BOOTSTRAPPING NON-CAUSAL AUTOREGRESSIONS: WITH AN APPLICATION TO EXPLOSIVE BUBBLE MODELLING

Giuseppe Cavaliere*, Heino Bohn Nielsen†, and Anders Rahbek†

Abstract

In this paper we develop bootstrap-based inference for noncausal autoregressions with infinite variance, or heavy tailed, innovations. This class of models is widely used for modelling bubbles and explosive dynamics in economic and financial time series. In the noncausal, heavy tail framework, a major drawback of asymptotic inference is that it is not feasible in most empirical applications, as the relevant limiting distributions depend crucially on nuisance parameters and on the exact distribution of the distribution of the innovations, which is generally unknown. Specifically, the limiting behavior (including rates of convergence) depends on the decay rate of the tails of the distribution of the innovations. In addition, even in the unrealistic case where the tail behavior is known, asymptotic inference suffers from small-sample issues in terms of size when testing hypotheses of interest. In contrast, we propose here to resort to bootstrap implementations of least squares based inference, based on parameter estimates obtained with the null imposed (restricted bootstrap). We discuss three different choices of bootstrap innovations: standard wild bootstrap based on Rademacher errors; permutation bootstrap; a combination of the two ('permutation wild bootstrap'). Crucially, implementation of these bootstraps do not required any a priori knowledge about the distribution of the innovation, such as the tail index or the convergence rates of the estimators. We establish sufficient conditions ensuring that, under the null hypothesis, the bootstrap statistics estimate consistently particular conditional distributions of the original statistics. In particular, we show that validity of the permutation bootstrap holds without any restriction on the distribution of the innovations, while the permutation wild and the standard wild bootstraps require further assumptions such as symmetry of the innovation distribution. Extensive Monte Carlo simulations show that the finite sample performance of the proposed bootstrap tests is exceptionally good, both in terms of size and of empirical rejection probabilities under the alternative hypothesis. We conclude by considering an application of the proposed bootstrap inference to Bitcoin/USD exchange rates, where we find that noncausal models with heavy tailed innovations fit the data well also in periods of bubble dynamics.

Keywords: Noncausal Autoregressions; Heavy Tails; Bubble Dynamics; Bootstrap.
JEL Classification: C32.

*Department of Statistical Sciences, University of Bologna, Italy. †Department of Economics, University of Copenhagen, Denmark. Rahbek is grateful for support from Center for Information and Bubble Studies (CIBS) at the University of Copenhagen. CIBS is sponsored by the Carlsberg Foundation. Correspondence to: Anders Rahbek, Department of Economics, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen K; Denmark; email: anders.rahbek@econ.ku.dk

1 Introduction

In the recent literature there has been an increasing interest in noncausal processes with heavy tailed innovations, see Gourieroux and Zakoian (2017) and references therein. Despite their simplicity, these models, even in their basic form, are capable of mimicking periods of explosive, bubble-type dynamics and other types of complex, nonlinear behavior as witnessed repeatedly in financial and economic series; see Hecq, Lieb and Telg (2016).

A major drawback of this class of models, which limits its application in empirical works, is that estimation and testing based on asymptotic inference is not feasible in practice. In fact, as is well-known from the seminal works by Davis and Resnick (1985, 1986a,b), the relevant limiting distributions depend crucially on nuisance parameters and on the exact distribution of the innovations. Specifically, rates of convergence of estimators and test statistics as well as the form of the corresponding limiting distributions depend on tails of the distribution of the innovations. In addition, as is well known also for causal processes, even in the unrealistic case where the relevant asymptotic distributions were known, asymptotic inference would suffer from small-sample issues in terms of size when testing hypotheses of interest and of empirical coverage when interval estimation is considered.

In contrast, we propose here to resort to bootstrap implementations of least squares [LS] based inference. To the best our knowledge, bootstrap in non-causal autoregressions with heavy-tailed case has not been pursued in the literature. In the standard, *causal* framework, the presence of heavy tails make the standard (Efron's) bootstrap (based on i.i.d. resampling) an inconsistent estimator of the null distribution of the test statistics of interest, as is well known from Athreya (1987) and Knight (1999). Other types of bootstrap, such as the 'm out of n' bootstrap has been proposed, but their performance in finite sample is not always satisfactory and, importantly, makes it necessary to arbitrarily select the size (m) of the bootstrap sample. Moreover, the properties of these bootstraps in the noncausal world is largely unknown.

In this paper we take a different route, based on two main ingredients. First, we rely on bootstrap algorithms to generate noncausal bootstrap data where parameter estimates obtained with the null hypothesis imposed on the bootstrap sample. This is the so-called 'restricted bootstrap' (see, *inter alia*, Davidson and MacKinnon, 2006 and Cavaliere, Nielsen and Rahbek, 2015), which we here adapt to the noncausal framework ('unrestricted bootstrap' algorithms are also considered in our analysis although – as it will be demonstrated – they tend to be inferior to restricted bootstraps). Second, following a recent proposal made by Cavaliere, Georgiev and Taylor (2016) for causal processes, we discuss three different choices of bootstrap innovations: standard wild bootstrap based on Rademacher errors; permutation bootstrap; a combination of the two ('permutation wild bootstrap'). In contrast to (plain or m out of n) i.i.d. bootstrap methods, the proposed bootstraps – instead of resampling with replacement from a set of residuals, say $\tilde{\varepsilon}_t$ – work as follows. The wild bootstrap generates the bootstrap errors as $\varepsilon_t^* = w_t^* \tilde{\varepsilon}_t$ with $\{w_t^*\}$ an i.i.d. sequence independent of Rademacher random variables, i.e. satisfying $P(w_t^* = 1) = P(w_t^* = -1) = \frac{1}{2}$ (in contrast to the finite variance case, other choices for the distribution of w_t^* do not work). The permutation bootstrap generates the bootstrap innovations by simply taking a (uniformly distributed) random permutation of the residuals $\tilde{\varepsilon}_t$. Finally, the permutation-wild bootstrap, combines the two schemes and multiply the randomly permuted residuals with the random, Rademacher sequence $\{w_t^*\}$.

A crucial feature of these bootstraps is that their implementation and related computation of e.g. bootstrap p-values do not required *any* a priori knowledge about the distribution of the innovation, such as the previously mentioned tail index or the convergence rates of the estimators.

Although these bootstraps do not estimate the unconditional distribution of the statistics of interest, we establish sufficient conditions ensuring that, under the null hypothesis, the bootstrap statistics estimate consistently particular *conditional* distributions of the original statistics. This is particularly important because, even if the use of LS inference does not take the tail behavior featured by heavy tailed processes into account, the use of the aforementioned bootstraps allow to restrict the reference population with respect to which the test statistics are compared. For instance, the wild bootstrap estimates the limit null distributions of the original test statistics conditional on the absolute values of the innovations. This restricts the reference population and consequently an increase of power (with respect to the unfeasible, unconditional inference) is expected; see Cavaliere *et al.* (2013) for a similar result in a simple location model with i.i.d. errors. Similarly, the permutation bootstrap estimates the limit null distributions of the original test statistics conditional on the *order* statistics of the original innovations.

Our theoretical analysis show that validity of the permutation bootstrap holds without any restriction on the distribution of the innovations, while the permutation wild and the standard wild bootstraps require further assumptions such as symmetry of the innovation distribution. Interestingly, in the special case of first-order noncausal autoregressions we show that the wild bootstrap mimics exactly (i.e. with no estimation error) the conditional distribution of the original statistic, conditional on the absolute values of the original innovations.

Extensive Monte Carlo simulations show that the finite sample performance of the proposed bootstrap tests is exceptionally good, both in terms of size and of empirical rejection probabilities under the alternative hypothesis.

We conclude by considering an application of the proposed bootstrap inference to Bitcoin/USD exchange rates. These data were analyzed in a recent paper by Gourieroux and Hencic (2016) using approximate likelihood methods based on the Cauchy density. Using our proposed bootstraps, we find that a non-causal model with heavy tailed innovations fits the data well also in periods of bubble dynamics.

The paper is organized as follows. Section 2 provides an introduction to non-causal autoregressive processes with heavy tailed innovations, with focus on non-causal autoregressive processes of order $k = 1$, which we denote as $AR^+(1)$. In Section 3 we derive validity of the three bootstrap schemes described above, with a detailed theory on both finite sample and asymptotic properties. Next, in Section 4 we extend our asymptotic results to the general case of higher order non-causal autoregressions, denoted $AR^+(k)$. Finite sample properties of our bootstrap schemes for $AR^+(k)$ inference are investigated by Monte Carlo simulation in Section 5. Finally, an illustrative empirical application using Bitcoin data is provided in Section 6, and Section 7 concludes. All mathematical proofs are relegated in the Appendix, along with a short review of relevant results on heavy tailed processes.

2 Non-causal stable first-order autoregressions

Before presenting our results on general non-causal autoregressions in Section 4, we start in this section by discussing the non-causal $\text{AR}^+(1)$ process with heavy-tailed innovations. We first define the model in Section 2.1 and compare it with the standard, causal processes (denoted by $\text{AR}^-(k)$ in what follows) with finite variance. The time series properties of the process are then discussed in Section 2.2. Estimation and testing (and related asymptotic inference) are reviewed in Section 2.3.

2.1 The $\text{AR}^+(1)$ model

Consider initially the non-causal autoregressive model of order one, $AR^+(1)$, as given by the forward recursions,

$$x_t = \rho x_{t+1} + \varepsilon_t,\ t \in \mathbb{Z} \tag{1}$$

where ε_t is an i.i.d. sequence of random variables with 'heavy tails'. The heavy tails, as detailed below, excludes the standard case of finite variance innovations, $E\varepsilon_t^2 < \infty$. Instead we require that that the tails of the distribution decay at a slow rate; more precisely,

$$P(|\varepsilon_t| > x) \sim cx^{-\alpha} L(x) \tag{2}$$

for some constant $c > 0$, $\alpha \in (0, 2)$ and $L(x)$ is a slowly varying function at infinity; see Definition A.1 of Appendix A.

As is clear from the definition of the evolution of x_t in (1), the process is non-causal or autoregressive forward in time. This formulation implies (see also Gourieroux and Zakoian, 2017) that periods of explosive type behavior will occur, as often witnessed in the (exponential) build up and (sudden) decay of bubble phenomena in economic time series; classic examples are stock returns, bit-coin data and social media type (e.g. Twitter or Google search) data. It is also worth noting that under the heavy tail assumption, the standard Gaussian case is ruled out, consequently excluding that the non-causal representation is identical to a standard causal autoregression with the same parameter ρ.

A key example of heavy tailed innovation obtains when the ε_t's are *stable* random variables. Stable distributions (see e.g. Andrews et al., 2009, and references therein) are indexed by an exponent α (here restricted to lie in the open interval $(0, 2)$), usually labelled 'characteristic exponent' or 'tail index', a skewness parameter $|\beta| \leq 1$, a strictly positive and finite scale parameter $\sigma > 0$, and a location parameter $\mu \in \mathbb{R}$, which is set to zero in the following. If $\beta = 0$, the stable distribution is symmetric (about μ). The Cauchy distribution, which is specifically investigated in Gourieroux and Zakoian (2017), is a special case of the stable distributions as seen by setting $\alpha = 1$ and $\beta = 0$. The Gaussian distribution corresponds to $\alpha = 2$; in this case its mean and variance are given by μ and $2\sigma^2$, respectively. Also, for stable random variables (2) holds with $L(x) = 1$ and $c = (1+\beta) C_\alpha$ for some constant $C_\alpha > 0$ depending on α alone (see e.g. Nolan, 2016):

$$P(|\varepsilon_t| > x) \sim cx^{-\alpha}$$

such that ε_t exhibits the so-called Pareto-type tails. Hence, $E|\varepsilon_t|^p = +\infty$ for all $p \geq \alpha$ while $E|\varepsilon_t|^p < +\infty$ for $p < \alpha$. Consequently, for any $\alpha \in (0,2)$, ε_t has infinite variance while, for $\alpha \in (0,1)$, ε_t also has infinite unconditional expectation.

2.2 Time series properties

As in the causal case, the non-causal AR$^+$ (1) process x_t is strictly stationary provided $|\rho| < 1$. In this case, the strictly stationary solution has the one-sided moving average representation

$$x_t = \sum\nolimits_{j=0}^{\infty} \rho^j \varepsilon_{t+j} \tag{3}$$

which is convergent a.s. for any $\alpha \in (0,2)$. Notice also that x_t inherits the same tail and moment properties of ε_t (cf. Brockwell and Davis, 1991). In particular, x_t has Pareto-type tails and satisfies $E|x_t|^p < \infty$ for $p \in [0,\alpha)$ and $E|x_t|^p = \infty$ for $p \geq \alpha$; see Appendix A for details.

From Gourieroux and Zakoian (2016), the AR$^+$(1) with stable errors has, for $\rho \neq 0$ and $\beta = 0$, the surprising property of being a Markov chain with expectation, conditional on the past, given by

$$E(x_t|x_{t-1}) = \text{sgn}(\rho) |\rho|^{(\alpha-1)} x_{t-1}$$

For the standard Cauchy case (that is, $\beta = 0$ and $\alpha = 1$) with $\rho > 0$, this leads to the martingale property

$$E(x_t|x_{t-1}) = x_{t-1}. \tag{4}$$

In this case the variance of x_t, conditional on the past, changes over time and is given by

$$V(x_t|x_{t-1}) = \frac{\sigma^2}{\rho(1-\rho)} + \frac{1-\rho}{\rho} x_{t-1}^2 =: \sigma_{t|t-1}^2. \tag{5}$$

Taken together, (4) and (5) implies that in the Cauchy case x_t can be given a semi-strong double autoregressive (DAR) representation (Ling, 2004, 2007; Ling and Li, 2008, Nielsen and Rahbek, 2014)

$$x_t = x_{t-1} + \sigma_{t|t-1} z_t$$

with z_t a martingale difference sequence with unit conditional variance and a conditional density depending on x_{t-1}.

2.3 Estimation

In line with Davis and Resnick (1985, 1986a,b) and Gourieroux and Zakoian (2017), we consider the empirical autocorrelation coefficient or equivalently the least-squares [LS] estimator of ρ in the non-causal AR(1) model. This estimator is non-parametric, in that it does not utilize the specification of the distribution of ε_t. In the special case where the distribution of the ε_t's is known, maximum likelihood [ML] can be employed. For instance, where the ε_t's are assumed to be stable random variables, MLE can be employed along the lines proposed in Andrews *et al.* (2009), although its implementation is challenging due to the lack of closed-form expressions for the likelihood function[1]. However, in the general case where as here the ε_t's are only known to be in the domain of attraction of a stable distribution, MLE cannot be implemented. Other estimators, such least absolute deviation (LAD) estimators can also be considered; here we focus on LS estimation only (for LAD estimation, see Hecq et al., 2011).

The LS estimator (which corresponds to the Gaussian QMLE conditional on x_T fixed), is given by[2]

$$\hat{\rho}_T = \frac{\sum_{t=1}^{T-1} x_t x_{t+1}}{\sum_{t=1}^{T-1} x_{t+1}^2} =: \frac{S_{01}}{S_{11}} \tag{6}$$

Crucially, and in contrast to the usual finite variance case, the asymptotic properties of $\hat{\rho}_T$ depend on several (unknown) features of the distribution of ε_t. Importantly, not only the speed of convergence of the estimator depend on the unknown tail index α, but the existence and the form of the asymptotic distribution depend on whether the α-th order moment of $|\varepsilon_t|$ is finite or not – a property which cannot be assessed in practice.

As an example, consider the special case where ε_t is symmetric about 0 and $E|\varepsilon_t|^\alpha = +\infty$, $\alpha \in (0, 2)$. Then, it follows by Theorem 4.4 in Davis and Resnick (1986a) that for some normalizing sequence $n_T \to \infty$, as $T \to \infty$

$$n_T\left(\hat{\rho}_T - \rho\right) \xrightarrow{w} Z := \frac{1-\rho^2}{\left(1-\rho^\alpha\right)^{1/\alpha}} \frac{\mathcal{S}_1}{\mathcal{S}_0} \tag{7}$$

where $\mathcal{S}_0$ and $\mathcal{S}_1$ are independent stable random variables with index $\alpha/2$ and α, respectively (for the general case, see Appendix A.2). In particular, for the case where ε_t is asymptotically equal to a Pareto distribution (see Nolan, 2016), then (7) holds with $n_T = (T/\log(T))^{1/\alpha}$. Also, when $\alpha = 1$ and ε_t is stable (such that ε_t is Cauchy), the

[1] See also Breidt, Davis, Lii and Rosenblatt (1991), Hecq, Lieb and Telg (2015) and Lanne and Saikkonen (2011) for further applications of MLE to non-causal heavy tailed processes.

[2] One may also consider the sample autocorrelation coefficient, as given by

$$\hat{\phi}_T := \frac{\sum_{t=2}^{T} x_t x_{t-1}}{\sum_{t=1}^{T} x_t^2} = \left(1 + \frac{x_1^2}{\sum_{t=2}^{T} x_t^2}\right)\hat{\rho}_T \ ,$$

see Davis and Resnick (1986). For the present purposes the two formulations are equivalent and leads to the same results.

previous expression reduces to

$$\frac{T}{\log T}(\hat{\rho}_T - \rho) \to_w Z = (1+\rho)\frac{\mathcal{S}_1}{\mathcal{S}_0}. \tag{8}$$

In this special case, $\mathcal{S}_1$ is standard Cauchy ($C(0,1)$) and $\mathcal{S}_0$ Lèvy distributed on $(0,\infty)$, which also implies that $\mathcal{S}_0^{-1}$ is χ_1^2. Hence, Z of (8) is distributed as $(1+\rho)\,C(0,1)\,\chi_1^2$.

Studentized statistics show similar properties. For instance, consider the standard t-ratio

$$t_T := \frac{\hat{\rho}_T - \rho_0}{\hat{\sigma}_T S_{11}^{-1/2}}$$

where $S_{11} := \sum_{t=1}^{T-1} x_{t+1}^2$ and $\hat{\sigma}_T^2$ is the residual variance, $\hat{\sigma}_T^2 := T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$, for $\hat{\varepsilon}_t := x_t - \hat{\rho}_T x_{t+1}$. In the special case considered above it is straightforward to prove, see Lemma A.5, that $t_T = O_p\left(T^{1/2} n_T^{-1}\right)$; hence, in practice it is not even obvious how to normalize the Student t-statistic in order to perform asymptotic inference.

As the examples above clearly indicate, inference based on LS estimation appears to be infeasible. Specifically, the related asymptotic distributions depend on α, which is unknown in practice. Moreover, even in cases where α is known, inference is not feasible in general, as the normalizing sequence n_T may depend on further unknown quantities, see the next section and Appendix A.2. Hence, we next investigate the usefulness of the bootstrap in approximating the distribution of the LS estimator in the non-causal case. Crucially, as we will argue next, the bootstrap allows for feasible inference even when the tail index α and/or the normalizing sequence n_T are not known.

3 The Bootstrap for the non-causal AR$^+$(1)

In this section we discuss the bootstrap in the non-causal AR$^+$(1) model by focusing on tests of the null hypothesis $\mathsf{H}_0 : \rho = \bar{\rho}$ against the two sided alternative $\rho \neq \bar{\rho}$. We consider test statistics of the form

$$\text{(i): } r_T = \hat{\rho}_T - \bar{\rho} \qquad \text{(ii) } t_T = T^{1/2}\frac{\hat{\rho}_T - \bar{\rho}}{\hat{\sigma}_T S_{11}^{-1/2}} \tag{9}$$

with S_{11} and $\hat{\sigma}_T^2$ as defined earlier. As discussed in the previous section, the asymptotic distributions of (appropriately normalized versions of) r_T and t_T are generally unknown and asymptotic inference is unfeasible. We now discuss how to construct bootstrap versions of the test statistics in (9). The true value of the AR parameter is denoted by ρ_0. Also, in what follows the unrestricted least squares residuals are denoted by $\hat{\varepsilon}_t = x_t - \hat{\rho}_T x_{t+1}$ while the restricted residuals are denoted by $\tilde{\varepsilon}_t = x_t - \bar{\rho} x_{t+1}$, such that, under the null hypothesis, $\tilde{\varepsilon}_t = \varepsilon_t$ $(t = 1, ..., T-1)$.

We initially describe in section 3.1 the two main bootstrap schemes which we propose in the paper. The first is a restricted bootstrap scheme, where the bootstrap DGP bootstrap satisfies the null hypothesis, while the second is a classic unrestricted bootstrap, where the null hypothesis is not imposed on the bootstrap sample. Then, in section 3.2 we discuss the finite sample and asymptotic properties of bootstrap tests based on both schemes.

3.1 Bootstrap schemes

Given the autoregressive structure of the data generating process in (1), we propose two recursive design bootstrap schemes. The first is a restricted bootstrap in the sense that the null is imposed throughout (for a recent application to restricted bootstrap testing in causal autoregressions, see Cavaliere, Nielsen and Rahbek, 2015), whereas the second is an unrestricted bootstrap where the null is not imposed. For the first we are able to provide exact results, establishing that under the null hypothesis the bootstrap statistics in finite samples has the same distribution as the corresponding original statistics; for the latter, the unrestricted bootstrap, we provide asymptotic results. As already emphasized the bootstrap here is used to provide approximations of particular *conditional* distributions of the original statistics (where the conditioning set depends on the actual construction of the bootstrap innovations), rather than estimating the unconditional null distribution as is usually the case in the bootstrap literature.

3.1.1 Restricted bootstrap

For some bootstrap innovations ε_t^* $(t = 1, ..., T-1)$ to be described below, we define the bootstrap process x_t^*, $t = 1, ..., T$ by

$$x_t^* = \bar{\rho} x_{t+1}^* + \varepsilon_t^*, \; t = 1, ..., T-1$$

initialized at $x_T^* = x_T$. Notice that this corresponds to $x_t^* = \sum_{i=0}^{T-t-1} \bar{\rho}^i \varepsilon_{t+i}^* + \bar{\rho}^{T-t} x_T$ for $t = 1, ..., T$, see also the MA representation in (3).

The choice of the bootstrap innovations ε_t^* is crucial in the heavy tail framework. In particular, it is well known that the i.i.d. bootstrap where the ε_t^*'s are i.i.d. draws from the empirical distribution function of the residuals renders the bootstrap inconsistent, see e.g. Athreya (1987) and Knight (1989). Hence, we consider three alternative bootstrap re-sampling methods, which were also considered in Cavaliere, Georgiev and Taylor (2016) for the sieve bootstrap in classic causal AR models. The first is the so-called *permutation* bootstrap, where $\varepsilon_t^* = \tilde{\varepsilon}_{\pi^*(t)}$, $t = 1, 2, ..., T-1$ with $\{\pi^*(i)\}_{i=1}^{T-1}$ a uniformly distributed random permutation of $\{1, 2, ..., T-1\}$. The second is a standard *wild* bootstrap based on Rademacher bootstrap innovations, where $\varepsilon_t^* = w_t^* \tilde{\varepsilon}_t$ with $\{w_t^*\}_{t=1}^{T-1}$ an i.i.d. sequence independent of the original data and $P(w_t^* = 1) = P(w_t^* = -1) = \frac{1}{2}$. The third, the *permutation-wild* bootstrap, combines the two and sets $\varepsilon_t^* = w_t^* \hat{\varepsilon}_{\pi^*(t)}$ with w_t^* and $\pi^*(t)$ as defined before.

We bootstrap the two statistics in (9) by considering their bootstrap analogs as given by:

$$\text{(i) } r_T^* = \hat{\rho}_T^* - \bar{\rho} \qquad \text{(ii) } t_T^* = \frac{\hat{\rho}^* - \bar{\rho}}{\hat{\sigma}_T^* S_{11}^{*-1/2}} \tag{10}$$

3.1.2 Unrestricted bootstrap

With $\hat{\rho}_T$ defined in (6) we define the unrestricted bootstrap process $x_t^\dagger$, $t = 1, ..., T$ by

$$x_t^\dagger = \hat{\rho}_T x_{t+1}^\dagger + \varepsilon_t^\dagger, \; t = 1, ..., T-1 \tag{11}$$

with $x_t^\dagger$ initialized at $x_T^\dagger = x_T$, such that $x_t^\dagger = \sum_{i=0}^{T-t-1} \hat{\rho}_T^i \varepsilon_{t+i}^\dagger + \hat{\rho}_T^{T-t} x_T$ for $t = 1, ..., T$ as noted for the restricted bootstrap. The resampling methods are as before (i.e., wild bootstrap, permutation bootstrap and combination thereof) but with $\tilde{\varepsilon}_t$ now replaced by the unrestricted residuals, $\hat{\varepsilon}_t = x_t - \hat{\rho}_T x_{t+1}$. Notice that while for the restricted bootstrap under the null hypothesis it holds that $\tilde{\varepsilon}_t = \varepsilon_t$ (and hence resampling is from the true errors), for the unrestricted bootstrap it holds that $\hat{\varepsilon}_t = \varepsilon_t - (\hat{\rho}_T - \rho_0) x_{t+1}$. While this difference is not generally crucial in the finite variance case, under heavy tailed innovations the asymptotic properties of the term $(\hat{\rho}_T - \rho_0) x_{t+1}$ are essential in order to assess the validity of the bootstrap.

For the unrestricted bootstrap, the reference bootstrap statistics are given by

$$\text{(i) } r_T^\dagger = \hat{\rho}_T^\dagger - \hat{\rho}_T \quad \text{(ii) } t_T^\dagger = \frac{\hat{\rho}_T^\dagger - \hat{\rho}_T}{\hat{\sigma}_T^\dagger S_{11}^{\dagger -1/2}}$$

We now turn to the finite sample and asymptotic properties of these bootstrap schemes.

3.2 Properties of the bootstrap

In order to investigate the properties of our bootstrap schemes in the AR$^+$(1) model, we first focus on some exact, finite sample results for the restricted bootstrap. In particular we initially show that under the null hypothesis this bootstrap replicates (with no estimation error) specific *conditional* distributions of the original statistics, irrespective of the dimension of the sample. Since the same exactness property does not hold when the unrestricted bootstrap is employed, we then proceed by establishing that the bootstrap asymptotically replicate such conditional distributions.

3.2.1 Finite sample properties (restricted bootstrap)

As observed above, under the null hypothesis $\tilde{\varepsilon}_t = \varepsilon_t$ such that for the restricted bootstrap the permutation, wild or the combined bootstraps resample from the true ε_t's. This implies that we can state the following exact finite sample result involving the distribution of the restricted bootstrap statistics, conditional on the data, and the distribution of the original statistics, conditional on suitable functions of the data, when the null hypothesis H_0: $\rho = \bar{\rho}$ holds. Here $\mathcal{D}(Y|X)$ denotes the (possibly random) cumulative distribution function of Y given X. Moreover, $\varepsilon_{(t)}$ $(t = 1, ..., T-1)$ denote the order statistics of $\{\varepsilon_t\}$ while $|\varepsilon|_{(t)}$ $(t = 1, ..., T-1)$ denote the order statistics of $\{|\varepsilon_t|\}$.

Lemma 1 (Restricted bootstrap) *With the 'restricted bootstrap' statistics defined in (10), under the null hypothesis $\rho_0 = \bar{\rho}$, restricted bootstrap inference is exact in the following sense:*

(i) For the permutation bootstrap,

$$\mathcal{D}(r_T^* | \{x_t\}_1^T) = \mathcal{D}(r_T | \{\varepsilon_{(t)}\}_1^{T-1}, x_T)$$

where $\{\varepsilon_{(t)}\}_1^{T-1}$ *denotes the order statistics of* $\{\varepsilon_t\}_1^{T-1}$*;*

(ii) For the wild bootstrap, if the distribution of ε_t *is symmetric about 0,*

$$\mathcal{D}(r_T^*|\{x_t\}_1^T) = \mathcal{D}(r_T|\{|\varepsilon_t|\}_1^{T-1}, x_T) \ ;$$

(iii) For the combined permutation-wild bootstrap, if the distribution of ε_t *is symmetric about 0,*

$$\mathcal{D}(r_T^*|\{x_t\}_1^T) = \mathcal{D}(r_T|\left\{|\varepsilon|_{(t)}\right\}_1^{T-1}, x_T) \ .$$

The same results hold with r_T, r_T^* *replaced by* t_T, t_T^**, respectively.*

Remark 3.1 Lemma 1 implies that the restricted bootstrap replicate particular conditional distributions of the original statistics. Specifically, denote (to simplify notation) by F_r^* the distribution function of r_T^* conditional on the original data, and by F_r the distribution function of the original statistic r_T, conditionally on the statistics specified in the lemma. Then, for any metric $\varrho(\cdot)$ on the space of distribution functions, such as the Lèvy metric (see below), the distance $\varrho(F_r, F_r^*)$ equals 0. As shown in the simulations of Section 5 this is reflected by the excellent size properties displayed by the restricted bootstrap in samples of $T = 100$ observations.

Remark 3.2 The fact the bootstrap reproduces with no estimation error a particular conditional distribution of the original statistic is not enough for establishing that the bootstrap p-value, $p_T^* := P^*(r_T^* \leq r_T)$, is uniformly distributed under the null hypothesis, at least asymptotically. A sufficient condition for asymptotic validity would be that, as $T \to \infty$, the limiting distribution of r_T^* given the data is continuous with probability 1, see Cavaliere and Georgiev (2017). In a simple location model, if the wild bootstrap is employed continuity can be established along the lines of Knight (1989, p.1173-4). We conjecture that the same result carries over to the $AR^+(1)$ case. The simulation results in section 5 support this conjecture.

Remark 3.3 The results in Lemma 1 do not hold for the unrestricted bootstrap, and consequently for this bootstrap we instead establish asymptotic results in the next part of this section. As already indicated, the reason is that the difference between unrestricted and true residuals, given by $\hat{\varepsilon}_t - \varepsilon_t = -(\hat{\rho}_T - \bar{\rho})\, x_{t+1}$, is not zero and instead vanishes only asymptotically (under suitable conditions).

Remark 3.4 The symmetry assumption in Lemma 1(ii) and (iii) is crucial for validity of the wild and permutation-wild bootstraps. This is because the exactitude of these two bootstraps are based on the fact that, under symmetry of ε_t, under the null hypothesis the original innovation ε_t and its bootstrap analog ε_t^* have the same distribution conditionally on the absolute value $|\varepsilon_t|$. This distributional equality, which is crucial to the proof of Lemma 1, is clearly violated when ε_t is not symmetrically distributed. It is also worth noticing that the permutation bootstrap does not require the symmetry assumption to be exact in finite samples.

Remark 3.5 In the finite variance case, it is well known that the requirement on the bootstrap shocks w_t^* is that they have zero mean, unit variance and finite fourth order moment. On the contrary, in the heavy tail case w_t^* needs to have the two-point (Rademacher) distribution; other choices (such as the Gaussian or the much used two point distributions of Liu and Mammen) would undermine bootstrap validity in the sense of Lemma 1, also asymptotically, because ε_t^*, conditionally on the data, would not have the same distribution as ε_t, conditionally on the absolute value $|\varepsilon_t|$.

Remark 3.6 Importantly, Lemma 1 only requires i.i.d.-ness of the sequence $\{\varepsilon_t\}$ (and symmetry for the permutation and permutation wild schemes, see the previous remark) and hence allows for any value of the non-causal autoregressive parameter ρ as well as of the tail index α.

3.2.2 Large sample properties

We now turn to the large sample properties of the bootstrap. Since under the null hypothesis the restricted bootstrap is exact in samples of finite size, we here focus on the properties of the unrestricted bootstrap. In particular, we established that, under the null hypothesis, the permutation and the permutation-wild bootstraps behave asymptotically as the restricted bootstrap statistics. The same result does not hold for the standard wild bootstrap, as it will be clarified below.

In order to state our results, we make use of the following assumption, see also Davis and Resnick (1986) and the discussion therein.

Assumption 1 *With $\{x_t\}_1^T$ as given in (1), assume that (i) $|\rho| < 1$; (ii) the tail decay of $\{\varepsilon_t\}$ in (2) holds with $\alpha \in (0,2)$ and $\lim_{x\to\infty} P(\varepsilon_t > x)/P(|\varepsilon_t| > x) =: p \in [0,1]$, $\lim_{x\to\infty} P(\varepsilon_t < -x)/P(|\varepsilon_t| > x) = 1 - p$; (iii) $E|\varepsilon_t|^\alpha = +\infty$; (iv) for $\alpha \in (1,2)$, $E(\varepsilon_t) = 0$ while, for $\alpha = 1$, ε_t is symmetrically distributed.*

Assumption 1(i) is a standard stationarity condition for non-causal autoregressions, see also Section 2. Assumption 1(ii) is classic in the heavy tail literature and corresponds to assuming that $\{\varepsilon_t\}$ is in the domain of attraction of an α-stable law; see Appendix A. Assumption 1(iii) connects the tail index α to the (in)finiteness of the moments of ε_t. Finally, Assumption 1(iv) is a very mild requirement which essentially rules out some pathological cases arising at the singularity $\alpha = 1$.

The following lemma provides the asymptotic properties under the null hypothesis. Specifically, recall that in Lemma 1 we show that under the null hypothesis the distribution of restricted bootstrap statistic r_T^*, conditional on the original data, is identical to the distribution of original statistic r_T, conditional on appropriate transformations of the original innovations $\{\varepsilon_t\}$. Recall, additionally, that the original statistic r_T converges weakly at the rate n_T, see (7) for the symmetric case and Lemma A.4 in Appendix for the general case. In the next lemma we show that under the null hypothesis the unrestricted bootstrap statistic $r_T^\dagger$ and restricted bootstrap statistic r_T^* are close in the sense that

$$n_T(r_T^\dagger - r_T^*) = o_p^*(1),$$

in probability. Importantly, this implies that (under the null) the distribution of unrestricted bootstrap statistic $r_T^{\dagger}$, conditional on the original data, asymptotically coincides with the distribution of original statistic r_T, conditional on appropriate transformations of the original innovations $\{\varepsilon_t\}$. More precisely, the Lèvy metric between these two conditional distributions tends to zero, in probability (see also Remark 3.7).

LEMMA 2 *Let $\{x_t^{\dagger}\}_{t=1}^{T}$ be the unrestricted bootstrap process defined in (11) using either permutation or permutation-wild bootstrap shocks. Then, under Assumption 1, if the null hypothesis holds the unrestricted bootstrap and restricted bootstrap statistics $r_T^{\dagger}$, r_T^{*} satisfy*

$$n_T(r_T^{\dagger} - r_T^{*}) = o_{P^*}(1)$$

with n_T as given in Appendix B. Likewise, for the studentized bootstrap statistics $t_T^{\dagger}, t_T^{*}$, it holds that

$$\frac{n_T}{T^{1/2}}(t_T^{\dagger} - t_T^{*}) = o_{P^*}(1)\,.$$

REMARK 3.7 From Lemma 2 we are able to conclude that asymptotically and conditionally on the data, the unrestricted bootstrap test statistics behave as the restricted bootstrap test statistics. In particular, this implies, for the permutation bootstrap, that in terms of the Levy-metric $\varrho_L(\cdot)$,[3]

$$\varrho_L\left(F_r^{\dagger}, F_r^{*}\right) \xrightarrow{p} 0$$

where F_r^{*} ($F_r^{\dagger}$) is now the distribution function of the normalized restricted bootstrap statistic $n_T r_T^{*}$ (unrestricted bootstrap statistic) conditional on the original data. Hence, using also Lemma 1,

$$\varrho_L\left(F_r^{\dagger}, F_r\right) \xrightarrow{p} 0$$

with F_r the distribution function of the normalized original statistic $n_T r_T$, conditionally on the statistics specified in Lemma 1. Under symmetry of ε_t, an equivalent result holds for the permutation-wild bootstrap.

REMARK 3.8 While the permutation and the permutation wild bootstraps are asymptotically valid (in the Levy metric), a similar result does not hold for the standard wild bootstrap, even under the symmetry assumption. However, we show in section 5 that this bootstrap (in the case of symmetric innovations) has good finite sample size.

REMARK 3.9 From the inspection of the proofs it can be noticed that the result in Lemma 2 would apply if in the construction of the bootstrap sample, the LS estimator is replaced by any consistent estimator, say $\check{\rho}_T$, such that $\check{\rho}_T - \rho_0 = O_p(\hat{\rho}_T - \rho_0) = O_p(n_T^{-1})$.

[3]For a given $\eta > 0$, η-proximity of two cumulative distribution functions, say F and F^*, at a point x can be evaluated by means of the indicator $I_\eta^{F,F^*}(x) := \mathbb{I}(F^*(x-\eta) - \eta \leq F(x) \leq F^*(x+\eta) + \eta)$. Then, the Lévy metric between F and F^* is defined as follows:

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

4 Higher-order dynamics

We now discuss how the results of the previous sections generalize to higher order processes. Specifically, we consider the AR$^+(k)$ process, which for $k \geq 1$ is given by

$$x_t = \rho_1 x_{t+1} + ... + \rho_k x_{t+k} + \varepsilon_t = \beta' z_{t+1} + \varepsilon_t \tag{12}$$

where $z_t := (x_t, ..., x_{t+k-1})'$, $\beta := (\rho_1, ..., \rho_k)'$ and the ε_t's as defined before. Interest lies in testing a linear hypothesis of the form $\mathsf{H}_0 : R'\beta = r$, where R is $k \times 1$ and r scalar (extension to multiple hypotheses are considered in Remark 4.4 below). As is standard, we focus on the test statistics

$$\text{(i) } r_T = R'\hat{\beta}_T - r \quad \text{(ii) } t_T = \frac{R'\hat{\beta}_T - r}{\hat{\sigma}_T (R' S_{11}^{-1} R)^{1/2}} \tag{13}$$

where, with $S_{11} := \sum_{t=1}^{T-k} z_{t+1} z_{t+1}'$ and $S_{10} := \sum_{t=1}^{T-k} z_{t+1} x_t$, $\hat{\beta}_T := S_{11}^{-1} S_{10}$ is the LS estimator of β and $\hat{\sigma}_T^2 := T^{-1} \sum_{t=1}^{T-k} \hat{\varepsilon}_t^2$, $\hat{\varepsilon}_t := x_t - \hat{\beta}_T' z_{t+1}$, is the residual variance.

Consider first the restricted bootstrap based, as for the AR$^+$ (1), on parameters estimated under the hypothesis. This requires to compute the restricted LS estimator, $\tilde{\beta}_T$, and associated residuals, $\tilde{\varepsilon}_t := x_t - \tilde{\beta}_T' z_{t+1}$, $t = 1, ..., T-k$. The bootstrap sample is next generated recursively with H_0 imposed as

$$x_t^* = \tilde{\beta}_T' z_{t+1}^* + \varepsilon_t^*, \; t = 1, ..., T-k \tag{14}$$

and initialized at $x_t^* = x_t$, $t = T-k+1, ..., T$. The ε_t^*'s are as before based on permutation and permutation-wild bootstrap resampling of the restricted residuals $\tilde{\varepsilon}_t$'s. The wild bootstrap is invalid in this case and hence it is not considered (see Remark 4.2 below). The restricted bootstrap statistics r_T^* and t_T^* are computed as

$$\text{(i) } r_T^* = R'\hat{\beta}_T^* - r \quad \text{(ii) } t_T^* = \frac{R'\hat{\beta}_T^* - r}{\hat{\sigma}_T^* (R' S_{11}^{*-1} R)^{1/2}}$$

where, with $S_{11}^* := \sum_{t=1}^{T-k} z_{t+1}^* z_{t+1}^{*\prime}$ and $S_{10}^* := \sum_{t=1}^{T-k} z_{t+1}^* x_t$, $\hat{\beta}_T^* := S_{11}^{*-1} S_{10}^*$ is the LS estimator of β obtained on the bootstrap sample and $\hat{\sigma}_T^{*2} := T^{-1} \sum_{t=1}^{T-k} \hat{\varepsilon}_t^{*2}$, $\hat{\varepsilon}_t^* := x_t^* - \hat{\beta}_T^{*\prime} z_{t+1}^*$, the corresponding residual variance.

As before, and taking the r_T statistic to illustrate, let $\varrho_L(\cdot, \cdot)$ denote the Levy metric in the space of distribution functions. Moreover, let F_r be the distribution function of (normalized) original statistic, $n_T r_T$, with r_T as in (13), conditionally on $\{\varepsilon_{(t)}\}_1^{T-k}, \{x_T\}_{t=T-k+1}^T$ for the permutation bootstrap, and conditionally on $\{|\varepsilon|_{(t)}\}_1^{T-k}, \{x_T\}_{t=T-k+1}^T$ for the permutation wild bootstrap. Finally, let F_r^* denote the distribution function of the normalized bootstrap statistic $n_T r_T^*$, conditionally on the original data. The following Theorem holds under the null hypothesis H_0 as the sample size diverges.

Theorem 1 *With $\{x_t\}_1^T$ as given in (12), assume that ε_t satisfy condition (ii)-(iv) in Assumption 1 and that all characteristic roots associated to (12) are outside the unit disk in the complex plan. Then, under the null hypothesis,*

$$\varrho_L\left(F_r^*, F_r\right) \xrightarrow{p} 0 \tag{15}$$

as $T \to \infty$.

Remark 4.1 Differently from the $\mathrm{AR}^+(1)$ case, for general linear hypotheses the restricted bootstrap is not exact under the null hypothesis. This happens because, even under the null hypothesis, the restricted residuals $\tilde{\varepsilon}_t$ differ from the true innovations, $\hat{\varepsilon}_t$. For the permutation and the permutation wild bootstrap, however, the difference $\tilde{\varepsilon}_t - \varepsilon_t$ vanishes sufficiently fast to ensure that the convergence in (15).

Remark 4.2 Because under the null hypothesis the difference $\tilde{\varepsilon}_t - \varepsilon_t$ does not equal zero, the restricted bootstrap for $\mathrm{AR}^+(k)$ models in the case of general hypotheses have similar features to the unrestricted bootstrap for $\mathrm{AR}^+(1)$ models in the case of simple hypotheses. For instance, as in the latter case the wild bootstrap scheme does not allow to mimic a proper conditional distribution of the original statistic, and hence fails to be asymptotically valid under the null hypothesis.

Remark 4.3 The unrestricted bootstrap based on the permutation and permutation wild schemes (the latter under symmetry) can be proved to be valid too. Simulation results provided in Section 5 confirm this.

Remark 4.4 The case of multivariate (q-dimensional, $q \geq 1$) hypotheses such as $\mathsf{H}_0 : R'\beta = r$, where R is now $k \times q$ (with full column rank q) and r is $q \times 1$ can be studied by bootstrapping F-type statistics of the form $F_T := \left(q\hat{\sigma}_T^2\right)^{-1}(R'\hat{\beta}_T - r)'(R'S_{11}^{-1}R)^{-1}(R'\hat{\beta}_T - r)$. With $\tilde{\beta}$ the LS estimator restricted by H_0, the bootstrap sample is generated recursively as (14) and the bootstrap statistic F_T^* corresponds to F_T computed on the bootstrap sample. Theorem 1 can be extended and proved to be valid in this case too.

5 Finite sample simulations

To illustrate the finite-sample properties of the proposed bootstrap tests in the non-causal autoregression framework, we present here a small set of Monte Carlo simulations. As in sections 2 and 3, we consider the non-causal $\mathrm{AR}^+(1)$ model (1). The data generating process in the simulations has $\rho_0 = 0.5$, and innovations distributed according to a stable law, $\varepsilon_t \sim S(\alpha, \beta)$, for the four combinations $(\alpha, \beta) \in (1, 1.5) \times (0, 0.75)$, and $T = 100$ observations. The hypothesis of interest is of the form

$$\mathsf{H}_0 : \rho = \bar{\rho},$$

against the two-sided alternative $\rho \neq \bar{\rho}$. The test statistics are r_T and its studentized version, t_T, as given in Section 2.

Table 1 presents results for the restricted bootstrap based on the three considered resampling schemes, i.e. the wild bootstrap, the permutation bootstrap and the permutation-wild bootstrap. All bootstrap tests are implemented with 999 bootstrap samples. Results for the corresponding unrestricted bootstraps are given in Table 2.

Section (A) in Table 1 focuses on the size properties and reports empirical rejection frequencies (ERFs) at nominal significance levels 2.5%, 5% and 10%, for tests of a true hypothesis, such that $\bar{\rho} = \rho_0 = 0.5$. We observe that the permutation bootstrap has excellent size properties, with ERFs close to the nominal levels, for all combinations of (α, β). The wild bootstrap and the permutation-wild bootstrap show inflated ERFs in cases with an asymmetric distribution, most severe when $\alpha = 1$. The differences between the statistic r_T and the studentized version, t_T, are small, although the bootstrap test based on the statistic r_T is generally preferable.

Section (B) of the table presents ERFs (at a nominal 5% level) for different values of $\bar{\rho}$, $\bar{\rho} \in \{0, 35, 0.40, ..., 0.65\}$, and focuses on the power-properties of the proposed test procedures. We observe that the permutation bootstrap test and the permutation-wild bootstrap test have reasonable power in all cases, while the wild bootstrap test, as expected, performs poorly when the error distribution are asymmetric. Again, the non-studentized statistic, r_T, is generally preferable, although the difference in power is not large.

The overall conclusion from Table 1 is that, in line with the theoretical claims in Section 3.2, all restricted bootstrap tests seem to work is cases of symmetric error distributions, while only the permutation bootstrap is valid in asymmetric cases.

Turning to the unrestricted bootstrap, results in Table 2 support the overall conclusion that the restricted bootstraps are slightly preferable over the corresponding unrestricted bootstraps in terms of size, in particular in the asymmetric cases.

In additional sets of unreported simulations[4], we have investigated the possible effects of different centering schemes for the estimated residuals, in particular using centering around the median of the residuals and around their sample mean (notice that our theory does not cover the use of re-centred residuals). On the one hand, these recentering schemes tend to inflate the empirical size of the permutation bootstrap in cases of asymmetric distributions, and in general do not lead to any improvement of the performance of the bootstrap tests in terms of size. On the other hand, in terms of (size-corrected) ERFs under the alternative, we observe that – only for the special case of the wild bootstrap – centering increases the rejection probabilities. The permutation bootstrap is not affected in terms of ERFs under the alternative by the possible recentering of the residuals.

[4] The full set of Monte Carlo simulations is available from the authors upon request.

				(A) ERF, Null			(B) ERF, Alternative, $\bar{\rho}$						
Bootstrap scheme	Statistic	α	β	2.5	5.0	10.0	0.35	0.40	0.45	0.50	0.55	0.60	0.65
Wild	r_T	1.0	0.00	2.3	4.7	9.6	34.2	25.5	16.1	4.7	31.8	53.1	69.3
	t_T	1.0	0.00	2.3	4.7	9.4	61.0	46.5	26.6	4.7	26.7	46.7	64.2
	r_T	1.0	0.75	26.0	34.7	45.6	70.0	67.4	63.6	34.7	10.0	18.7	31.4
	t_T	1.0	0.75	33.8	41.8	50.9	91.1	86.2	76.9	41.8	10.1	16.1	27.3
	r_T	1.5	0.00	2.6	5.3	10.0	32.6	18.3	8.0	5.3	15.9	34.2	55.0
	t_T	1.5	0.00	2.7	5.3	9.9	44.9	27.2	12.1	5.3	12.5	28.0	47.7
	r_T	1.5	0.75	4.0	7.5	13.4	36.0	20.4	9.4	7.5	17.2	32.1	50.5
	t_T	1.5	0.75	4.5	8.0	13.8	47.2	28.8	13.9	8.0	14.4	26.9	43.8
Permutation	r_T	1.0	0.00	2.7	5.4	10.6	61.0	37.5	18.1	5.4	20.1	43.9	70.9
	t_T	1.0	0.00	2.8	5.2	10.4	65.2	41.3	20.0	5.2	19.2	40.1	66.2
	r_T	1.0	0.75	2.6	5.0	9.5	68.9	44.3	21.1	5.0	12.3	25.7	49.3
	t_T	1.0	0.75	2.5	4.9	9.4	70.7	47.0	22.7	4.9	12.2	23.1	40.2
	r_T	1.5	0.00	2.7	5.5	10.3	44.8	23.8	9.3	5.5	13.2	32.4	57.9
	t_T	1.5	0.00	2.7	5.4	10.2	49.9	27.9	11.6	5.4	11.5	28.5	52.8
	r_T	1.5	0.75	2.6	5.4	10.4	47.2	26.0	10.8	5.4	14.8	35.0	58.5
	t_T	1.5	0.75	2.7	5.2	10.4	50.9	29.3	12.4	5.2	12.5	29.8	52.7
Permutation-wild	r_T	1.0	0.00	2.7	5.3	10.4	60.9	36.9	17.7	5.3	20.6	45.0	75.6
	t_T	1.0	0.00	2.7	5.3	10.3	66.1	41.7	20.1	5.3	19.2	39.5	66.8
	r_T	1.0	0.75	4.1	9.6	21.4	90.1	71.0	37.7	9.6	13.8	24.8	43.7
	t_T	1.0	0.75	6.9	13.3	25.1	92.3	76.9	45.3	13.3	15.1	23.5	38.1
	r_T	1.5	0.00	2.7	5.3	10.2	45.3	23.2	9.2	5.3	14.2	34.2	61.2
	t_T	1.5	0.00	2.8	5.5	10.0	51.1	28.5	11.7	5.5	11.6	28.2	53.1
	r_T	1.5	0.75	2.4	5.2	11.3	47.5	26.3	11.3	5.2	13.6	33.6	57.6
	t_T	1.5	0.75	3.2	6.2	11.9	53.3	31.6	14.5	6.2	11.9	27.9	50.9

TABLE 1: Simulation results for the Wild bootstrap, the Permutation bootstrap and the Permutation-wild bootstrap for the statistics r_T and t_T. Section (A) reports empirical rejection frequencies (ERF) for the bootstrap test of a true null hypothesis, $\bar{\rho} = \rho_0 = 0.5$, for significance levels 2.5%, 5%, and 10%. Section (B) reports ERF for the bootstrap tests for different values of $\bar{\rho}$, $\bar{\rho} \in \{0.35, 0.40, ..., 0.65\}$. The innovations of the DGP are drawn from a stable distribution, $S(\alpha, \beta)$, and $T = 100$. Results are based on 999 bootstrap samples and 10000 Monte Carlo replications.

				(A) ERF, Null			(B) ERF, Alternative, $\bar{\rho}$						
Bootstrap scheme	Statistic	α	β	2.5	5.0	10.0	0.35	0.40	0.45	0.50	0.55	0.60	0.65
Wild	r_T	1.0	0.00	2.6	5.0	9.9	35.5	26.7	17.6	5.0	32.1	53.5	69.7
	t_T	1.0	0.00	2.7	5.1	9.8	61.4	47.2	27.5	5.1	27.5	47.5	64.7
	r_T	1.0	0.75	29.2	37.3	47.0	73.0	71.2	67.3	37.3	10.0	18.8	32.4
	t_T	1.0	0.75	36.7	43.9	52.2	91.3	86.9	78.3	43.9	10.4	16.2	28.1
	r_T	1.5	0.00	2.8	5.5	10.5	33.8	19.3	8.7	5.5	16.3	34.6	55.3
	t_T	1.5	0.00	2.9	5.5	10.1	45.3	28.2	12.7	5.5	13.0	28.2	48.3
	r_T	1.5	0.75	7.1	10.3	16.2	31.0	18.2	9.5	10.3	21.8	36.7	54.2
	t_T	1.5	0.75	7.4	10.8	16.6	41.9	25.4	12.8	10.8	19.6	32.2	48.4
Permutation	r_T	1.0	0.00	2.7	5.4	10.6	61.6	38.0	18.3	5.4	20.2	44.3	71.4
	t_T	1.0	0.00	2.8	5.2	10.4	66.1	42.2	20.3	5.2	19.3	40.0	66.2
	r_T	1.0	0.75	4.5	8.4	16.1	89.0	68.0	34.0	8.4	13.3	25.7	48.3
	t_T	1.0	0.75	4.7	8.6	16.1	90.2	70.3	35.9	8.6	13.1	23.2	39.4
	r_T	1.5	0.00	2.7	5.1	10.2	44.9	23.4	9.2	5.1	13.6	33.0	58.5
	t_T	1.5	0.00	2.8	5.5	10.3	51.1	28.8	12.0	5.5	11.6	28.7	53.3
	r_T	1.5	0.75	3.3	6.6	12.8	49.2	27.7	12.6	6.6	16.5	38.9	63.0
	t_T	1.5	0.75	3.3	6.4	12.3	52.5	30.9	14.6	6.4	12.0	28.6	52.0
Permutation-wild	r_T	1.0	0.00	2.7	5.3	10.5	61.1	36.8	17.8	5.3	20.5	45.0	75.5
	t_T	1.0	0.00	2.7	5.3	10.3	66.0	41.8	20.1	5.3	19.2	39.4	66.9
	r_T	1.0	0.75	4.2	9.8	21.9	90.2	71.7	38.6	9.8	13.8	24.8	43.7
	t_T	1.0	0.75	6.9	13.4	25.6	92.4	77.3	46.2	13.4	15.3	23.6	38.2
	r_T	1.5	0.00	2.8	5.3	10.2	45.1	23.2	9.1	5.3	14.2	34.2	61.3
	t_T	1.5	0.00	2.8	5.5	10.2	51.0	28.5	11.7	5.5	11.6	28.1	53.1
	r_T	1.5	0.75	2.4	5.3	11.6	47.4	26.3	11.3	5.3	13.7	33.7	57.8
	t_T	1.5	0.75	3.1	6.3	12.2	53.3	31.6	14.4	6.3	11.9	28.0	51.0

TABLE 2: Simulation results for the unrestricted Wild bootstrap, the Permutation bootstrap and the Permutation-wild bootstrap for the statistics r_T and t_T. Section (A) reports empirical rejection frequencies (ERF) for the bootstrap test of a true null hypothesis, $\rho = 0.5$, for significance levels 2.5%, 5%, and 10%. Section (B) reports ERF for the bootstrap tests for different values of ρ, $\rho \in \{0.35, 0.40, ..., 0.65\}$. The innovations of the DGP are drawn from a stable distribution, $S(\alpha, \beta)$, and $T = 100$. Results are based on 999 bootstrap samples and 10000 Monte Carlo replications.

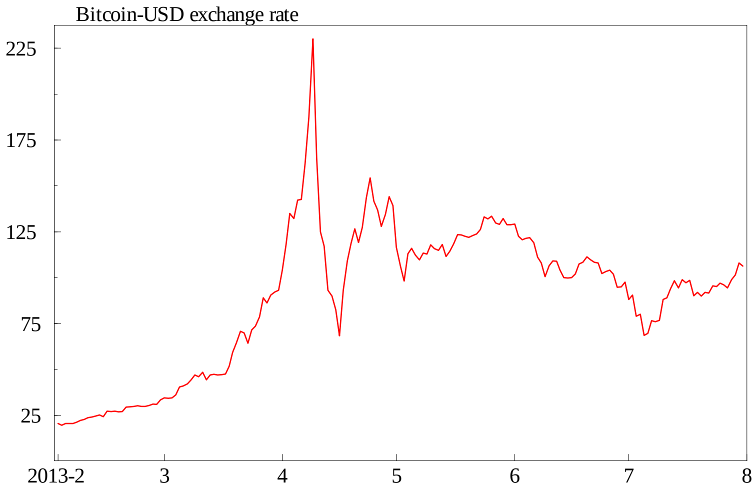

FIGURE 1: Daily Bitcoin-USD exchange rate from February 1 – July 31, 2013.

6 Empirical illustration

We consider daily Bitcoin-USD exchange rates from the MtGox Bitcoin exchange in Figure 1. Gourieroux and Hencic (2015) provide a detailed description and discussion of the data and find that an $\text{AR}^+(k)$ model with $k = 2$ is adequate for the sample period February–July 2013 with a total of $T = 181$ observations.[5]

In line with our derived results we proceed by analyzing the Bitcoin data using a general to specific approach by initially estimating an $\text{AR}^+(5)$ model applying the analyzed least squares estimator. The $\text{AR}^+(5)$ model is given by (12) with $k = 5$ and consequently $\beta = (\rho_1, ..., \rho_5)'$. Our theory allows us step-wise to test $k - 1$ against k, and also a joint test for several restrictions. As an example, the joint test for $k = 2$ against $k = 5$ corresponds to the hypothesis $R'\beta = 0$ with

$$R' = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

As is clear from the reported bootstrap p-values in Table 3, both the stepwise and joint tests allow a reduction to an $\text{AR}^+(2)$ model, thus confirming the findings of Hencic and Gourieroux (2015). The validity of the reduction from an $\text{AR}^+(k)$ model of order $k = 5$ to $k = 2$ holds for all variants of the bootstrap implementations, while the wild bootstrap (borderline) suggests a further reduction to an $\text{AR}^+(1)$.

The fitted $\text{AR}^+(2)$ model is given by

$$x_t = \hat{\rho}_1 x_{t+1} + \hat{\rho}_2 x_{t+2} + \hat{\varepsilon}_t,$$

[5] We use daily closing prices obtained from www.quandl.com/collections/markets/bitcoin-data.

k	AR$^+(k)$ 5	4	3	2
$\hat{\rho}_1$	1.212	1.214	1.216	1.203
$\hat{\rho}_2$	−0.287	−0.285	−0.296	−0.230
$\hat{\rho}_3$	0.028	0.011	0.055	
$\hat{\rho}_4$	−0.026	0.036		
$\hat{\rho}_5$	0.051			
	p-value for stepwise test AR$^+(k-1)$ against AR$^+(k)$			
Wild	0.651	0.783	0.760	0.143
Permutation	0.453	0.591	0.406	0.000
Permutation-wild	0.457	0.572	0.392	0.000
	p-value for joint test AR$^+(2)$ against AR$^+(k)$			
Wild	0.968	0.969	0.760	
Permutation	0.877	0.860	0.406	
Permutation-wild	0.857	0.859	0.392	

TABLE 3: Empirical analysis of Bitcoin data. Bootstrap p-values are based on 999 bootstrap replications.

where x_t denote the deviation of the Bitcoin/USD exchange rate from its sample average; $\hat{\varepsilon}_t = x_t - \hat{\rho}_1 x_{t+1} - \hat{\rho}_2 x_{t+2}$ denotes the estimated residuals, and parameter estimates are as given in Table 3.

In order to study the properties of the dependence structure of estimated residuals Gourieroux and Zakoian (2016) propose to study the empirical correlation of the $\hat{\varepsilon}_t$'s for an AR$^+(1)$ model,

$$\sum_{t=2}^{T-1} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} / \sum_{t=1}^{T-1} \hat{\varepsilon}_t^2, \tag{16}$$

whose (upon normalization) asymptotic distribution is found by simulation for the special case where the distribution of the ε_t's is known. The empirical correlation in (16) is closely related to the LS estimator in (6) and therefore, upon an appropriate normalization, has a limiting distribution which can be represented in terms of a ratio of stable distributions, see (A.8). Our proposed strategy of estimating an AR$^+(k)$ model first, and then testing by bootstrap p-values whether a reduction to an AR$^+(k-1)$ model is valid, is equivalent in the sense that the test statistic for a reduction from the AR$^+(2)$ to the AR$^+(1)$ has similar asymptotic properties to (16). Our results are more general, however, as they allow testing the AR$^+(k)$ structure for any $k \geq 1$, and importantly it is made feasible by the proposed bootstrap inference which can be used independently of the tail index (and, more generally, of the distribution) of the innovations ε_t.

As emphasized in Gourieroux and Zakoian (2016), the $\mathrm{AR}^+(1)$ model with Cauchy innovations has a causal recursive double autoregressive structure, hence motivating a misspecification analysis for both autocorrelation and ARCH-type effects in the causal $\mathrm{AR}^-(1)$ model. Diagnostics of 'causal vs. non-causal' can be based on the estimated non-causal residuals, $\hat{\varepsilon}_t$, and estimated causal residuals, $\hat{\varepsilon}_t^-$ say, for autocorrelation and ARCH-type effects. However, the result that the empirical autocovariance between $\hat{\varepsilon}_t$ and $\hat{\varepsilon}_{t+1}$, and hence the autocorrelation, have a non-standard non-pivotal limiting distribution, implies that standard Lagrange multiplier type tests for autocovariance are not valid (when based on conventional asymptotic Gaussian-based p-values), and likewise for the standard Lagrange multiplier statistics for ARCH effects, as these are based on covariances between $\hat{\varepsilon}_t^2$ and its lag(s).

To overcome this difficulty we propose to use Spearman's rank statistic, as given by

$$S_T\left((v_t, w_t)_{t=1}^T\right) = R_T\left(\frac{T-2}{1-R_T^2}\right)^{1/2},$$

where $R_T = 1 - 6\sum_{i=1}^T d_i^2/\left(T\left(T^2-1\right)\right)$ with $d_i = \mathrm{rk}_i\left((v_t)_{t=1}^T\right) - \mathrm{rk}_i\left((w_t)_{t=1}^T\right)$ the difference between the two ranks of the observations $(v_t, w_t)_{t=1}^T$ and S_T is approximately $t(m)$-distributed with $m = T-2$ degrees of freedom, see e.g. Moran (1950). The results in Table 4 apply for the non-casual case $(v_t, w_t) = (\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1})$ (and for the causal case $(\hat{\varepsilon}_t^-, \hat{\varepsilon}_{t-1}^-)$) to test for zero autocorrelation or level dependence, and $(v_t, w_t) = (\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}^2)$ $((\hat{\varepsilon}_t^-, (\hat{\varepsilon}_{t-1}^-)^2))$ to test for double autoregressive ARCH type dependence. We find that for both the non-causal $\mathrm{AR}^+(2)$ and causal $\mathrm{AR}^-(2)$ there is no autocorrelation, and stronger ARCH effects for the causal $\mathrm{AR}^-(2)$ model, supporting further the $\mathrm{AR}^+(2)$ model.

7 Concluding remarks

This paper establishes validity of bootstrap-based inference for pure noncausal $\mathrm{AR}^+(k)$ processes, thereby making inference feasible with general heavy tailed innovations. In terms of the bootstrap algorithm, we propose to apply the restricted bootstrap, *i.e.* parameters estimated under the null hypothesis are used for generating the bootstrap sample, together with innovations resampled by permutation bootstrap or wild bootstrap, or a combination thereof. From a detailed simulation study as well as the empirical illustration it can be seen that the proposed bootstrap inference is not only simple to implement in practice but also works very well in finite samples. In order to distinguish noncausal from causal processes, we apply diagnostics checking based on Spearman rank statistics in terms of both causally and noncausally estimated autoregressive residuals. As suggested by Gourieroux and Zakoian (2017), a pure noncausal process will upon estimation have residuals which are uncorrelated in levels and squares over time, while

	(v_t, w_t)	Spearman corr., R_T	Test statistic, S_T	p-value
Autocorrelation				
Non-causal	$(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1})$	0.017	0.228	0.820
Causal	$(\hat{\varepsilon}_t^-, \hat{\varepsilon}_{t-1}^-)$	−0.010	−0.127	0.899
ARCH-type				
Non-causal	$(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}^2)$	0.023	0.309	0.758
Causal	$(\hat{\varepsilon}_t^-, (\hat{\varepsilon}_{t-1}^-)^2)$	0.130	1.747	0.082

Table 4: Misspecification tests for non-causal residuals, $\hat{\epsilon}_t = x_t - \hat{\rho}_1 x_{t+1} - \hat{\rho}_2 x_{t+2}$, and causal residuals, $\hat{\epsilon}_t^- = x_t - \hat{\rho}_1^- x_{t-1} - \hat{\rho}_2^- x_{t-2}$, where $\hat{\rho}_1^-$ and $\hat{\rho}_2^-$ denote the LS estimates in the causal $AR^-(2)$ model.

if estimated by a causal autoregression, correlation in the squares will be detectable, as was confirmed in our included illustration based on Bitcoin data. Finally, as is well-known we stress that it is also of interest to consider *mixed* causal and non-causal processes. For these however, due to identification issues addressed in *e.g.* Hecq *et al.* (2016), it is required to extend our bootstrap theory to include (quasi-)maximum likelihood or LAD-type estimation and inference, which we leave for future research.

References

Andrews B., M. Calder and R.A. Davis (2009). Maximum Likelihood Estimation for α-Stable Autoregressive Processes, *The Annals of Statistics* 37, 1946-1982

Athreya, K.B. (1987) Bootstrap of the Mean in the Infinite Variance Case, *The Annals of Statistics* 15, 724–731.

Berk K. (1974). Consistent autoregressive spectral estimates, *The Annals of Statistics* 2, 489-502.

Breid F.J., R.A. Davis, K.-S. Lh and M. Rosenblatt (1991), Maximum likelihood estimation for noncausal autoregressive processes, *Journal of Multivariate Analysis* 36, 175–198

Brockwell, P.J. and R. Davis (1991). *Time Series: Theory and Methods.* New York: Springer.

Cavaliere G. and I. Georgiev (2017). Bootstrap Inference Under Random Distributional Limits, Working paper.

Cavaliere G., I. Georgiev and A.M.R. Taylor (2016). Sieve-based inference for infinite-variance linear processes, *The Annals of Statistics* 44, 1467–1494 (with Supplement).

Cavaliere, G., H.B. Nielsen and A. Rahbek (2015): "Bootstrap Testing of Hypotheses on Co-Integration Relations in Vector Autoregressive Models", *Econometrica,* 83(2), 813–831.

Chan, N.H. and Tran, L.T. (1989). On the first order autoregressive process with infinite variance. *Econometric Theory* 5, 354-362.

Davidson R. and J. MacKinnon (2006). The Power of Bootstrap and Asymptotic Tests, *Journal of Econometrics* 133, 421–441.

Davis R. and S. Resnick (1985a). Limit theory for moving averages of random variables with regularly varying tail probabilities, *The Annals of Probability* 13, 179-195.

Davis R. and S. Resnick (1985b). More limit theory for the sample correlation function of moving averages, *Stochastic Processes and their Applications* 20, 257-279.

Davis R. and S. Resnick (1986). Limit theory for the sample covariance and correlation functions of moving averages, *The Annals of Statistics* 14, 533-558.

Feller W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 2, John Wiley & Sons.

Gourieroux C. and A. Hencic (2016). Noncausal Autoregressive Modelling in Application to Bitcoin/USD exchange rates, in V.-N. Huynh, V. Kreinovich, S. Sriboonchitta and K. Suriya (eds.) *Econometrics of Risk*, 17–40.

Gourieroux C. and J.-M. Zakoïan (2016). Local explosion modelling by noncausal process, *Journal of the Royal Statistical Society B*, forthcoming.

Hecq A., L. Lieb and S. Telg (2015). Identification of Mixed Causal-Noncausal Models in Finite Samples, *Annals of Economics & Statistics* 123/124, 307–331.

Knight K. (1989). On the bootstrap of the sample mean in the infinite variance case, *The Annals of Statistics* 17, 1168-1175.

Lanne M. and P. Saikkonen (2011), Noncausal Autoregressions for Economic Time Series, *Journal of Time Series Econometrics* 3, Issue 3, Article 2.

Ling, S. (2004), Estimation and Testing of Stationarity for Double-Autoregressive Models, *Journal of the Royal Statistical Society* B 66, 63–78.

Ling, S. (2007). A Double AR(p) Model: Structure and Estimation, *Statistica Sinica* 17, 161–175.

Ling, S. and D. Li (2008). Asymptotic Inference for a Nonstationary Double AR(1) Model, *Biometrika* 95, 257–263.

Mikosh, T. (1999). Regular Variation, Subexponentiality and their Applications in Probability Theory. Lecture notes, University of Copenhagen.

Moran, P.A.P. (1950). Recent Developments in Ranking Theory, *Journal of the Royal Statistical Society B* 12, 153–162.

Nielsen H.B. and A. Rahbek (2014), Unit Root Vector Autoregression with Volatility Induced Stationarity, *Journal of Empirical Finance* 29, 144–167.

Nolan, J.P. (2016). *Stable Distributions: Models for Heavy-Tailed Data*, New York: Springer.

A Appendix

A.1 Large T results for heavy tail sequences

We start by introducing the concept of regularly and slowly varying functions at infinity, which are useful in the characterization of distributions with heavy tails such as Pareto-type tails. These functions are defined as follows.

Definition A.1 *A (positive and measurable) function $g(x)$ is said to be* regularly varying *at infinity with index $\alpha \in \mathbb{R}$ if it satisfies*

$$\lim_{\lambda \to \infty} \frac{g(\lambda x)}{g(\lambda)} = x^{-\alpha}$$

for all $x > 0$. If $\alpha = 0$, g is said to be slowly varying *at infinity.*

Distributions with Pareto-Lèvy type tails such as Pareto and Stable (with $\alpha < 2$) distributions satisfy

$$\bar{F}(x) := P(X > x) \sim cx^{-\alpha},\ \alpha \in (0, 2)$$

This implies, in particular, that $\bar{F}(x)$ is regularly varying with index α, that is

$$\lim_{\lambda \to \infty} \frac{\bar{F}(\lambda x)}{\bar{F}(\lambda)} = x^{-\alpha}$$

From Mikosch (1999, Remark 1.1.3) it follows that any regularly varying function has representation

$$\bar{F}(x) = cx^{-\alpha} L(x),\ \alpha \in (0, 2) \tag{A.1}$$

for some slowly varying function. Thus, Pareto and Stable distributions are special cases of (A.1) where the slowly varying component $L(x)$ is constant.

An important and related concept is the domain of attraction of a stable law.

Definition A.2 *Consider an i.i.d. sequence $\{\varepsilon_t\}_{t \geq 1}$, each of them with distribution function F. Then, F is said to belong to the domain of attraction of an α-stable distribution S_α if there exist constants $a_T > 0$ and d_T such that*

$$\frac{1}{a_T} \sum\nolimits_{t=1}^{T} (\varepsilon_t - d_T) \to_d S_\alpha$$

as $T \to \infty$. If $m_T = T^{1/\alpha}$, F is said to belong to the 'normal' domain of attraction of an α-stable distribution S_α.

The relation between the property of being in the domain of attraction of a stable law and the behavior of the tails is given in the following Lemma.

LEMMA A.1 (CHAN-TRAN, 1990) *A random variable with distribution function $F(x) = 1 - \bar{F}(x)$ is in the domain of attraction of a stable r.v. with index $\alpha \in (0,2)$* if and only if

$$\bar{F}(x) \sim px^{-\alpha} L(x), \ \alpha \in (0,2), \ as \ x \to \infty$$

where

$$p = \lim_{x \to \infty} \frac{\bar{F}(x)}{\bar{F}(x) + F(-x)} = \lim_{x \to \infty} \frac{P(X > x)}{P(|X| > x)} \in [0,1]$$

In order to characterize the asymptotic behavior of sample mean, variances and autocovariances of heavy tailed, i.i.d. sequence $\{\varepsilon_t\}_{t \geq 1}$ (see the next section), as well as to prove the main theory provided in this paper, it is necessary to define the following two normalizing sequences:

$$a_T := \inf_{x \in \mathbb{R}^+} \left(P(|\varepsilon_t| > x) \leq T^{-1} \right), \ \tilde{a}_T := \inf_{x \in \mathbb{R}^+} \left(P(|\varepsilon_t \varepsilon_{t+1}| > x) \leq T^{-1} \right) \qquad \text{(A.2)}$$

where it also holds that $TP(|\varepsilon_t| > a_T x) \leq x^{-\alpha}$ for all $x > 0$. The relation between the order of magnitudes of a_T and $\tilde{a}_T$ are given next.

LEMMA A.2 (DAVIS AND RESNIK, 1986A) *It holds that*

$$a_T, \tilde{a}_T \to \infty$$

and, in particular, that the sequences $a_T, \tilde{a}_T$ are regularly varying with index α^{-1}. Moreover, as $T \to \infty$,

$$\begin{aligned} a_T &= o(\tilde{a}_T) \\ \tilde{a}_T &= o(a_T^{1+\epsilon}) \ \textit{for any} \ \epsilon > 0 \\ \tilde{a}_T^{-1} a_T^2 &= T^{1/\alpha} L(T) \end{aligned}$$

for L a slowly varying function at infinity. $a_T = T^\alpha L(T)$

Let us turn to the properties of sample second order moments and cross product moments of an heavy tailed sequence. The next theorem holds under the assumptions made in Section 3.2 for the analysis of the asymptotic properties of the bootstrap.

THEOREM A.1 (DAVIS AND RESNICK, 1986A, THEOREM 3.3) *Let $\{\varepsilon_t\}$ satisfy Assumption 1(ii),(iii) of Section 3.2. Then, with $\mu_T := E(\varepsilon_t \varepsilon_{t+1} \mathbb{I}(|\varepsilon_t \varepsilon_{t+1}| \leq \tilde{a}_T)$, under the*

$$\left(\frac{1}{a_T^2} \sum_{t=1}^{T} \varepsilon_t^2, \frac{1}{\tilde{a}_T} \sum_{t=1}^{T-h} (\varepsilon_t \varepsilon_{t+1} - \mu_T), \ldots\ldots, \frac{1}{\tilde{a}_T} \sum_{t=1}^{T-h} (\varepsilon_t \varepsilon_{t+h} - \mu_T) \right) \xrightarrow{w} \qquad \text{(A.3)}$$
$$(\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_h)$$

where $\mathcal{S}_0$ is stable with index $\alpha/2$, $\mathcal{S}_i$ are stable with index α for $i = 1, .., h$ and $\mathcal{S}_j$ are independent for $j = 0, 1, \ldots, h$. The term μ_T can be omitted if Assumption 1(iv) also holds.

The proof of this theorem exploits properties of regular variation and Karamata's theorem (see Feller, 1971 p.283) as detailed in Davis and Resnick (1985, 1986a). In particular, we have

$$T\tilde{a}_T^{-1}E|\varepsilon_t\varepsilon_{t+1}|\mathbb{I}\left(|\varepsilon_t\varepsilon_{t+1}|<\tilde{a}_T\right) \rightarrow \alpha/\left(1-\alpha\right) \tag{A.4}$$

$$T\tilde{a}_T^{-2}E\left(\varepsilon_t^2\varepsilon_{t+1}^2\mathbb{I}\left(|\varepsilon_t^2\varepsilon_{t+1}^2|<\tilde{a}_T\right)\right) = O\left(1\right) \tag{A.5}$$

and, for $\eta<\alpha$,

$$\limsup_{T\to\infty} T\tilde{a}_T^{-\eta}E|\varepsilon_t\varepsilon_{t+1}|^{\eta}\mathbb{I}\left(|\varepsilon_t\varepsilon_{t+1}|\leq\tilde{a}_T\right)<\infty \tag{A.6}$$

and

$$TE\left(|\varepsilon_{t+i}\varepsilon_{t-1}|^{\eta}\,\mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}|>\tilde{a}_T\right)\right)=O\left(\tilde{a}_T^{\eta}\right). \tag{A.7}$$

A.2 On the empirical autocorrelation function

With $\hat{\rho}_T$ denoting the empirical autocorrelation function of order one, or equivalently the LS estimator of ρ in the AR$^+$ (1) model, properties of $\hat{\rho}_T$ have been investigated in Davis and Resnick (1986a). In particular, consistency and asymptotic distributional behaviour require assumptions 1(ii) and (iii), i.e. that the innovations ε_t's are in the domain of attraction of an α-stable distribution. We collect the consistency property in the following lemma, where $\hat{\rho}_T=\sum_{t=1}^{T-1}x_tx_{t+1}/\sum_{t=2}^{T}x_t^2$.

Lemma A.3 (Davis and Resnick, 1986) *Under Assumption 1(i)–(iii), then* $\hat{\rho}_T\rightarrow_p \rho_0$.

To state the asymptotic distribution of $\hat{\rho}_T$, we let $n_T:=\tilde{a}_T^{-1}a_T^2$, with $\tilde{a}_T$ and a_T as given in the previous section. The following lemma holds.

Lemma A.4 (Davis and Resnick, 1986) *Under the assumption of Lemma A.3,*

$$n_T\left(\hat{\rho}_T-\rho_0-d_T\right)\rightarrow_w Z:=\frac{1-\rho_0^2}{\left(1-\rho_0^{\alpha}\right)^{1/\alpha}}\frac{\mathcal{S}_1}{\mathcal{S}_0} \tag{A.8}$$

where $\mathcal{S}_0$ and $\mathcal{S}_1$ are independent stable *random variables with index $\alpha/2$ and α, respectively and $a_T,\tilde{a}_T$ are defined in (A.2). The centering factor d_T satisfies*

$$d_T:=\frac{T}{1-\rho}E\left(\varepsilon_1\varepsilon_2\mathbb{I}(|\varepsilon_1\varepsilon_2|\leq\tilde{a}_T)\right)\left(\sum_{t=1}^{T-1}x_{t+1}^2\right)^{-1} \tag{A.9}$$

and can be omitted if Assumption 1(iv) also holds.

We end this section by presenting the following lemma, which provides the order of magnitude of the studentized sample autocorrelation.

Lemma A.5 *Under Assumption 1,*

$$t_T:=\frac{\hat{\rho}_T-\rho_0}{\hat{\sigma}_T S_{xx}^{-1/2}}=O_P(T^{1/2}n_T^{-1}).$$

B Proofs

In this appendix, we use the following notation. For a given sequence X_T^* computed on the bootstrap data, $X_T^* = o_p^*(1)$ means that for any $\epsilon > 0$, $P^*(||X_T^*|| > \epsilon) \xrightarrow{p} 0$, as $T \to \infty$. Similarly, $X_T^* = O_p^*(1)$ means that, for every $\epsilon > 0$, there exists a constant $M > 0$ such that, for all large T, $P(P^*(||X_T^*|| > M) < \epsilon)$ is arbitrarily close to one. Also, $\mathbb{I}(\cdot)$ denotes the indicator function; $\lfloor \cdot \rfloor$ denotes the integer part of its argument.

Proof of Lemma 1

Consider first the wild bootstrap (ii): Initially note under symmetry that $\varepsilon_t = |\varepsilon_t| w_t$ with w_t an i.i.d. sequence, $P(w_t = \pm 1) = \frac{1}{2}$, and recall that $\varepsilon_t^* = |\varepsilon_t| w_t^*$. Then as,

$$\begin{aligned} \mathcal{D}\left(\varepsilon_1^*, ..., \varepsilon_{T-1}^* | \{x_t\}_{t=1}^T\right) &= \mathcal{D}(\varepsilon_1^*, ..., \varepsilon_{T-1}^* | \{\varepsilon_t\}_{t=1}^{T-1}, x_T) = \mathcal{D}(\varepsilon_1^*, ..., \varepsilon_{T-1}^* | \{|\varepsilon_t|\}_{t=1}^{T-1}, x_T) \\ &= \mathcal{D}(\varepsilon_1, ..., \varepsilon_{T-1} | \{|\varepsilon_t|\}_{t=1}^{T-1}, x_T) \end{aligned}$$

it follows that,

$$\mathcal{D}\left(r_T^* | \{x_t\}_1^T\right) = \mathcal{D}(r_T^* | \{|\varepsilon_t|\}_1^{T-1}, x_T) = \mathcal{D}(r_T | \{|\varepsilon_t|\}_1^{T-1}, x_T).$$

Consider next the permutation bootstrap (i) and observe that

$$\begin{aligned} \mathcal{D}\left(\varepsilon_1^*, ..., \varepsilon_{T-1}^* | \{x_t\}_{t=1}^T\right) &= \mathcal{D}(\varepsilon_1^*, ..., \varepsilon_{T-1}^* | \{\varepsilon_t\}_{t=1}^{T-1}, x_T) \\ &= \mathcal{D}(\varepsilon_{\pi(1)}, ..., \varepsilon_{\pi(T-1)} | \{\varepsilon_t\}_{t=1}^{T-1}, x_T) \\ &= \mathcal{D}(\varepsilon_{\pi(1)}, ..., \varepsilon_{\pi(T-1)} | \left\{\varepsilon_{(t)}\right\}_{t=1}^{T-1}, x_T) \\ &= \mathcal{D}(\varepsilon_{\pi\circ(1)}, ..., \varepsilon_{\pi\circ(T-1)} | \left\{\varepsilon_{(t)}\right\}_{t=1}^{T-1}, x_T) \\ &= \mathcal{D}(\varepsilon_1, ..., \varepsilon_{T-1} | \left\{\varepsilon_{(t)}\right\}_{t=1}^{T-1}, x_T). \end{aligned}$$

where $\pi \circ (t)$ denotes a permutation of the order statistic (t). Next, as for the wild bootstrap,

$$\mathcal{D}\left(r_T^* | \{x_t\}_1^T\right) = \mathcal{D}(r_T^* | \left\{\varepsilon_{(t)}\right\}_1^{T-1}, x_T) = \mathcal{D}(r_T | \left\{\varepsilon_{(t)}\right\}_1^{T-1}, x_T).$$

The results for the permutation-wild bootstrap and for the t_T^* statistics follow similarly. □

Proof of Lemma 2

Consider the difference between $r_T^\dagger$ (based on estimated ρ and unrestricted residuals) and r_T^* (based on true ρ and true errors), where

$$r_T^\dagger = \hat{\rho}_T^\dagger - \hat{\rho}_T = S_{1\varepsilon}^\dagger S_{11}^{\dagger -1},$$

such that by Berk (1974, eq. (2.15))

$$|r_T^\dagger - r_T^*| \le \frac{|S_{11}^{*-1}|^2|S_{11}^\dagger - S_{11}^*|}{1-|S_{11}^{\dagger-1}||S_{11}^\dagger - S_{11}^*|}\left(|S_{1\varepsilon}^*| + |S_{1\varepsilon}^\dagger - S_{1\varepsilon}^*|\right) + |S_{11}^{*-1}||S_{1\varepsilon}^\dagger - S_{1\varepsilon}^*|$$

With $a_T, \tilde{a}_T$ as defined in Section A, the result follows by establishing: *(i)* $|S_{11}^\dagger - S_{11}^*| = o_{P^*}\left(a_T^{1+\varepsilon}\right)$, $a_T^2 S_{11}^{*-1} = O_{P^*}(1)$, *(ii)* $|S_{1\varepsilon}^\dagger - S_{1\varepsilon}^*| = o_{P^*}(\tilde{a}_T)$ and finally *(iii)* $|S_{1\varepsilon}^*| = O_{P^*}(\tilde{a}_T)$. In this case, we find using also that $\tilde{a}_T = o\left(a_T^{1+\varepsilon}\right)$ for all $\varepsilon > 0$,

$$\begin{aligned}|r_T^\dagger - r_T^*| &= \frac{O_{P^*}\left(a_T^{-3+\varepsilon}\right)}{1+o_{P^*}(1)}[O_{P^*}(\tilde{a}_T) + o_{P^*}(\tilde{a}_T)] + o_{P^*}\left(\tilde{a}_T a_T^{-2}\right)\\ &= O_{P^*}\left(a_T^{-3+\varepsilon}\right)O_{P^*}(\tilde{a}_T) + o_{P^*}\left(\tilde{a}_T a_T^{-2}\right)\\ &= O_{P^*}\left(a_T^{-1+\varepsilon}\right) + o_{P^*}\left(\tilde{a}_T a_T^{-2}\right) = o_{P^*}\left(\tilde{a}_T a_T^{-2}\right)\end{aligned}$$

for $0 < \varepsilon < 1$. Hence, as claimed, $n_T(r_T^\dagger - r_T^*) = o_{P^*}(1)$ for $n_T := a_T^2 \tilde{a}_T^{-1}$.We proceed by establishing *(i)*-*(iii)*.

Establishing *(i)* $|S_{11}^\dagger - S_{11}^*| = O_{P^*}\left(a_T^{1+\varepsilon}\right)$, $a_T^2 S_{11}^{*-1} = O_{P^*}(1)$:

We present here the result for the case of the wild bootstrap as this is the most involved case. By definition, $x_t^\dagger = \hat{\rho}_T^{T-t} x_T^\dagger + \sum_{i=0}^{T-t-1} \hat{\rho}_T^i \varepsilon_{t+i}^\dagger$, and hence with $x_T^\star = 0$ without loss of generality we find

$$\begin{aligned}S_{11}^\dagger &= \sum_{t=2}^{T-1} x_t^{\dagger 2} = \sum_{t=2}^{T-1}\left(\sum_{i=0}^{T-t-1}\hat{\rho}_T^i \varepsilon_{t+i}^\dagger\right)^2 = \sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\hat{\rho}_T^{2i}\varepsilon_{t+i}^{\dagger 2} + \sum_{t=2}^{T-1}\sum_{i\neq j, j=0}^{T-t-1}\sum_{i=0}^{T-t-1}\hat{\rho}_T^{i+j}\varepsilon_{t+i}^\dagger\varepsilon_{t+j}^\dagger\\ &= \underbrace{\sum_{m=1}^{T-1}\hat{\varepsilon}_m^2\left(\sum_{i=0}^{m-1}\hat{\rho}_T^{2i}\right)}_{(\text{sq}^\dagger)} + \underbrace{2\sum_{m=1}^{T-1}\sum_{k=1}^{T-1-m}\varepsilon_m^\dagger\varepsilon_{m+k}^\dagger\left(\sum_{i=0}^{m-1}\hat{\rho}_T^{2i+k}\right)}_{(\text{cp}^\dagger)}\end{aligned}$$

Similarly, $x_t^* = \rho^{T-t}x_T^* + \sum_{i=0}^{T-t-1}\rho^i\varepsilon_{t+i}^*$, and hence with $x_T^\star = 0$ without loss of generality we find

$$\begin{aligned}S_{11}^* &= \sum_{t=2}^{T-1} x_t^{*2} = \sum_{t=2}^{T-1}\left(\sum_{i=0}^{T-t-1}\rho_0^i \varepsilon_{t+i}^*\right)^2 = \sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\rho_0^{2i}\varepsilon_{t+i}^{*2} + \sum_{t=2}^{T-1}\sum_{i\neq j, j=0}^{T-t-1}\sum_{i=0}^{T-t-1}\rho_0^{i+j}\varepsilon_{t+i}^*\varepsilon_{t+j}^*\\ &= \underbrace{\sum_{m=1}^{T-1}\varepsilon_m^2\left(\sum_{i=0}^{m-1}\rho_0^{2i}\right)}_{(\text{sq}^*)} + \underbrace{2\sum_{m=1}^{T-1}\sum_{k=1}^{T-1-m}\varepsilon_m^*\varepsilon_{m+k}^*\left(\sum_{i=0}^{m-1}\rho_0^{2i+k}\right)}_{(\text{cp}^*)}.\end{aligned}$$

Here we have used the fact that $\varepsilon_t^{\dagger 2} = \hat{\varepsilon}_t^2 w_t^2 = \hat{\varepsilon}_t^2$ and similarly $\varepsilon_t^{*2} = \varepsilon_t^2$. Consider first the difference sq$^\dagger -$sq*, which equals

$$\sum_{m=1}^{T-1}\varepsilon_m^2\left(\sum_{i=0}^{m-1}(\hat{\rho}_T^{2i} - \rho_0^{2i})\right) + \sum_{m=1}^{T-1}(\hat{\varepsilon}_m^2 - \varepsilon_m^2)\left(\sum_{i=0}^{m-1}\hat{\rho}_T^{2i}\right)$$

where

$$\left|\sum_{m=1}^{T-1}\varepsilon_m^2\left(\sum_{i=0}^{m-1}(\hat{\rho}_T^{2i}-\rho_0^{2i})\right)\right|\leq c_T\left(\sum_{m=1}^{T-1}\varepsilon_m^2\right)|\hat{\rho}_T-\rho_0|=O_P\left(\tilde{a}_T\right)=o_P\left(a_T^{1+\varepsilon}\right),$$

with $c_T = O_P(1)$ by the mean-value theorem and the fact that $|\hat{\rho}_T| < 1$ for T large enough. Similarly,

$$\left|\sum_{m=1}^{T-1}(\hat{\varepsilon}_m^2-\varepsilon_m^2)\left(\sum_{i=0}^{m-1}\hat{\rho}_T^{2i}\right)\right|\leq c\left(\hat{\rho}_T-\rho_0\right)^2\sum_{t=1}^{T-1}x_{t+1}^2=O_p\left(\tilde{a}_T^2a_T^{-2}\right)=o_P\left(a_T^{1+\varepsilon}\right)$$

using the fact that $\hat{\varepsilon}_t-\varepsilon_t=-\left(\hat{\rho}_T-\rho_0\right)x_{t+1}$.

Now we consider the difference of the cross-product terms, $\mathrm{cp}^{\dagger}-\mathrm{cp}^{*}$:

$$\sum_{m=1}^{T-1}\sum_{k=1}^{T-1-m}\varepsilon_m^*\varepsilon_{m+k}^*\left(\sum_{i=0}^{m-1}\hat{\rho}_T^{2i+k}-\rho_0^{2i+k}\right)+\sum_{m=1}^{T-1}\sum_{k=1}^{T-1-m}\left(\varepsilon_m^{\dagger}\varepsilon_{m+k}^{\dagger}-\varepsilon_m^*\varepsilon_{m+k}^*\right)\left(\sum_{i=0}^{m-1}\hat{\rho}_T^{2i+k}\right)=:\xi_{1T}+\xi_{2T}.$$

Rewrite ξ_{1T} as follows,

$$\begin{aligned}\xi_{1T} &= \sum_{m=1}^{T-1}\sum_{n=m+1}^{T-1}\varepsilon_m\varepsilon_n w_n w_m\sum_{i=0}^{m-1}\left(\hat{\rho}_T^{2i+n-m}-\rho_0^{2i+n-m}\right)\\ &= \sum_{m=1}^{T-1}\sum_{n=m+1}^{T-1}\varepsilon_m\varepsilon_n w_n w_m R_{T,m,n}\end{aligned}$$

such that consequently,

$$\begin{aligned}E^*\left(\xi_T^2\right) &= E^*(\sum_{m=1}^{T-1}\sum_{n=m+1}^{T-1}\varepsilon_m\varepsilon_n w_n w_m R_{T,m,n})^2\\ &= \sum_{m=1}^{T-1}\sum_{n=m+1}^{T-1}\sum_{p=1}^{T-1}\sum_{q=p+1}^{T-1}\varepsilon_m\varepsilon_n\varepsilon_p\varepsilon_q E^*\left(w_m w_n w_p w_q\right)R_{T,m,n}R_{T,p,q}\\ &= c\sum_{m=1}^{T-1}\sum_{n=m+1}^{T-1}\varepsilon_m^2\varepsilon_n^2R_{T,m,n}^2\leq c_T\left(\hat{\rho}_T-\rho_0\right)^2(\sum_{m=2}^{T-1}\varepsilon_m^2)^2=O_P\left(\tilde{a}_T^2\right)=o_P(a_T^{2(1+\varepsilon))})\end{aligned}$$

where, $c_T = O_P(1)$ and, for T large enough, $R_{T,m}^2 = O_P((\hat{\rho}_T-\rho_0)^2)$. Regarding ξ_{2T} we have

$$\xi_{2T}=\sum_{m=1}^{T-1}\sum_{k=1}^{T-1-m}\left(\hat{\varepsilon}_m\hat{\varepsilon}_{m+k}-\varepsilon_m\varepsilon_{m+k}\right)w_m w_{m+k}\underbrace{(\sum_{i=0}^{m-1}\hat{\rho}_T^{2i+k})}_{\gamma_{T,m}}$$

and hence

$$E^*(\xi_{2T}^2) = \sum_{m=1}^{T-1}\sum_{k=1}^{T-1-m}\left((\hat{\varepsilon}_m-\varepsilon_m)\hat{\varepsilon}_{m+k}+\varepsilon_m(\hat{\varepsilon}_{m+k}-\varepsilon_{m+k})\right)^2\gamma_{T,m}^2$$

$$
\begin{aligned}
&\leq c(\hat{\rho}_T-\rho_0)^2(\sum_{m=1}^{T-1}\sum_{k=1}^{T-1-m} x_{m+1}^2\hat{\varepsilon}_{m+k}^2+\sum_{m=1}^{T-1}\sum_{k=1}^{T-1-m} x_{m+k+1}^2\varepsilon_m^2) \\
&\leq c(\hat{\rho}_T-\rho_0)^2(\sum_{m=1}^{T-1} x_{m+1}^2)(\sum_{m=1}^{T-1}\hat{\varepsilon}_m^2+\sum_{m=1}^{T-1}\varepsilon_m^2) \\
&= O_p(\tilde{a}_T^2 a_T^{-4})O_p(a_T^4)=O_p(\tilde{a}_T^2)=o_p(a_T^{2(1+\varepsilon)}).
\end{aligned}
$$

Collecting terms, we find $S_{11}^{\dagger}-S_{11}^{*}=o_{P^*}\left(a_T^{1+\varepsilon}\right)$, as desired.

Establishing *(ii)* $\left|S_{1\varepsilon}^{\dagger}-S_{1\varepsilon}^{*}\right|=o_{P^*}(\tilde{a}_T)$:

By definition, using $x_T^{\dagger}=0$,

$$
\begin{aligned}
S_{1\varepsilon}^{*} &= \sum_{t=1}^{T-1} x_{t+1}^{*}\varepsilon_t^{*}=\sum_{t=2}^{T-1} x_t^{*}\varepsilon_{t-1}^{*}=\sum_{t=2}^{T-1}\left(\sum_{i=0}^{T-t-1}\rho_0^i\varepsilon_{t+i}^{*}\right)\varepsilon_{t-1}^{*} \\
&= \sum_{t=2}^{T-1}\varepsilon_{t-1}w_{t-1}\left(\sum_{i=0}^{T-t-1}\rho_0^i\varepsilon_{t+i}w_{t+i}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
S_{1\varepsilon}^{\dagger} &= \sum_{t=1}^{T-1} x_{t+1}^{\dagger}\varepsilon_t^{\dagger}=\sum_{t=2}^{T-1} x_t^{\dagger}\varepsilon_{t-1}^{\dagger}=\sum_{t=2}^{T-1}\left(\sum_{i=0}^{T-t-1}\hat{\rho}_T^i\varepsilon_{t+i}^{\dagger}\right)\varepsilon_{t-1}^{\dagger} \\
&= \sum_{t=2}^{T-1}\hat{\varepsilon}_{t-1}w_{t-1}\left(\sum_{i=0}^{T-t-1}\hat{\rho}_T^i\hat{\varepsilon}_{t+i}w_{t+i}\right)
\end{aligned}
$$

Hence,

$$
S_{1\varepsilon}^{\dagger}-S_{1\varepsilon}^{*}=\sum_{t=1}^{T-1}x_{t+1}^{\dagger}\varepsilon_t^{\dagger}-\sum_{t=1}^{T-1}x_{t+1}^{*}\varepsilon_t^{*}=\underbrace{\sum_{t=1}^{T-1}\left(x_{t+1}^{\dagger}-x_{t+1}^{*}\right)\varepsilon_t^{*}}_{A_T}+\underbrace{\sum_{t=1}^{T-1}x_{t+1}^{\dagger}\left(\varepsilon_t^{\dagger}-\varepsilon_t^{*}\right)}_{B_T}
$$

Regarding A_T, notice that

$$
A_T=\sum_{t=1}^{T-1}\left(x_{t+1}^{\dagger}-x_{t+1}^{\#}\right)\varepsilon_t^{*}+\sum_{t=1}^{T-1}\left(x_{t+1}^{\#}-x_{t+1}^{*}\right)\varepsilon_t^{*}=:A_{1T}+A_{2T}
$$

with $\{x_t^{\#}\}$ being a bootstrap sample based on ρ_0 and resample from $\hat{\varepsilon}_t$. First,

$$
\begin{aligned}
E^*\left(A_{1T}\right)^2 &= E^*(\sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\left(\hat{\rho}_T^i-\rho_0^i\right)\hat{\varepsilon}_{t+i}w_{t+i}\varepsilon_{t-1}w_{t-1})^2 \\
&= \sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\sum_{m=2}^{T-1}\sum_{j=0}^{T-t-1}(\hat{\rho}_T^i-\rho_0^i)(\hat{\rho}_T^j-\rho_0^j)\hat{\varepsilon}_{t+i}\varepsilon_{t-1}\hat{\varepsilon}_{m+j}\varepsilon_{m-1}E^*\left(w_{m+j}w_{t+i}w_{m-1}w_{t-1}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\left(\hat{\rho}_T^i-\rho_0^i\right)^2\hat{\varepsilon}_{t+i}^2\varepsilon_{t-1}^2 \le c\left(\hat{\rho}_T-\rho_0\right)^2\sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\gamma^{2i}\hat{\varepsilon}_{t+i}^2\varepsilon_{t-1}^2 \\
&= c\left(\hat{\rho}_T-\rho_0\right)^2 Q_T = o_p\left(\tilde{a}_T^2\right)
\end{aligned}
$$

where for the last inequality we have used that, for T large enough and for some positive constants $\gamma < 1$ and generic c,

$$
|\hat{\rho}_T^i-\rho_0^i| = |\hat{\rho}_T-\rho_0|\left|\sum_{k=1}^{i}\hat{\rho}_T^{i-k}\rho_0^{k-1}\right| \le c|\hat{\rho}_T-\rho_0|\gamma^i\ . \qquad \text{(B.10)}
$$

The order of magnitude $o_p\left(\tilde{a}_T^2\right)$ follows by observing first,

$$
\begin{aligned}
Q_T &\le c\sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\gamma^{2i}\varepsilon_{t+i}^2\varepsilon_{t-1}^2 + c\sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\gamma^{2i}(\hat{\varepsilon}_{t+i}-\varepsilon_{t+i})^2\varepsilon_{t-1}^2 \\
&\le c\sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\gamma^{2i}\varepsilon_{t+i}^2\varepsilon_{t-1}^2 + c\left(\hat{\rho}_T-\rho_0\right)^2\sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\gamma^{2i}x_{t+1+i}^2\varepsilon_{t-1}^2 \\
&= Q_{1T} + c\left(\hat{\rho}_T-\rho_0\right)^2 Q_{2T},
\end{aligned}
$$

with Q_{1T} and Q_{2T} implicitly defined. Decompose Q_{1T} as follows,

$$
Q_{1T} = \sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\gamma^{2i}\varepsilon_{t+i}^2\varepsilon_{t-1}^2\left(\mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| \le \tilde{a}_T\right) + \mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| > \tilde{a}_T\right)\right),
$$

and note that

$$
\begin{aligned}
E\left(\sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\gamma^{2i}\varepsilon_{t+i}^2\varepsilon_{t-1}^2\mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| \le \tilde{a}_T\right)\right) &= \sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\gamma^{2i}E\left(\varepsilon_{t+i}^2\varepsilon_{t-1}^2\mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| \le \tilde{a}_T\right)\right) \\
&\le cTE\left(\varepsilon_t^2\varepsilon_{t-1}^2\mathbb{I}\left(|\varepsilon_t\varepsilon_{t-1}| \le \tilde{a}_T\right)\right) = O\left(\tilde{a}_T^2\right)
\end{aligned}
$$

by Karamata's Theorem, see (A.5). Likewise, for some $\eta \in (0,\alpha)$ such that $\eta/2 < 1$, and using again Karamata's Theorem, see (A.7),

$$
E\Big(\sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\gamma^{2i}\varepsilon_{t+i}^2\varepsilon_{t-1}^2\mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| > \tilde{a}_T\right)\Big)^{\frac{\eta}{2}} \le cTE\left(|\varepsilon_{t+i}\varepsilon_{t-1}|^{\eta}\,\mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| > \tilde{a}_T\right)\right) = O\left(\tilde{a}_T^{\eta}\right).
$$

Collecting terms, $Q_{1T} = O_p\left(\tilde{a}_T^2\right)$. Likewise $Q_{2T} = \sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\gamma^{2i}x_{t+i}^2\varepsilon_{t-1}^2 = O_p\left(\tilde{a}_T^2\right)$, and hence

$$
\left(\hat{\rho}_T-\rho\right)^2 Q_T = o_p(1)\left(O_p\left(\tilde{a}_T^2\right) + o_p(1)\,O_p\left(\tilde{a}_T^2\right)\right) = o_p\left(\tilde{a}_T^2\right).
$$

Next, consider $B_T = \sum_{t=1}^{T-1}x_{t+1}^{\dagger}\left(\varepsilon_t^{\dagger}-\varepsilon_t^{*}\right)$:

Remark B.1 For our main result to hold we require B_T to be of order $o_P^*(\tilde{a}_T)$ in P-probability. However, unless the bootstrap shocks involve the permutation (i.e. $\{\pi(t)\}_1^{T-1}$ is a uniformly distributed random permutation of $\{1,...T-1\}$), such rate cannot be achieved. To see this, notice that without permutation

$$\varepsilon_t^{\dagger} - \varepsilon_t^* = -(\hat{\rho}_T - \rho_0)\, x_{t+1} w_t$$

such that $B_T = -(\hat{\rho}_T - \rho_0)\sum_{t=1}^{T-1}\left(\sum_{i=0}^{T-t-1}\hat{\rho}_T^i \hat{\varepsilon}_{t+1+i} w_{t+1+i}\right) x_{t+1} w_t$. Hence,

$$E^*\left(B_T^2\right) = (\hat{\rho}_T - \rho_0)^2 \sum_{t=1}^{T-1}\sum_{i=0}^{T-t-1}\hat{\rho}_T^{2i}\hat{\varepsilon}_{t+1+i}^2 x_{t+1}^2 = O_p\left(\frac{\tilde{a}_T^2}{a_T^4}\right) O\left(a_T^4\right) = O_p\left(\tilde{a}_T^2\right)$$

and B_T is of order $O_p^*(\tilde{a}_T)$ rather than $o_p^*(\tilde{a}_T)$.

We prove that the required rate is obtained for the permutation-wild bootstrap. In this case, $x_t^{\dagger}$ and x_t^* are generated with bootstrap shocks defined as $\varepsilon_t^{\dagger} = \hat{\varepsilon}_{\pi(t)} w_t$ and $\varepsilon_t^* = \varepsilon_{\pi(t)} w_t$, respectively. For this choice we find

$$B_T = \sum_{t=1}^{T-1}\left(\sum_{i=0}^{T-t-1}\hat{\rho}_T^i \hat{\varepsilon}_{\pi(t+1+i)} w_{t+1+i}\right)\left(\hat{\varepsilon}_{\pi(t)} - \varepsilon_{\pi(t)}\right) w_t$$

and

$$E^*\left(B_T^2\right) = E^*(\hat{\varepsilon}_{\pi(t+1+i)}^2\left(\hat{\varepsilon}_{\pi(t)} - \varepsilon_{\pi(t)}\right)^2)\sum_{t=1}^{T-1}\sum_{i=0}^{T-t-1}\hat{\rho}_T^{2i}$$

where the double summation term is of order T. Next, using standard properties of expectations under random permutations,

$$\begin{aligned} E^*(\hat{\varepsilon}_{\pi(t+1+i)}^2\left(\hat{\varepsilon}_{\pi(t)} - \varepsilon_{\pi(t)}\right)^2) &\leq \frac{1}{T^2}\sum_{t=1}^{T-1}\hat{\varepsilon}_t^2 \sum_{t=1}^{T}(\hat{\varepsilon}_t - \varepsilon_t)^2 \\ &= O_p\left(\frac{a_T^2}{T^2}\right)(\hat{\rho} - \rho_0)^2 \sum_{t=1}^{T} x_{t+1}^2 = O_p\left(\frac{a_T^2}{T^2}\right) O_p\left(\frac{\tilde{a}_T^2}{a_T^4}\right) O_p\left(a_T^2\right) \\ &= O_p\left(T^{-2}\tilde{a}_T^2\right) \end{aligned}$$

Hence, in P-probability, $B_T = O_p^*\left(T^{-1/2}\tilde{a}_T\right) = o_p^*(\tilde{a}_T)$.

Establishing *(iii)* $|S_{1\varepsilon}^*| = O_{P^*}(\tilde{a}_T)$:

We omit without loss of generality the permutation and consider instead the wild bootstrap which is the most involved case. By definition of the wild bootstrap,

$$S_{1\varepsilon}^* = \sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1}\rho_0^i \varepsilon_{t+i}\varepsilon_{t-1} w_{t-1} w_{t+i}$$

such that

$$E^*\left(S_{1\varepsilon}^{*2}\right) = \sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1} \rho_0^{2i}\varepsilon_{t+i}^2\varepsilon_{t-1}^2\left(\mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| \leq \tilde{a}_T\right) + \mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| > \tilde{a}_T\right)\right)$$

Next, as under *(ii)* using Karamata's Theorem (A.5),

$$E(\sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1} \rho_0^{2i}\varepsilon_{t+i}^2\varepsilon_{t-1}^2\mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| \leq \tilde{a}_T\right)) \leq cTE\left(\varepsilon_t^2\varepsilon_{t-1}^2\mathbb{I}\left(|\varepsilon_t\varepsilon_{t-1}| \leq \tilde{a}_T\right)\right) = O\left(\tilde{a}_T^2\right).$$

Likewise for some $\eta \in (0, \alpha)$ such that $\eta/2 < 1$, and using Karamata's Theorem (A.7) again,

$$\begin{aligned} E(\sum_{t=2}^{T-1}\sum_{i=0}^{T-t-1} \rho_0^{2i}\varepsilon_{t+i}^2\varepsilon_{t-1}^2\mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| > \tilde{a}_T\right))^{\frac{\eta}{2}} &\leq \\ cTE\left(|\varepsilon_{t+i}\varepsilon_{t-1}|^\eta\, \mathbb{I}\left(|\varepsilon_{t+i}\varepsilon_{t-1}| > \tilde{a}_T\right)\right)\sum_{i=0}^{\infty}\left|\rho^i\right|^\eta &= O\left(\tilde{a}_T^\eta\right) \end{aligned}$$

Collecting terms, $E^*\left(S_{1\varepsilon}^{*2}\right) = O_P\left(\tilde{a}_T^2\right)$ as desired. □

Proof of Theorem 1

In order to prove the theorem, we need to introduce an *infeasible restricted* bootstrap, based on the true parameter estimates. For this bootstrap, the bootstrap sample is generated recursively as

$$x_t^\star = \beta_0' z_{t+1}^* + \varepsilon_t^\star,\ t = 1, ..., T-k,$$

again initialized at $x_t^\star = x_t$, $t = T-k+1, ..., T$. The bootstrap shocks are generated as $\varepsilon_t^\star := \varepsilon_{\pi^*(t)}w_t$, where as before $\pi^*(t)$, $t = 1, ..., T-k$ is a uniformly distributed random permutation of $\{1, ..., T-k\}$, while w_t is either equal to 1 (wild bootstrap) or i.i.d. Rademacher (permutation wild bootstrap). The bootstrap statistic (based on unrestricted parameter estimation) is then given by

$$r_T^\star = R'\hat{\beta}_T^\star - r$$

where $\hat{\beta}_T^\star$ is the LS estimator of β computed on $\{x_t^\star\}$. This bootstrap satisfies the following lemma (its proof mimics the one given for Lemma 1 and is therefore omitted).

Lemma B.6 (Restricted bootstrap) *With $r_T^\star$ the infeasible restricted bootstrap, and r_T the original statistic, it follows that for the permutation, permutation-wild and wild bootstrap schemes the following holds:*

(i) For the permutation bootstrap,

$$\mathcal{D}(r_T^\star|\{x_t\}_1^T) = \mathcal{D}(r_T|\{\varepsilon_{(t)}\}_1^{T-1}, x_T)$$

where $\{\varepsilon_{(t)}\}_1^{T-1}$ *denotes the order statistics of* $\{\varepsilon_t\}_1^{T-1}$*;*

(ii) For the wild bootstrap, if the distribution of ε_t *is symmetric,*

$$\mathcal{D}(r_T^\star|\{x_t\}_1^T) = \mathcal{D}(r_T|\{|\varepsilon_t|\}_1^{T-1}, x_T) \ ;$$

(iii) For the combined permutation-wild bootstrap, if the distribution of ε_t *is symmetric,*

$$\mathcal{D}(r_T^\star|\{x_t\}_1^T) = \mathcal{D}(r_T|\left\{|\varepsilon_{(t)}|\right\}_1^{T-1}, x_T) \ .$$

The same results hold with $r_T, r_T^\star$ *replaced by* $t_T, t_T^\star$*, respectively.*

Given the results in Lemma B.6, we now proceed by showing that in P-probability

$$n_T(r_T^* - r_T^\star) = o_p^*(1) \tag{B.11}$$

where $n_T := \tilde{a}_T^{-1}a_T^2$. This implies the desired result that the (normalized) bootstrap statistics, conditionally on the data, mimic the distribution of the (normalized) original statistic, conditional on suitable statistics.

Without loss of generality, we set $z_{T-k+1} = 0$. As in the proof of Lemma 2 it suffices to establish that: *(i)* $||S_{11}^\star - S_{11}^*|| = o_{P^*}\left(a_T^{1+\varepsilon}\right)$, $a_T^2||S_{11}^{*-1}|| = O_{P^*}(1)$, *(ii)* $||S_{1\varepsilon}^\star - S_{1\varepsilon}^*|| = o_{P^*}(\tilde{a}_T)$ and finally *(iii)* $||S_{1\varepsilon}^\star|| = O_{P^*}(\tilde{a}_T)$. In order to do so, one can follow exactly the same steps as done there, where the results in the following lemmas are now required (their proof is omitted for the sake of brevity).

LEMMA B.7 *Under the assumptions of Theorem 1, the restricted and unrestricted LS estimators* $\tilde{\beta}_T$ *and* $\hat{\beta}_T$ *satisfy, under the null hypothesis,*

$$||\tilde{\beta}_T - \beta_0|| = O_p(||\hat{\beta}_T - \beta_0||) = O_p\left(\tilde{a}_T a_T^{-2}\right) = o_P(1)$$

Moreover, with $\tilde{\varepsilon}_t - \varepsilon_t = -(\tilde{\beta}_T - \beta_0)z_{t+1}$*,*

$$\sum |\tilde{\varepsilon}_t^2 - \varepsilon_t^2| = o_p\left(\tilde{a}_T^2\right)$$

LEMMA B.8 *The bootstrap process* x_t^* *satisfies*

$$x_t^* = \sum_{i=0}^{t-k-1} \tilde{\theta}_{T,i}\varepsilon_{t+i}^* \tag{B.12}$$

where $|\tilde{\theta}_{T,i}| \le c_T\rho^i$ *for some* $\rho \in [0,1)$ *and* $c_T = O_P(1)$*. The unfeasible bootstrap process* $x_t^\star$ *satisfies (B.12) with* ε_{t+i}^* *replaced by* $\varepsilon_{t+i}^\star$ *and* $\tilde{\theta}_{T,i}$ *replaced by* θ_i*, with* $|\theta_i| \le c\rho^i$*.*

FURTHER PROOFS

PROOF OF LEMMA A.5. Recall that $\hat{\rho}_T - \rho_0 = O_P(\tilde{a}_T a_T^{-2})$, and that $S_{xx}^{-1/2} = O_P(a_T^{-1})$. Next,

$$\hat{\sigma}_T^2 = T^{-1}\left(S_{\varepsilon\varepsilon} - S_{\varepsilon x}^2 S_{xx}^{-1}\right),$$

where $S_{\varepsilon\varepsilon} = O_P\left(a_T^2\right)$ and $S_{\varepsilon x} = O_P\left(\tilde{a}_T\right)$, and consequently $\hat{\sigma}_T^2 = O_P\left(T^{-1}\left(a_T^2 + \tilde{a}_T^2 a_T^{-2}\right)\right) = O_P\left(T^{-1}a_T^2\right)$. Collecting terms the result holds. □