated in a simulation experiment in order to evaluate its effect on the finite-sample performance of the bootstrap, as we shall do in Section 5 below.

(v) Under IV, without permuting the residuals, the term $\Delta\hat{\boldsymbol{\beta}}_k^*$ is not asymptotically negligible compared to $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k$. In fact, conditionally on the data, the terms S_{00}^{*k} and $\sum_{t=k+1}^T w_t\mathbf{X}_{t-1}^{*k}(\mathbf{X}_{t-1}^k)'$ have the same order of magnitude in probability as the squared extremes of the data, and so $\Delta\hat{\boldsymbol{\beta}}_k^*$ has the same order of magnitude as $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k$.

(vi) The term $\Delta\hat{\boldsymbol{\beta}}_k^*$ is not related to the fact that we do not centre the residuals, nor to the approximate nature of the autoregressions we estimate. The same correction would be necessary even for an i.i.d. process with IV errors ($\gamma_i = 0$, $i \in \mathbb{N}$) to which an exact finite-order autoregression is fitted (say, with $k = 1$ and coefficient β). Without permuting, in that ($k = 1$) case

$$\Delta\beta^* = \hat{\beta}\left(\sum_{t=3}^T \varepsilon_{t-1}^2\right)^{-1}\left(\sum_{t=3}^T \varepsilon_{t-1}^2 w_{t-1}w_t\right) + o_{P^*}(\hat{\beta}) \tag{4.2}$$

in P-probability, where P^* denotes probability conditional on the data. Then, in the sense of weak convergence of random measures,

$$\mathcal{L}\left(\frac{\sum_{t=3}^{T}\varepsilon_{t-1}^2 w_{t-1}w_t}{\sum_{t=3}^{T}\varepsilon_{t-1}^2}\middle|\{\varepsilon_t\}_{t=2}^{T-1}\right)\xrightarrow{w}\mathcal{L}\left(\frac{\sum_{i=1}^{\infty}\tau_i^{-2/\alpha}\delta_i}{\sum_{t=1}^{\infty}\tau_i^{-2/\alpha}}\middle|\{\tau_i\}_{i=1}^{\infty}\right)$$

with τ_i ($i\in\mathbb{N}$) distributed as the arrival times of a Poisson process with intensity one, and $\{\delta_i\}$ an i.i.d. sequence of Rademacher r.v.'s jointly independent of $\{\tau_i\}_{i\in\mathbb{N}}$ [LePage, Woodroffe and Zinn (1981)]. Rather, the correction is made necessary by the IV of the regressors. This is in contrast to the case where $\sum_{t=3}^{T}\varepsilon_{t-1}^4(\sum_{t=3}^{T}\varepsilon_{t-1}^2)^{-2}=o_P(1)$, for example, when $\mathrm{E}\varepsilon_1^4<\infty$.

(vii) In the context of Remark 4.2(vi), if a random permutation, say π, is applied to the residuals, then the following expansion holds in place of (4.2):

$$\Delta\beta^*=\hat{\beta}\left(\sum_{t=3}^{T}\varepsilon_{t-1}^2\right)^{-1}\left(\sum_{t=3}^{T}\varepsilon_{\pi(t-1)}\varepsilon_{\pi(t)-1}w_{t-1}w_t\right)+o_{P^*}(\hat{\beta}) \tag{4.3}$$

in P-probability. Because $\pi(t-1)\neq\pi(t)-1$ with high probability, the random permutation avoids, with high probability, the squaring of errors in the numerator of $\Delta\beta^*$, in contradistinction to (4.2). Intuitively, since the cumulation of mixed products of errors is of lower stochastic magnitude order than the cumulation of squared errors, the order of magnitude of the leading term in expansion (4.3) is lower than in (4.2). A rough but sufficient formal estimate confirming this is

$$\mathrm{E}^*\left|\frac{\sum_{t=3}^{T}\varepsilon_{\pi(t-1)}\varepsilon_{\pi(t)-1}w_{t-1}w_t}{\sum_{t=3}^{T}\varepsilon_{t-1}^2}\right|\le\frac{1}{T-3}\frac{(\sum_{t=1}^{T}|\varepsilon_t|)^2}{\sum_{t=3}^{T}\varepsilon_{t-1}^2}=o_P(1),$$

where $\sum_{t=3}^{T}\varepsilon_{t-1}^2=O_P(a_T^2)$, and $\sum_{t=1}^{T}|\varepsilon_t|$ is $O_P(a_T)$, $O_P(Tl_T)$ and $O_P(T)$, respectively, for $\alpha\in(0,1)$, $\alpha=1$ and $\alpha\in(1,2)$, with l_T slowly varying at infinity. Hence, by Markov's inequality, $\Delta\beta^*=o_{P^*}(\hat{\beta})$ in P-probability and no correction of the bootstrap statistic is necessary.

(viii) Theorem 3 employs normalisation by the rates from the unconditional analysis of mean corrected estimators (see Lemma 1 and Theorem 2), but does not employ the mean correction itself. Omitting the mean correction may affect the order of magnitude of the estimators (by a multiplicative slowly varying factor) only for the case of asymmetric errors with $\alpha=1$ [see Remark 3.2(iii)]. The bootstrap approximations remain valid also if the statistics are divided by this extra factor because the conclusions of Theorem 3 can be shown to hold also if the normalisation sequence $a_T^2\tilde{a}_T^{-1}$ is replaced by $a_T^{1+\epsilon}$ for small $\epsilon>0$ (for the wild bootstrap, under the extra condition that $a_T^{1+\epsilon}\sum_{i=k+1}^{\infty}|\beta_i|\to 0$ as $k\to\infty$).

(ix) To test m linear restrictions on $\boldsymbol{\beta}$, written as $L\boldsymbol{\beta}=l$, where L satisfies the assumptions of Lemma 1, a Wald statistic can be used:

$$W:=T\hat{\sigma}_T^{-2}(L_k\hat{\boldsymbol{\beta}}_k-l)'\left[L_k\left(S_{00}^k\right)^{-1}L_k'\right]^{-1}(L_k\hat{\boldsymbol{\beta}}_k-l),$$

where $\hat{\sigma}_T^2 := \sum_{t=k+1}^T \hat{\varepsilon}_t^2$. Critical values from conditional distributions of W [with the three conditioning options as in parts (a)–(c) of Theorem 3] can be approximated using the bootstrap distribution of the bootstrap counterparts (resp., feasible and infeasible),

$$W^* := T(\sigma_T^{*2})^{-1}(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k)' L_k' [L_k (S_{00}^{*k})^{-1} L_k']^{-1} L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k),$$
$$W_\Delta^* := T(\sigma_T^{*2})^{-1}(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k + \Delta\hat{\boldsymbol{\beta}}_k^*)' L_k' [L_k (S_{00}^{*k})^{-1} L_k']^{-1} L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k + \Delta\hat{\boldsymbol{\beta}}_k^*),$$

where σ_T^{*2} is the sum of squared residuals for the bootstrap data. The properties of the bootstrap approximation are analogous to those stated in Theorem 3 for the univariate non-Studentised statistics. More specifically, in the supplement [Cavaliere, Georgiev and Taylor (2016)] we show that if $a_T^2 \tilde{a}_T^{-1} L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$, $a_T^2 \tilde{a}_T^{-1} L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k)$ and $a_T^2 \tilde{a}_T^{-1} L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k + \Delta\hat{\boldsymbol{\beta}}_k^*)$ are replaced by $a_T^4 \tilde{a}_T^{-2} T^{-1} W$, $a_T^4 \tilde{a}_T^{-2} T^{-1} W^*$ and $a_T^4 \tilde{a}_T^{-2} T^{-1} W_\Delta^*$, respectively, the conclusions of Theorem 3 remain valid, provided the row sequences of L decay sufficiently fast (cf. Lemma 1).

5. Finite sample properties. We now present results from a small Monte Carlo simulation study comparing the finite sample size and power properties of the three bootstrap procedures from Algorithm 1, together with a standard i.i.d. bootstrap, an "m out of n" bootstrap and a non-bootstrap test which uses a critical value from the standard Gaussian distribution. Throughout the section, the wild bootstrap is based on centred residuals, as we found that centring tends to attenuate the size distortions due to the inconsistency [see Remark 4.2(i)] of this bootstrap method. As a benchmark for comparison, results for the infeasible corrected version of the wild bootstrap discussed in Remark 4.2(iv) are also included in the cases of symmetric errors (where the correction is asymptotically valid). The reference DGP is the MA(1)

$$X_t = \varepsilon_t + \gamma \varepsilon_{t-1}, \qquad t = 1, \dots, T \tag{5.1}$$

with $\gamma \in \{\pm 0.4, 0\}$ and $T \in \{100, 500\}$. The errors $\{\varepsilon_t\}_{t=0}^T$ are i.i.d. draws from one of the following *stable* distributions: (1) symmetric with $\alpha = 1$ (Cauchy); (2) asymmetric with $\alpha = 1$ and asymmetry parameter 0.75; (3) symmetric with $\alpha = 1.5$; (4) asymmetric with $\alpha = 1.5$ and asymmetry parameter 0.75. As a benchmark case, we also include: (5) the standard Gaussian distribution ($\alpha = 2$).

We evaluate the finite-sample size and power properties of tests for the null hypothesis $\mathsf{H}_0 : \beta_1 = \bar{\beta}$ against the two-sided alternative $\mathsf{H}_1 : \beta_1 \neq \bar{\beta}$ in the context of the sieve autoregression (2.3) with k chosen such that the condition in Theorem 3 that $1/k + k^5/T \to 0$ as $T \to \infty$ is satisfied. The same value of k is used in step (iii) of Algorithm 1. Results are reported for the (two-sided) Studentised t-type version of the bootstrap tests [see Remark 4.2(ix)] at the nominal 5% level (tests based on non-Studentised statistics behave very similarly and so are not reported). The results are based on 10,000 Monte Carlo and $B = 1499$ bootstrap repetitions.

TABLE 1
Empirical rejection frequencies under the null hypothesis: *DGP* (5.1)

γ	−0.4					0.0					0.4				
Case	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Panel A: $T = 100$															
Wild	6.4	6.1	6.7	7.0	5.6	6.3	6.4	6.7	7.1	5.5	6.1	5.8	6.6	6.9	5.4
Wild (corr.)	3.5	–	2.8	–	3.2	3.5	–	2.9	–	3.3	3.6	–	2.8	–	3.4
Perm.-Wild	5.2	6.0	4.5	5.1	4.8	5.1	5.8	4.6	4.8	4.9	5.1	5.6	4.7	4.8	4.9
Permutation	5.2	6.1	4.7	5.2	5.0	5.1	5.9	4.7	5.0	5.0	5.2	5.8	4.8	5.0	5.0
i.i.d.	3.7	4.4	3.8	4.1	4.7	3.8	4.4	4.6	4.0	4.8	3.8	4.4	4.0	4.1	4.9
m/n	3.3	4.0	3.3	3.4	3.9	3.3	3.8	3.3	3.4	4.4	3.3	3.7	3.4	3.6	3.9
t_N	3.7	4.4	4.6	4.7	6.9	3.6	4.2	4.6	4.7	6.8	3.7	4.0	4.6	4.7	6.7
Panel B: $T = 500$															
Wild	5.8	6.0	7.2	7.9	4.7	5.9	5.9	7.2	8.0	4.8	5.8	5.6	7.2	8.0	4.8
Wild (corr.)	4.2	–	4.4	–	4.3	4.3	–	4.4	–	4.3	4.1	–	4.3	–	4.3
Perm.-Wild	4.7	6.9	5.1	5.0	4.8	4.8	6.8	5.1	4.9	4.8	4.7	6.5	5.0	4.9	4.8
Permutation	4.7	6.1	4.9	5.3	4.7	4.9	6.2	4.9	5.2	4.6	4.7	5.9	4.9	5.1	4.7
i.i.d.	2.8	3.7	4.0	3.8	4.6	3.0	3.8	4.0	3.8	4.6	2.9	3.4	4.0	3.9	4.7
m/n	2.9	3.9	3.9	3.7	4.4	2.9	4.0	3.9	3.7	4.4	2.9	3.5	3.9	3.6	4.4
t_N	2.4	3.5	3.6	3.5	5.0	4.0	3.4	3.6	3.4	5.0	2.5	3.0	3.7	3.4	5.0

Notes: (i) Tests of $\mathsf{H}_0 : \beta_1 = \bar{\beta}$ with $\bar{\beta} = \gamma$ under the null hypothesis. (ii) "Wild", "Perm.-Wild", "Permutation", "i.i.d.", and "m/n", denote the wild, permuted wild hybrid, permutation, i.i.d. and "m out of n" bootstraps, respectively, based on Studentised tests; (iii) "t_N" denotes the (non-bootstrapped) Studentised test based on standard Gaussian critical values; (iv) "Wild (corr.)" indicates the infeasible wild bootstrap with the correction term included, see Remark 4.2(iv); (v) the lag truncation in both the sieve regression (2.3) and its bootstrap analogue is set to $k = \lfloor 25T^{1/5}/\ln T \rfloor$; the size of the "$m$ out of n" bootstrap samples is set to $m = \lfloor 3T/\ln T \rfloor$.

Empirical rejection frequencies (ERFs) under the true null hypothesis, $\mathsf{H}_0 : \beta_1 = \gamma$, are reported in Table 1. The results for stable symmetric cases (1) and (3) suggest that the permutation and hybrid bootstraps outperform the wild bootstrap, "m out of n" bootstrap and i.i.d. bootstrap procedures as well as the t_N test in terms of finite-sample size control. The same observation can be made for the asymmetric stable cases (2) and (4), although here we do observe a degree of oversizing for the hybrid bootstrap in case (2) when $T = 500$ (recall that the hybrid bootstrap was not shown to be theoretically valid under asymmetry). As expected on theoretical grounds, the wild bootstrap test under cases (1)–(4) is oversized, with size distortions appearing to increase slightly, other things being equal, as the sample size is increased. Notice also that the infeasible corrected wild bootstrap appears to, if anything, overcorrect in small samples. Under case (5), where the errors are Gaussian, all of the procedures are asymptotically valid and little is seen between them, save to note that the "m out of n" bootstrap remains moderately undersized.

TABLE 2
Empirical rejection frequencies under the alternative hypothesis

$\bar{\beta}$	−0.1					0.1				
Case	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Panel A: $T = 100$										
Wild	62.5	68.9	32.3	29.5	15.2	62.4	52.1	33.2	33.5	15.2
Wild (corr.)	62.9	–	30.1	–	11.2	63.9	–	29.2	–	10.9
Perm.-Wild	31.3	38.3	19.2	19.5	14.1	31.3	24.3	18.1	18.4	13.7
Permutation	31.3	38.9	19.2	20.1	14.4	31.0	24.2	18.4	19.0	14.0
i.i.d.	7.7	12.9	11.3	12.5	14.1	7.7	4.2	10.6	9.7	13.4
m/n	6.4	11.3	9.7	10.9	12.1	6.3	3.4	8.9	7.9	11.7
t_N	7.5	13.0	12.8	13.8	17.7	7.3	3.9	12.5	11.4	17.5
Panel B: $T = 500$										
Wild	94.7	98.0	83.2	84.9	59.4	95.1	89.6	83.3	80.4	58.6
Wild (corr.)	95.7	–	83.3	–	58.0	95.2	–	82.9	–	58.4
Perm.-Wild	91.7	97.5	75.8	74.7	58.7	91.2	80.8	75.1	74.4	59.2
Permutation	91.8	97.2	75.4	75.3	58.9	91.1	78.8	74.9	74.9	59.1
i.i.d.	89.9	96.0	73.0	71.6	58.4	89.1	71.9	72.1	72.0	58.9
m/n	90.8	96.8	72.0	69.8	57.4	89.9	74.4	71.0	70.7	57.8
t_N	84.6	96.1	67.1	64.4	59.9	83.4	63.6	65.7	67.0	60.1

Notes: (i) Tests of $\mathsf{H}_0 : \beta_1 = \bar{\beta}$ under the alternative hypothesis $\mathsf{H}_1 : \beta_1 = \gamma = 0$. See also notes to Table 1.

ERFs under the false null hypotheses $\mathsf{H}_0 : \beta_1 = \bar{\beta}$ for $\bar{\beta} \in \{-0.1, 0.1\}$ when in fact $\gamma = 0$ are provided in Table 2. The reported results show that under cases (1)–(4) the permutation and hybrid bootstraps, and to an even greater extent, the wild bootstrap, can lead to significantly more powerful tests than their i.i.d. and "m out of n" bootstrap counterparts, as well as the t_N test. Power gains are particularly apparent for $T = 100$ and are considerably greater for $\alpha = 1$ *vis-à-vis* $\alpha = 1.5$, other things equal. These results are consistent with previous evidence in the literature [LePage (1992); see also the first two paragraphs of Section 4] documenting that in the IV case, inference based on conditional distributions tends to be more precise relative to unconditional inference. The precision gains decrease when α approaches 2, as the conditional distributions get closer to the corresponding unconditional distributions (at least in large samples).

It should be recalled, however, that the wild bootstrap is not size controlled, and this is the price one pays for the additional finite sample power it displays relative to the permuted and hybrid bootstraps under cases (1)–(4). Taking these two aspects of the wild bootstrap together, these results are arguably in accordance with a strand in the recent literature on the possible finite-sample advantages of inconsistent bootstrap procedures with respect to their consistent modifications [Samworth (2003); cf. Pötscher and Leeb (2009), for inconsistent model selection]. The permutation and hybrid bootstraps display almost identical power throughout, suggesting that the permutation bootstrap should be preferred, given its theoretical validity under both symmetric and asymmetric stable cases.

Overall, our Monte Carlo results suggest that, in particular in situations where the practitioner desires inference robust to the possibility of IV, rather than inference specifically designed for the case of IV, the implementation of OLS estimation in conjunction with the permutation and hybrid bootstrap methods appears to be very useful. Furthermore, the wild bootstrap may also constitute a relevant inference device, given its validity in finite variance autoregressive models [Kreiss (1997); Gonçalves and Kilian (2007)] and its superior power properties under IV.

6. Further applications. In this section, we briefly discuss how our results can be applied to the examples of analysing the impulse response (MA) coefficients and the power transfer function (scale-free spectral density) of the process. Proofs of the results in Corollaries 4 and 5 can be found in the accompanying supplement, Cavaliere, Georgiev and Taylor (2016).

Theorem 1 can be used to obtain the consistency properties of the associated sieve-based estimates of the impulse response (MA) coefficients in (2.1). To that end, let $\boldsymbol{\gamma}_k := (\gamma_1, \dots, \gamma_k)'$ denote the vector formed from the first k MA coefficients from (2.1). It is well known that $\boldsymbol{\gamma}_k$ and $\boldsymbol{\beta}_k$ are related via the recursive relation $\boldsymbol{\gamma}_k \equiv \Gamma_k \boldsymbol{\beta}_k$, where Γ_k is the lower triangular Toeplitz matrix with first column $(1 : \boldsymbol{\gamma}'_{k-1})'$. Given $\hat{\boldsymbol{\beta}}_k$, a sieve-based estimator of $\boldsymbol{\gamma}_k$ can therefore by obtained via the recursive relations $\hat{\boldsymbol{\gamma}}_k \equiv \hat{\Gamma}_k \hat{\boldsymbol{\beta}}_k$, where $\hat{\Gamma}_k$ is the lower triangular Toeplitz matrix with first column $(1 : \hat{\boldsymbol{\gamma}}'_{k-1})'$. The consistency of $\hat{\boldsymbol{\gamma}}_k =: (\hat{\gamma}_1, \dots, \hat{\gamma}_k)'$ for $\boldsymbol{\gamma}_k$ is established in the following corollary of Theorem 1.

COROLLARY 4. *Let the conditions of Theorem 1 hold. Then, provided* $1/k + k^2/T \to 0$ *as* $T \to \infty$, *it follows that* $\|\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\| = o_P(1)$.

The impulse response estimates, $\hat{\gamma}_1, \dots, \hat{\gamma}_k$, can in turn be used to obtain a sieve-based estimate of the power transfer function (scale-free spectral density) of the process X_t, $C(\lambda) := |1 + \sum_{j=1}^{\infty} \gamma_j e^{\mathbf{i}j\lambda}|^2$, $\lambda \in (-\pi, \pi]$, where $\mathbf{i}$ is the imaginary unit. Specifically, the sieve-based estimator of $C(\lambda)$ is given by $\hat{C}_T(\lambda) := |1 + \sum_{j=1}^{k} \hat{\gamma}_j e^{\mathbf{i}j\lambda}|^2$, $\lambda \in (-\pi, \pi]$. The following corollary of Theorem 1 establishes the uniform consistency of this estimator.

COROLLARY 5. *Let the conditions of Theorem 1 hold. Then, provided* $1/k + k^2/T \to 0$ *as* $T \to \infty$, *it follows that* $\sup_{\lambda \in (-\pi,\pi]} |\hat{C}_T(\lambda) - C(\lambda)| = o_P(1)$.

We now use Theorem 2 to obtain the asymptotic distribution of the sieve-based estimate $\hat{\boldsymbol{\gamma}}_k$, introduced above, of the vector of the first k impulse response coefficients in (2.1), $\boldsymbol{\gamma}_k$, via the relation $\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k \equiv \hat{\Gamma}_k\hat{\boldsymbol{\beta}}_k - \Gamma_k\boldsymbol{\beta}_k$.

COROLLARY 6. *Let the conditions of Theorem 2 hold, including the rate conditions on k stated therein. Then* $a_T^2\tilde{a}_T^{-1}L_k\{(\hat{\boldsymbol{\gamma}}_k-\boldsymbol{\gamma}_k)-\hat{\Gamma}_k^2 d_T\} \xrightarrow{w} S^{-1}\sum_{j=1}^{\infty}\tilde{A}_j S_j$, *where* S *and* $\{S_j\}_{j=1}^{\infty}$ *are as under cases* (i) *and* (ii) *in Theorem 2, and where* $\tilde{A}_j := \sum_{i=1}^{j} L\Gamma^2(\Sigma^{-1})_{\cdot i}\gamma_{j-i}$ $(j \in \mathbb{N})$ *and* Γ *is the infinite-order lower triangular Toeplitz matrix with first column* $\boldsymbol{\gamma} := (1,\gamma_1,\gamma_2,\ldots)'$. *The centering term,* $\hat{\Gamma}_k^2 d_T$, *can again be omitted under the circumstances outlined in Remark 3.2*(ii).

The bootstrap schemes outlined in Section 4 extend naturally to tests of restrictions on the MA coefficients $\boldsymbol{\gamma}^1 := (\gamma_1,\gamma_2,\ldots)'$. For example, for a null hypothesis of the form $L\boldsymbol{\gamma}^1 = l$, where L satisfies the assumptions of Lemma 1, in the Wald statistics W and W^* one should replace $L_k\hat{\boldsymbol{\beta}}_k - l$, $L_k(S_{00}^k)^{-1}L_k'$, $\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k$ and $L_k(S_{00}^{*k})^{-1}L_k'$ by $L_k\hat{\boldsymbol{\gamma}}_k - l$, $L_k\hat{\Gamma}_k^2(S_{00}^k)^{-1}\hat{\Gamma}_k'^2 L_k'$, $\hat{\boldsymbol{\gamma}}_k^* - \hat{\boldsymbol{\gamma}}_k$ and $L_k(\hat{\Gamma}_k^*)^2(S_{00}^{*k})^{-1}(\hat{\Gamma}_k^{*\prime})^2L_k'$, respectively, where $\hat{\boldsymbol{\gamma}}_k^*$ is obtained from $\hat{\boldsymbol{\beta}}_k^*$ through the recursive relations $\hat{\boldsymbol{\gamma}}_k^* := \hat{\Gamma}_k^*\hat{\boldsymbol{\beta}}_k^*$, with $\hat{\Gamma}_k^*$ denoting a Toeplitz lower triangular matrix with first column $(1 : \hat{\boldsymbol{\gamma}}_{k-1}^{*\prime})'$.

7. Proofs. We employ the matrix norms $\|\cdot\|_2 := \sup_{\|x\|=1}\|(\cdot)x\|$ induced by the Euclidean vector norm and $\|\cdot\| := [\mathrm{tr}\{(\cdot)'(\cdot)\}]^{1/2}$, with tr denoting the trace operator. In particular, for square positive semi-definite matrices, $\|\cdot\|_2 = \lambda_{\max}(\cdot)$, the largest eigenvalue.

7.1. *Consistency of OLS sieve estimators.* The proof of Theorem 1 uses the next lemma [proved in the supplement, Cavaliere, Georgiev and Taylor (2016)]. Let $\Sigma_k := (r_{|i-j|})_{k\times k}$, with $r_{|i-j|} := \sum_{s=0}^{\infty}\gamma_s\gamma_{s+|i-j|}$; under Assumption 1, the eigenvalues of Σ_k are bounded and bounded away from zero as $k \to \infty$ [see Berk (1974), and the related discussion of Σ preceding Lemma 1].

LEMMA 2. *Under Assumption 1 and the condition* $k^2/T + 1/k \to 0$ *as* $T \to \infty$, *for every* $\epsilon > 0$,

(a) $S_{00}^k := \sum_{t=k+1}^{T}\mathbf{X}_{t-1}^k(\mathbf{X}_{t-1}^k)'$ *has* $\|S_{00}^k - \Sigma_k\sigma_T^2\|_2 = O_P(l_T\tilde{a}_T)\max\{k^\epsilon a_k, k\}$, *where* $\sigma_T^2 := \sum_{t=k+1}^{T}\varepsilon_t^2$, $l_T = 1$ *for* $\alpha \neq 1$ *and* l_T *is slowly varying for* $\alpha = 1$; *Moreover,* $(S_{00}^k)^{-1}$ *exists with probability approaching one and* $\|(S_{00}^k)^{-1} - \Sigma_k^{-1}\sigma_T^{-2}\|_2 = O_P(l_T\tilde{a}_T a_T^{-4})\max\{k^\epsilon a_k, k\}$.

(b) $S_{0\varepsilon}^k := \sum_{t=k+1}^{T}\mathbf{X}_{t-1}^k\varepsilon_{t,k}$ *satisfies* $\|S_{0\varepsilon}^k - \sum_{t=k+1}^{T}\mathbf{X}_{t-1}^k\varepsilon_t\| = o_P(a_T^{1-\zeta}) + O_P(a_T^2)\sum_{j=k+1}^{\infty}|\beta_j|$ *and* $\|S_{0\varepsilon}^k\| = o_P(k^\epsilon a_k l_T\tilde{a}_T) + O_P(a_T^2)\sum_{j=k+1}^{\infty}|\beta_j|$ *with* $\zeta > 0$ *sufficiently small and* l_T *as in* (a).

PROOF OF THEOREM 1. From $\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = \|(S_{00}^k)^{-1} S_{0\varepsilon}^k\| \le \|(S_{00}^k)^{-1}\|_2 \times \|S_{0\varepsilon}^k\|$ and the triangle inequality we obtain that, for $\epsilon > 0$,

$$\begin{aligned} \|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| &\le \bigl[\{\lambda_{\min}(\Sigma_k)\}^{-1}\sigma_T^{-2} + \bigl\|(S_{00}^k)^{-1} - \Sigma_k^{-1}\sigma_T^{-2}\bigr\|_2\bigr]\|S_{0\varepsilon}^k\| \\ &= O_P\bigl(a_T^{-2}\bigr)\Biggl[o_P\bigl(k^{\epsilon} a_k l_T \tilde{a}_T\bigr) + O_P\bigl(a_T^2\bigr)\sum_{j=k+1}^{\infty}|\beta_j|\Biggr] \qquad (7.1) \\ &= O_P\bigl(a_k a_T^{\epsilon-1}\bigr) + O_P\Biggl(\sum_{j=k+1}^{\infty}|\beta_j|\Biggr) = o_P(1) \end{aligned}$$

using Lemma 2(a), (b), the stochastic boundedness of $a_T^2\sigma_T^{-2}$ (which converges weakly to an a.s. finite r.v.), the convergence of $\sum_{j=1}^{\infty}|\beta_j|$ and the condition $k^2/T + 1/k \to 0$. □

7.2. *Asymptotic α-stability.* As the row dimension m of the restriction design matrix L is fixed, it is enough to provide proofs for $m = 1$ (in the case of limiting distributions, by the Cramér–Wold device). Thus, $L = (l_1, l_2, \ldots)$ is $1 \times \infty$ in this section.

First, using estimates of the decay rates of the off-diagonal elements of Σ and Σ^{-1}, we discuss the well-definition of the random series in Lemma 1 and its proof. As assumed in that lemma, let $\delta' \in (\delta, \frac{2\alpha}{2+\alpha})$ be such that $\sum_{j=1}^{\infty} j^s |l_j|^{\delta'} < \infty$.

Regarding Σ, the estimate $r_i \le c(1+i)^{-2/\delta}$ for some $c > 0$, δ as in Assumption 1(b) and all $i \in \mathbb{N}$ is implied by the convergence of the series $\sum_{i=1}^{\infty} i^{2/\delta}|r_i|$ which is straightforward to establish under Assumption 1(b). This estimate of r_i implies that $\Sigma \in Q_{2/\delta}$, an algebra studied by Jaffard (1990), and by his Proposition 3, also $\Sigma^{-1} \in Q_{2/\delta}$. Equivalently, if $\Sigma^{-1} =: (s_{ij})_{i,j\in\mathbb{N}}$, then there exists a $c' > 0$ such that $|s_{ij}| \le c'(1 + |i-j|)^{-2/\delta}$ $(i, j \in \mathbb{N})$. As a consequence, for $L\Sigma^{-1} =: (\tilde{l}_1, \tilde{l}_2, \ldots)$, $s \in \{0, 1\}$ and δ' as previously, we find that

$$\sum_{i=1}^{\infty} i^s |\tilde{l}_i|^{\delta'} \le \sum_{i=1}^{\infty} i^s \sum_{j=1}^{\infty} |s_{ji}|^{\delta'} |l_j|^{\delta'} \le (c')^{\delta'} \sum_{j=1}^{\infty}\Biggl[\sum_{i=1}^{\infty}\frac{i^s}{j^s}\bigl(1+|i-j|\bigr)^{-2\delta'/\delta}\Biggr] j^s |l_j|^{\delta'},$$

where

$$\begin{aligned} &\sum_{i=1}^{\infty}\frac{i^s}{j^s}\bigl(1+|i-j|\bigr)^{-2\delta'/\delta} \\ &\qquad = \sum_{i=1}^{j}\frac{i^s}{j^s}\bigl(1+|i-j|\bigr)^{-2\delta'/\delta} + \sum_{i=j+1}^{\infty}\frac{i^s}{j^s}\bigl(1+|i-j|\bigr)^{-2\delta'/\delta} \\ &\qquad < \sum_{i=1}^{j} i^{-2\delta'/\delta} + \sum_{i=1}^{\infty}\frac{(i+j)^s}{j^s} i^{-2\delta'/\delta} < 3\sum_{i=1}^{\infty} i^{s-2\delta'/\delta} < \infty \end{aligned}$$

because $\delta' \in (\delta, 1)$. Hence, with $c_s'' := 3(c')^{\delta'} \sum_{i=1}^{\infty} i^{s-2\delta'/\delta} < \infty$, it holds that $\sum_{i=1}^{\infty} i^s |\tilde{l}_i|^{\delta'} < c_s'' \sum_{j=1}^{\infty} j^s |l_j|^{\delta'}$, and further, $\sum_{i=1}^{\infty} i^s |\tilde{l}_i|^{\delta'} < \infty$ because $\sum_{j=1}^{\infty} j^s |l_j|^{\delta'} < \infty$. Finally, regarding A_j, for $s \in \{0, 1\}$ and δ' as previously,

$$\sum_{j=1}^{\infty} j^s |A_j|^{\delta'} \le \sum_{j=1}^{\infty} \sum_{i=1}^{j} j^s |\tilde{l}_i|^{\delta'} |\gamma_{j-i}|^{\delta'} \le 2 \sum_{j=1}^{\infty} \sum_{i=1}^{j} \{i^s |\tilde{l}_i|^{\delta'}\}\{(j-i+1)^s |\gamma_{j-i}|^{\delta'}\}$$
$$\le 2 \left\{ \sum_{i=1}^{\infty} i^s |\tilde{l}_i|^{\delta'} \right\} \left\{ \sum_{j=0}^{\infty} (j+1)^s |\gamma_j|^{\delta'} \right\},$$

so $\sum_{j=1}^{\infty} j^s |A_j|^{\delta'} < \infty$ holds given that $\sum_{i=1}^{\infty} i^s |\tilde{l}_i|^{\delta'} < \infty$ and $\sum_{j=1}^{\infty} j^s |\gamma_j|^{\delta'} < \infty$. This guarantees that the series below are absolutely convergent a.s. (with $s = 0$) and asymptotic results of Davis and Resnick can be invoked (with $s = 1$, for use in the proof of Theorem 2).

PROOF OF LEMMA 1. Let $\tilde{L}_k := (\tilde{l}_1, \tilde{l}_2, \ldots, \tilde{l}_k) = (L\Sigma^{-1})_k$ consist of the first k entries of $L\Sigma^{-1}$. Then $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k = (S_{00}^k)^{-1} S_{0\varepsilon}^k$ satisfies

$$\left| L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) - \sigma_T^{-2} \tilde{L}_k \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^k \varepsilon_t \right| \le \|L_k\|(B_1 + B_2) + B_3, \tag{7.2}$$

where $\|L_k\|^2 \le \sum_{i=1}^{\infty} l_i^2 < \infty$ for all k, and B_i $(i = 1, 2, 3)$ are defined next and shown to be $o_P(a_T^{-2}\tilde{a}_T)$ provided $k^3/T + 1/k \to 0$. First,

$$B_1 := \left\| \hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k - (S_{00}^k)^{-1} \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^k \varepsilon_t \right\| \le \|(S_{00}^k)^{-1}\|_2 \left\| S_{0\varepsilon}^k - \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^k \varepsilon_t \right\|,$$

where $\|(S_{00}^k)^{-1}\|_2 = O_P(a_T^{-2})$ as in (7.1). Thus, using also Lemma 2(b), $B_1 = o_P(a_T^{-1}) + O_P(1) \sum_{j=k+1}^{\infty} |\beta_j| = o_P(a_T^{-1}) = o_P(a_T^{-2}\tilde{a}_T)$, given that $a_T \sum_{j=k+1}^{\infty} |\beta_j| \to 0$. Second, by Lemma 2(a) and because $\| \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^k \varepsilon_t \| = o_P(k^{\epsilon} a_k l_T \tilde{a}_T)$ for all $\epsilon > 0$ [see the proof of Lemma 2(b)], it holds that

$$B_2 := \|(S_{00}^k)^{-1} - \Sigma_k^{-1}\sigma_T^{-2}\|_2 \left\| \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^k \varepsilon_t \right\| = o_P(\tilde{a}_T^2 a_T^{-4} l_T a_k k^{\epsilon} \max\{a_k, k\}),$$

using the property that multiplication preserves slow variation. Under $k^3/T \to 0$ it is checked directly that $\tilde{a}_T^2 a_T^{-4} l_T k^{\epsilon} = o(a_T^{\epsilon-2})$ for all $\epsilon > 0$ and that $a_T^{\epsilon-2} a_k \max\{a_k, k\} = o(a_T^{-2}\tilde{a}_T)$ for small enough $\epsilon > 0$, so $B_2 = o_P(a_T^{-2}\tilde{a}_T)$. Third,

$$B_3 := \sigma_T^{-2} \left\| (L_k \Sigma_k^{-1})^{\infty} - L\Sigma^{-1} \|_2 \| \sum_{t=k+1}^{T} (\mathbf{X}_{t-1}^k)^{\infty} \varepsilon_t \right\|,$$

where $(\cdot)^\infty$ denotes the infinite sequence (or infinite matrix) obtained from a vector (resp. a matrix) by appending a tail of zeroes (resp., in both dimensions), and the norm is in ℓ_2 (resp., its dual space) in order to comply with the notation of Theorem 3.1 of Strohmer (2002); that theorem ensures the estimate $\|(L_k\Sigma_k^{-1})^\infty - L\Sigma^{-1}\|_2 = O(k^{1/2-s})$ provided that $r_i \le c(1+i)^{-s}$ and $l_i \le c(1+i)^{-s}$ for some $s > 1$, $c > 0$ and all $i \in \mathbb{N}$. Under Assumption 1(b), $r_i \le c(1+i)^{-s}$ was shown above to hold for $s = 2/\delta > 1/2 + 1/\alpha$ and some $c > 0$. As, under the hypotheses of Lemma 1, also $l_i \le c(1+i)^{-s}$ ($i \in \mathbb{N}$) for some $s > 1/2 + 1/\alpha$, $c > 0$, we can define a new $s > 1/2 + 1/\alpha$ such that $\|(L_k\Sigma_k^{-1})^\infty - L\Sigma^{-1}\|_2 = O(k^{1/2-s})$ does hold. Using also that $\|\sum_{t=k+1}^T (\mathbf{X}_{t-1}^k)^\infty \varepsilon_t\| = \|\sum_{t=k+1}^T \mathbf{X}_{t-1}^k \varepsilon_t\| = o_P(k^\epsilon a_k l_T \tilde{a}_T)$ for all $\epsilon > 0$, we find the product of norms in the definition of B_3 to be $o_P(k^{1/2-s+\epsilon} a_k l_T \tilde{a}_T) = o_P(k^{-\omega} l_T \tilde{a}_T)$ for some $\omega > 0$ defined by fixing a sufficiently small $\epsilon > 0$. For $\alpha \neq 1$ (and $l_T = 1$), this magnitude order is $o_P(\tilde{a}_T)$, whereas for $\alpha = 1$ the extra assumption that k grows faster than any slowly varying function of T yields the same magnitude order, so $B_3 = o_P(a_T^{-2}\tilde{a}_T)$ because $\sigma_T^{-2} = O_P(a_T^{-2})$. Hence, from (7.2),

$$L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) = \sigma_T^{-2} \tilde{L}_k \sum_{t=k+1}^T \mathbf{X}_{t-1}^k \varepsilon_t + o_P\big(a_T^{-2}\tilde{a}_T\big). \tag{7.3}$$

Next, define $J_1 := \sum_{j=1}^\infty (\sum_{i=k+1}^j \tilde{l}_i \gamma_{j-i}) \sum_{t=k+1}^T \varepsilon_{t-j}\varepsilon_t \mathbb{I}_{|\varepsilon_{t-j}\varepsilon_t| > \tilde{a}_T}$ and $J_2 := \sum_{j=1}^\infty (\sum_{i=k+1}^j \tilde{l}_i \gamma_{j-i}) \sum_{t=k+1}^T (\varepsilon_{t-j}\varepsilon_t \mathbb{I}_{|\varepsilon_{t-j}\varepsilon_t| \le \tilde{a}_T} - \mu_T)$. Then we can write

$$\begin{aligned}
&\tilde{L}_k \sum_{t=k+1}^T \big(\mathbf{X}_{t-1}^k \varepsilon_t - \mathbf{u}_k \mu_T \gamma(1)\big) \\
&\quad = \sum_{j=1}^\infty \left(\sum_{i=1}^{\min(k,j)} \tilde{l}_i \gamma_{j-i} \right) \sum_{t=k+1}^T (\varepsilon_{t-j}\varepsilon_t - \mu_T) \\
&\quad = \sum_{j=1}^\infty A_j \sum_{t=k+1}^T (\varepsilon_{t-j}\varepsilon_t - \mu_T) + J_1 + J_2.
\end{aligned} \tag{7.4}$$

First, observe that $J_1 = o_P(\tilde{a}_T)$, since

$$\begin{aligned}
\mathrm{E}\big|\tilde{a}_T^{-1} J_1\big|^{\delta'} &\le a_T^{-\delta'} T \mathrm{E}\big(|\varepsilon_1\varepsilon_2|^{\delta'} \mathbb{I}_{|\varepsilon_1\varepsilon_2| > \tilde{a}_T}\big) \sum_{j=1}^\infty \left| \sum_{i=k+1}^j \tilde{l}_i \gamma_{j-i} \right|^{\delta'} \\
&\le O(1) \left(\sum_{j=1}^\infty |\gamma_j|^{\delta'} \right) \left(\sum_{i=k+1}^j |\tilde{l}_i|^{\delta'} \right) = o(1)
\end{aligned}$$

by Karamata's theorem [KT] and the fact that $\sum_{i=k+1}^{\infty}|\tilde{l}_i|^{\delta'} \to 0$ as $k \to \infty$. Second, $J_2 = o_P(\tilde{a}_T)$, since

$$\begin{aligned}
\mathrm{E}J_2^2 = {} & \sum_{j,h=1}^{\infty}\left(\sum_{i=k+1}^{j}\tilde{l}_i\gamma_{j-i}\right)\left(\sum_{i=k+1}^{h}\tilde{l}_i\gamma_{h-i}\right) \\
& \times \sum_{t,s=k+1}^{T}\mathrm{E}\{(\varepsilon_{t-j}\varepsilon_t\mathbb{I}_{|\varepsilon_{t-j}\varepsilon_t|\le\tilde{a}_T}-\mu_T)(\varepsilon_{s-h}\varepsilon_s\mathbb{I}_{|\varepsilon_{s-h}\varepsilon_s|\le\tilde{a}_T}-\mu_T)\} \\
& \le 4T\mathrm{E}(\varepsilon_1^2\varepsilon_2^2\mathbb{I}_{|\varepsilon_1\varepsilon_2|\le\tilde{a}_T})\left(\sum_{j=1}^{\infty}|\gamma_j|\right)^2\left(\sum_{i=k+1}^{\infty}|\tilde{l}_i|\right)^2 = o(\tilde{a}_T^2)
\end{aligned}$$

by KT and because $\sum_{i=k+1}^{\infty}|\tilde{l}_i| \to 0$ as $k \to \infty$. The lemma then follows by combining (7.3) and (7.4) with $J_1 + J_2 = o_P(\tilde{a}_T)$. □

PROOF OF THEOREM 2. Given that $a_k/a_T = o(1)$ and $\tilde{a}_k/\tilde{a}_T = o(1)$ as $T \to \infty$, and $\sum_{j=1}^{\infty} j|A_j|^{\delta'} < \infty$ by the previous argument, it follows respectively from the proof of Theorem 4.4 of Davis and Resnick (1986) and from their Theorem 3.5 in (1985b) that

$$\left(a_T^{-2}\sum_{t=k+1}^{T}\varepsilon_t^2, \tilde{a}_T^{-1}\sum_{j=1}^{\infty}A_j\left(\sum_{t=k+1}^{T}\varepsilon_{t-j}\varepsilon_t - \mu_T\right)\right) \xrightarrow{w} \left(S, \sum_{j=1}^{\infty}A_jS_j\right),$$

with the limit distribution respectively as in the two parts of Theorem 2. This convergence and Lemma 1 prove Theorem 2. □

7.3. *Bootstrap approximations*.

7.3.1. *Notation and preparatory results*. Without loss of generality, in this section we set $X_t^* = 0$, $t = 1, \ldots, k$. Let $\hat{V}_k$ be the $k \times k$ matrix $\hat{V}_k := (\hat{\boldsymbol{\beta}}_k, \mathbf{u}_1, \ldots, \mathbf{u}_{k-1})'$, where $\mathbf{u}_i$ is the ith canonical basis vector in $\mathbb{R}^k$, and let $\hat{\boldsymbol{\gamma}}_{j:k} := \hat{V}_k^j\mathbf{u}_1 = (\hat{\gamma}_j, \ldots, \hat{\gamma}_{j-k+1})'$, $\hat{\gamma}_i := 0$ $(i < 0)$. Then $\mathbf{X}_t^{*k} = \sum_{j=0}^{t-k-1}\hat{\boldsymbol{\gamma}}_{j:k}\varepsilon_{t-j}^*$, $t = k+1, \ldots, T$.

Further, as a benchmark, we introduce the (infeasible) bootstrap errors $\varepsilon_t^\dagger := \varepsilon_{\pi(t)}w_t$ $(t = k+1, \ldots, T)$, which are a transformation of the true errors ε_t instead of the residuals $\hat{\varepsilon}_t$, with π and $\{w_t\}_{t=k+1}^{\infty}$ defined respectively as in the wild, permutations or permuted wild bootstrap. Associated with $\varepsilon_t^\dagger$ we define the infeasible bootstrap sample $X_t^\dagger = 0$ $(t = 1, \ldots, k)$ and $\mathbf{X}_t^{\dagger k} = \sum_{j=0}^{t-k-1}\boldsymbol{\gamma}_{j:k}\varepsilon_{t-j}^\dagger$, $t = k+1, \ldots, T$, where $\mathbf{X}_t^{\dagger k} := (X_t^\dagger, \ldots, X_{t-k+1}^\dagger)'$, $\boldsymbol{\gamma}_{j:k} := (\gamma_j, \ldots, \gamma_{j-k+1})'$, $\gamma_i := 0$ $(i < 0)$, as well as the product moments matrices $S_{00}^{\dagger k} := \sum_{t=k+1}^{T}\mathbf{X}_{t-1}^{\dagger k}(\mathbf{X}_{t-1}^{\dagger k})'$ and

$S_{0\varepsilon}^{\dagger k} := \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^{\dagger k} \varepsilon_{t,k}^{\dagger}$, where $\varepsilon_{t,k}^{\dagger} := \varepsilon_t^{\dagger} + \rho_{t,k}^{\dagger}$, $\rho_{t,k}^{\dagger} := \sum_{i=k+1}^{t-k-1} \beta_i X_{t-i}^{\dagger}$. Equivalently,

$$X_t^{\dagger} = \begin{cases} 0, & t = 1, \dots, k, \\ \boldsymbol{\beta}_k' \mathbf{X}_{t-1}^{\dagger k} + \varepsilon_{t,k}^{\dagger}, & t = k+1, \dots, T. \end{cases}$$

In order to prove Theorem 3, we will need the following Lemma 3 [proved in the supplement, Cavaliere, Georgiev and Taylor (2016)]. We denote by $P^{\dagger}$ probability conditional on $\{\varepsilon_t\}_{t=-\infty}^{T}$ (equivalently, on $\{\varepsilon_t\}_{t=-\infty}^{T}$ and the data, as the latter are a measurable function of $\{\varepsilon_t\}_{t=-\infty}^{T}$).

LEMMA 3. *Under Assumption 1 and the conditions $k^4/T + 1/k \to 0$, $a_T \sum_{i=k+1}^{\infty} |\beta_i| \to 0$ as $T \to \infty$, it holds in P-probability that*:

(a) $\|S_{00}^{*k} - S_{00}^{\dagger k}\| = o_{P^{\dagger}}(k a_k a_T^{1+\epsilon})$ *for all $\epsilon > 0$ and $\lambda_{\min}^{-1}(a_T^{-2} S_{00}^{*k}) = O_{P^{\dagger}}(1)$ in P-probability.*

(b) *If π is the identity, then $\|S_{0\varepsilon}^{*k} - S_{0\varepsilon}^{\dagger k} - \sigma_1\| = o_{P^{\dagger}}(a_T^1)$, where $\sigma_1 := \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^{*k}(\hat{\varepsilon}_t - \varepsilon_{\pi(t),k}) w_t$. If π is a random permutation* [*r.p.*], *then $\|S_{0\varepsilon}^{*k} - S_{0\varepsilon}^{\dagger k}\| = o_{P^{\dagger}}(T^{\epsilon - 1/2} k^{1/2} a_k a_T)$ for all $\epsilon > 0$.*

(c) $\|S_{0\varepsilon}^{\dagger k}\| = O_{P^{\dagger}}(a_k^{1+\epsilon} \tilde{a}_T)$ *for π equal to the identity and all $\epsilon > 0$, and $\|S_{0\varepsilon}^{\dagger k}\| = O_{P^{\dagger}}(h_{Tk})$ for an r.p. π, with $h_{Tk} = \min\{\frac{k^{1/2}}{T^{1/2}} a_T^2, \frac{k}{T} \max\{T^2, a_T^2 l_T^2\}\}$ and l_T as in Lemma 2.*

7.3.2. *Proof of Theorem* 3. Let the bootstrap statistic and its corrected version be $\tau^* := a_T^2 \tilde{a}_T^{-1} L_k(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k)$ and $\tau_c^* := a_T^2 \tilde{a}_T^{-1} L_k\{(\hat{\boldsymbol{\beta}}_k^* - \hat{\boldsymbol{\beta}}_k) - (S_{00}^{*k})^{-1}\sigma_1\}$, where $\sigma_1 = \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^{*k}(\hat{\varepsilon}_{\pi(t)} - \varepsilon_{\pi(t),k}) w_t$. We need to evaluate the Lévy distance between the distribution of τ^* and τ_c^* conditional of the data and three conditional distributions of $\tau := a_T^2 \tilde{a}_T^{-1} L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)$. To this end, we introduce some auxiliary r.v.'s and evaluate sequentially several distances involving them as well as τ^*, τ_c^* and τ, such that our desired evaluation then follows by the triangle inequality. The auxiliary r.v.'s are $\tau^{\dagger} := a_T^2 \tilde{a}_T^{-1} L_k(S_{00}^{\dagger k})^{-1} S_{0\varepsilon}^{\dagger k}$, $\breve{\tau} := a_T^2 \tilde{a}_T^{-1} L_k(\breve{S}_{00}^{k})^{-1} \breve{S}_{0\varepsilon}^{k}$ and $\tau^{\pi} := a_T^2 \tilde{a}_T^{-1} L_k(S_{00}^{\pi k})^{-1} S_{0\varepsilon}^{\pi k}$, where we define $\breve{S}_{00}^{k} := \sum_{t=k+2}^{T} \breve{\boldsymbol{X}}_{t-1}^{k}(\breve{\boldsymbol{X}}_{t-1}^{k})'$ and $\breve{S}_{0\varepsilon}^{k} := \sum_{t=k+2}^{T} \breve{\boldsymbol{X}}_{t-1}^{k} \varepsilon_{\pi(t)}$ with $\breve{\boldsymbol{X}}_{t-1}^{k} := \sum_{j=0}^{t-k-2} \boldsymbol{\gamma}_{j:k} \varepsilon_{\pi(t-j-1)}$, while $S_{00}^{\pi k} := \sum_{t=k+2}^{T} \mathbf{X}_{t-1}^{\pi k}(\mathbf{X}_{t-1}^{\pi k})'$ and $S_{0\varepsilon}^{\pi k} := \sum_{t=k+2}^{T} \mathbf{X}_{t-1}^{\pi k}(\varepsilon_{\pi(t)} + \rho_{t,k}^{\pi})$ with, $\mathbf{X}_{t-1}^{\pi k} := \sum_{j=0}^{t-k-2} \boldsymbol{\gamma}_{j:k} \varepsilon_{\pi(t-j-1)} + \sum_{j=t-k}^{\infty} \boldsymbol{\gamma}_{j:k} \varepsilon_{t-j}$ and finally $\rho_{t,k}^{\pi} := \sum_{i=k+1}^{\infty} \beta_i \times (\sum_{m=0}^{t-k-i-1} \gamma_m \varepsilon_{\pi(t-m-i)} + \sum_{m=t-k-i}^{\infty} \gamma_m \varepsilon_{t-m-i})$. The sequential distances are as follows.

1. The bootstrap statistics τ^* and τ_c^*, as measurable functions of the data, π and $\{w_t\}_{t=k+1}^{T}$, have the same distribution conditional on the data. Specifically, conditional on the data and all past $\{\varepsilon_t\}_{t=-\infty}^{T}$ it holds that $\rho_{1c} := \rho_L(\mathcal{L}^*(\tau_c^*), \mathcal{L}^{\dagger}(\tau_c^*)) =$

0 and $\rho_1 := \rho_L(\mathcal{L}^*(\tau^*), \mathcal{L}^\dagger(\tau^*)) = 0$, where ρ_L denotes Lévy distance, $\mathcal{L}$ stands for law, the $*$ superscript for conditioning on the data, and the $\dagger$ superscript for conditioning on the data and $\{\varepsilon_t\}_{t=-\infty}^T$.

2. If π is the identity, then $\tau_c^* = a_T^2 \tilde{a}_T^{-1} L_k (S_{00}^{*k})^{-1}(S_{0\varepsilon}^{*k} - \sigma_1) = \tau^\dagger + o_{P^\dagger}(1)$ in P-probability (proved below), whereas if π is an r.p., also $\tau^* = \tau^\dagger + o_{P^\dagger}(1)$ in P-probability independently of the specification of $\{w_t\}_{t=k+1}^T$ (proved below). Hence,

$$\rho_{2c} := \rho_L\bigl(\mathcal{L}^\dagger(\tau_c^*), \mathcal{L}^\dagger(\tau^\dagger)\bigr) = o_P(1),$$

$$\rho_2 := \rho_L\bigl(\mathcal{L}^\dagger(\tau^*), \mathcal{L}^\dagger(\tau^\dagger)\bigr) = o_P(1).$$

3(a) Under bootstrap schemes $\mathbf{w}_R$ and symmetry of ε_t, it holds that $\rho_3^{|\boldsymbol{\varepsilon}|} := \rho_L(\mathcal{L}^\dagger(\tau^\dagger), \mathcal{L}^{|\boldsymbol{\varepsilon}|}(\breve{\tau})) = 0$, where the $|\boldsymbol{\varepsilon}|$ subscript denotes conditioning on $= \{|\varepsilon_t|\}_{t=-\infty}^T$.

3(b) Under scheme $(\boldsymbol{\pi}_R, \mathbf{w}_1)$, $\tau^\dagger = \breve{\tau}$ so $\rho_3^\dagger := \rho_L(\mathcal{L}^\dagger(\tau^\dagger), \mathcal{L}^\dagger(\breve{\tau})) = 0$.

4(a) Under symmetry of ε_t, $\breve{\tau} = \tau^\pi + o_{P^{|\boldsymbol{\varepsilon}|}}(1)$ in P-probability (discussed below), resulting in $\rho_L(\mathcal{L}^{|\boldsymbol{\varepsilon}|}(\breve{\tau}), \mathcal{L}^{|\boldsymbol{\varepsilon}|}(\tau^\pi)) = o_P(1)$. Two conclusions follow.

Where π is the identity, it holds that $\tau^\pi = \tau$, so the previous convergence becomes $\rho_4^{|\boldsymbol{\varepsilon}|} := \rho_L(\mathcal{L}^{|\boldsymbol{\varepsilon}|}(\breve{\tau}), \mathcal{L}^{|\boldsymbol{\varepsilon}|}(\tau)) = o_P(1)$.

Instead, where π is an r.p., τ conditional on $\{|\varepsilon_t|\}_{t=-\infty}^k$ and the order statistics of $\{|\varepsilon_t|\}_{t=k+1}^T$ is distributed like τ^π conditional on $\{|\varepsilon_t|\}_{t=-\infty}^T$, so now $\rho_4^{|\mathbf{e}|} := \rho_L(\mathcal{L}^{|\boldsymbol{\varepsilon}|}(\breve{\tau}), \mathcal{L}^{|\mathbf{e}|}(\tau)) = o_P(1)$, with $|\mathbf{e}|$ standing for conditioning on $\{|\varepsilon_t|\}_{t=-\infty}^k$ and the order statistics of $\{|\varepsilon_t|\}_{t=k+1}^T$.

4(b) Generally, $\breve{\tau} = \tau^\pi + o_{P^\dagger}(1)$ in P-probability (discussed below). As τ conditional on $\{\varepsilon_t\}_{t=-\infty}^k$ and the order statistics of $\{\varepsilon_t\}_{t=k+1}^T$ is distributed like τ^π conditional on $\{\varepsilon_t\}_{t=-\infty}^T$ (equivalently, under $P^\dagger$), it follows that $\rho_4^{\mathbf{e}} := \rho_L(\mathcal{L}^\dagger(\breve{\tau}), \mathcal{L}^{\mathbf{e}}(\tau)) = o_P(1)$, with $\mathbf{e}$ standing for conditioning on $\{\varepsilon_t\}_{t=-\infty}^k$ and the order statistics of $\{\varepsilon_t\}_{t=k+1}^T$.

Next we combine the previous evaluations. First, we can conclude that, for π equal to the identity (scheme $\boldsymbol{\pi}_{\mathrm{id}}$, wild bootstrap),

$$\rho_L\bigl(\mathcal{L}^*(\tau_c^*), \mathcal{L}^{|\boldsymbol{\varepsilon}|}(\tau)\bigr) \le \rho_{1c} + \rho_{2c} + \rho_3^{|\boldsymbol{\varepsilon}|} + \rho_4^{|\boldsymbol{\varepsilon}|} = o_P(1),$$

which is equivalent to the convergence in Theorem 3(a). On the other hand, for an r.p. π,

$$\rho_L\bigl(\mathcal{L}^*(\tau^*), \mathcal{L}^{\mathbf{e}}(\tau)\bigr) \le \rho_1 + \rho_2 + \rho_3^\dagger + \rho_4^{\mathbf{e}} = o_P(1) \qquad \text{for } (\boldsymbol{\pi}_R, \mathbf{w}_1),$$

$$\rho_L\bigl(\mathcal{L}^*(\tau^*), \mathcal{L}^{|\mathbf{e}|}(\tau)\bigr) \le \rho_1 + \rho_2 + \rho_3^{|\boldsymbol{\varepsilon}|} + \rho_4^{|\mathbf{e}|} = o_P(1) \qquad \text{for } (\boldsymbol{\pi}_R, \mathbf{w}_R);$$

hence, Theorem 3(b), (c).

It remains to complete steps 2 and 4 outlined above.

Step 2. Let $\sigma_1^{\mathrm{id}} := \mathbb{I}_{\pi=\mathrm{id}}\sigma_1$. The next evaluation is valid for $L_k \neq 0$ of type $m \times k$, $m \in \mathbb{N}$:

$$\begin{aligned}
&\|L_k\|^{-1}\|L_k(S_{00}^{*k})^{-1}(S_{0\varepsilon}^{*k} - \sigma_1^{\mathrm{id}}) - L_k(S_{00}^{\dagger k})^{-1}S_{0\varepsilon}^{\dagger k}\| \\
&\quad \le \|(S_{00}^{*k})^{-1} - (S_{00}^{\dagger k})^{-1}\|\|S_{0\varepsilon}^{*k} - \sigma_1^{\mathrm{id}}\| + \|(S_{00}^{\dagger k})^{-1}\|_2\|S_{0\varepsilon}^{*k} - \sigma_1^{\mathrm{id}} - S_{0\varepsilon}^{\dagger k}\| \\
&\quad \le \frac{\|(S_{00}^{\dagger k})^{-1}\|_2^2\|S_{00}^{*k} - S_{00}^{\dagger k}\|}{1 - \|(S_{00}^{\dagger k})^{-1}\|_2\|S_{00}^{*k} - S_{00}^{\dagger k}\|}(\|S_{0\varepsilon}^{\dagger k}\| + \|S_{0\varepsilon}^{*k} - \sigma_1^{\mathrm{id}} - S_{0\varepsilon}^{\dagger k}\|) \\
&\qquad + \|(S_{00}^{\dagger k})^{-1}\|_2\|S_{0\varepsilon}^{*k} - \sigma_1^{\mathrm{id}} - S_{0\varepsilon}^{\dagger k}\|
\end{aligned}$$

with $P^\dagger$-probability approaching 1 in P-probability, as $\|(S_{00}^{\dagger k})^{-1}\|_2\|S_{00}^{*k} - S_{00}^{\dagger k}\| = o_{P^\dagger}(1)$ in P-probability by Lemma 3. Using again Lemma 3 and the conditions $k^4/T \to 0$ (for $\boldsymbol{\pi}_{\mathrm{id}}$) and $k^5/T \to 0$, $k^{1+2/\alpha+\zeta}/T \to 0$ (for $\boldsymbol{\pi}_R$ and some $\zeta > 0$) it follows that for small enough $\epsilon > 0$,

$$\begin{aligned}
&\|L_k\|^{-1}|L_k(S_{00}^{*k})^{-1}(S_{0\varepsilon}^{*k} - \sigma_1^{\mathrm{id}}) - L_k(S_{00}^{\dagger k})^{-1}S_{0\varepsilon}^{\dagger k}| \\
&\quad \le \frac{o_{P^\dagger}(ka_k a_T^{\epsilon-3})}{1 + o_{P^\dagger}(1)}(\mathbb{I}_{\pi=\mathrm{id}}a_k a_T^{1+\epsilon} + \mathbb{I}_{\pi=\mathrm{r.p.}}\{h_{Tk} + T^{\epsilon-1/2}k^{1/2}a_k a_T\}) \\
&\qquad + a_T^{-2}o_{P^\dagger}(\mathbb{I}_{\pi=\mathrm{id}}a_T + \mathbb{I}_{\pi=\mathrm{r.p.}}T^{\epsilon-1/2}k^{1/2}a_k a_T) \\
&\quad = o_{P^\dagger}(a_T^{-1}) \qquad \text{in } P\text{-probability}
\end{aligned}$$

for all the three bootstrap schemes, from where step 2 follows.

Step 4. This step is analogous to step 2, prepared by Lemma 3's estimates involving ρ, with $\operatorname{sgn}\varepsilon_t$ playing the role of w_t.

Acknowledgements. We thank Runze Li (Editor), an Associate Editor and three anonymous referees for their very helpful and constructive comments on a previous version of this paper.

Parts of this paper were written while Giuseppe Cavaliere was also affiliated with the University of Copenhagen.

SUPPLEMENTARY MATERIAL

Supplement to "Sieve-based inference for infinite-variance linear processes" (DOI: 10.1214/15-AOS1419SUPP; .pdf). In this supplement, which contains additional theoretical results and proofs, we provide: a lemma with two tail inequalities regarding the series of the coefficients from the AR(∞) representations; a proof of Lemma 2 and corollaries from Section 6; proofs of the results given in Section 7.3.1; a discussion of multiple restrictions.

REFERENCES

ARCONES, M. A. and GINÉ, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Ann. Inst. Henri Poincaré Probab. Stat.* **25** 457–481. MR1045246

ATHREYA, K. B. (1987). Bootstrap of the mean in the infinite variance case. *Ann. Statist.* **15** 724–731. MR0888436

BERK, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489–502. MR0421010

BÖTTCHER, A. and SILBERMANN, B. (1999). *Introduction to Large Truncated Toeplitz Matrices*. Springer, New York. MR1724795

BRILLINGER, D. R. (2001). *Time Series: Data Analysis and Theory. Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. MR1853554

CALDER, M. and DAVIS, R. A. (1998). Inference for linear processes with stable noise. In *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (R. Adler, R. Feldman and M. Taqqu, eds.) 159–176. Birkhäuser, Basel.

CAVALIERE, G., GEORGIEV, I. and TAYLOR, A. M. R. (2013). Wild bootstrap of the sample mean in the infinite variance case. *Econometric Rev.* **32** 204–219. MR2988925

CAVALIERE, G., GEORGIEV, I. and TAYLOR, A. (2016). Supplement to "Sieve-based inference for infinite-variance linear processes." DOI:10.1214/15-AOS1419SUPP.

CORNEA-MADEIRA, A. and DAVIDSON, R. (2015). A parametric bootstrap for heavy-tailed distributions. *Econometric Theory* **31** 449–470. MR3348454

DAVIS, R. A. (2010). Heavy tails in financial time series. In *Encyclopedia Quantitative Finance* (R. Cont, ed.). Wiley, New York.

DAVIS, R. A., KNIGHT, K. and LIU, J. (1992). M-estimation for autoregressions with infinite variance. *Stochastic Process. Appl.* **40** 145–180. MR1145464

DAVIS, R. and RESNICK, S. (1985a). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13** 179–195. MR0770636

DAVIS, R. and RESNICK, S. (1985b). More limit theory for the sample correlation function of moving averages. *Stochastic Process. Appl.* **20** 257–279. MR0808161

DAVIS, R. and RESNICK, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14** 533–558. MR0840513

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events: For Insurance and Finance. Applications of Mathematics (New York)* **33**. Springer, Berlin. MR1458613

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd ed. Wiley, New York. MR0270403

FINKENSTÄDT, B. and ROOTZÉN, H. (2003). *Extreme Values in Finance, Telecommunication and the Environment*. Chapman & Hall, London.

GABAIX, X. (2009). Power laws in economics and finance. *Annual Review of Economics* **1** 255–293.

GONÇALVES, S. and KILIAN, L. (2007). Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity. *Econometric Rev.* **26** 609–641. MR2415671

HANNAN, E. J. and KANTER, M. (1977). Autoregressive processes with infinite variance. *J. Appl. Probab.* **14** 411–415. MR0436502

HILL, J. B. (2013). Least tail-trimmed squares for infinite variance autoregressions. *J. Time Series Anal.* **34** 168–186. MR3028364

JAFFARD, S. (1990). Propriétés des matrices "bien localisées" près de leur diagonale et quelques applications. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **7** 461–476. MR1138533

KNIGHT, K. (1987). Rate of convergence of centred estimates of autoregressive parameters for infinite variance autoregressions. *J. Time Series Anal.* **8** 51–60. MR0868017

KNIGHT, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *Ann. Statist.* **17** 1168–1175. MR1015144

KREISS, J.-P. (1997). Asymptotic properties of residual bootstrap for autoregression. Technical report, TU Braunschweig. Available at https://www.tu-braunschweig.de/Medien-DB/stochastik/kreiss-1997.pdf.

KREISS, J.-P., PAPARODITIS, E. and POLITIS, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.* **39** 2103–2130. MR2893863

LEPAGE, R. (1992). Bootstrapping signs. In *Exploring the Limits of the Bootstrap* (R. Lepage and L. Billard, eds.) 215–224. Wiley, New York.

LEPAGE, R. and PODGÓRSKI, K. (1996). Resampling permutations in regression without second moments. *J. Multivariate Anal.* **57** 119–141. MR1392581

LEPAGE, R., WOODROOFE, M. and ZINN, J. (1981). Convergence to a stable distribution via order statistics. *Ann. Probab.* **9** 624–632. MR0624688

LEWIS, R. and REINSEL, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *J. Multivariate Anal.* **16** 393–411. MR0793499

MARONNA, R. A., MARTIN, R. D. and YOHAI, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley, Chichester. MR2238141

MCCULLOCH, J. H. (1997). Measuring tail thickness to estimate the stable index α: A critique. *J. Bus. Econom. Statist.* **15** 74–81. MR1435386

PAPARODITIS, E. (1996). Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *J. Multivariate Anal.* **57** 277–296. MR1391173

PÖTSCHER, B. M. and LEEB, H. (2009). On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding. *J. Multivariate Anal.* **100** 2065–2082. MR2543087

RESNICK, S. I. (1997). Heavy tail modeling and teletraffic data. *Ann. Statist.* **25** 1805–1869. MR1474072

SAMWORTH, R. (2003). A note on methods of restoring consistency to the bootstrap. *Biometrika* **90** 985–990. MR2024773

SILVERBERG, G. and VERSPAGEN, B. (2007). The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance. *J. Econometrics* **139** 318–339. MR2380742

STROHMER, T. (2002). Four short stories about Toeplitz matrix calculations. *Linear Algebra Appl.* **343/344** 321–344. MR1878948

G. CAVALIERE
DEPARTMENT OF STATISTICAL SCIENCES
UNIVERSITY OF BOLOGNA
VIA DELLE BELLE ARTI 41
I-40126 BOLOGNA
ITALY
E-MAIL: giuseppe.cavaliere@unibo.it

I. GEORGIEV
DEPARTMENT OF STATISTICAL SCIENCES
UNIVERSITY OF BOLOGNA
VIA DELLE BELLE ARTI 41
I-40126 BOLOGNA
ITALY
AND
NOVA SCHOOL OF BUSINESS AND ECONOMICS
UNIVERSIDADE NOVA DE LISBOA
CAMPUS DE CAMPOLIDE
1099-032 LISBON
PORTUGAL
E-MAIL: i.georgiev@unibo.it

A. M. R. TAYLOR
ESSEX BUSINESS SCHOOL
UNIVERSITY OF ESSEX
WIVENHOE PARK
CO4 3SQ COLCHESTER
UNITED KINGDOM
E-MAIL: rtaylor@essex.ac.uk