Testing the Fractionally Integrated Hypothesis using M Estimation

Matei Demetrescu[a], Paulo M.M. Rodrigues[b] and Antonio Rubia[c]

[a] Christian-Albrechts-University of Kiel

[b] Banco de Portugal and Nova School of Business and Economics, Universidade Nova de Lisboa

[c] University of Alicante

Work in Progress. This version: December 8, 2017

Abstract

This paper develops a class of tests for fractional integration in the time domain based on M estimation. This testing strategy offers more robust properties against non-Gaussian errors than least squares (LS) or other estimation principles. We discuss the asymptotic properties of the new tests under fairly general assumptions, and for different estimation approaches based on direct optimization of the M loss-function and on iterated k-step and reweighted LS numeric algorithms. Monte Carlo experimentation shows that M tests for fractional integration exhibit empirical size close to the nominal level in finite samples and enhanced power in relation to alternative procedures, such as LS and even quantile regression, when innovations are drawn from heavy-tailed distributions. An application to daily volatility of several stock market indices, proxied by log absolute returns and log high-low price ranges, shows the empirical performance of the new tests.

Keywords: Fractional integration, M estimation, heavy-tails, long memory, volatility.
JEL: C12, C22.

1 Introduction

Most unit-root and cointegration tests build on least squares (LS) estimation. This principle ensures efficiency under Gaussian conditions. In practice, however, macroeconomic and financial variables are usually driven by heavy-tailed distributions. In this context, LS-based tests remain asymptotically valid under appropriate conditions, but are no longer efficient. Suitable alternatives, which either accommodate the true likelihood of the data or ensure robust properties against deviations from normality, can exhibit improved power. This consideration becomes particularly relevant in stochastic-trend detection and persistence analysis because the relative losses in power from using inefficient tests tend to be greater under nonstationarity; see, among others, Rothenberg and Stock (1997) and Georgiev, Rodrigues and Taylor (2017).

This concern is extensible to fractional integrated models (Beran, 1994; Haldrup and Nielsen, 2003; Tolvi, 2003), which generalize the unit-root setting and provide a convenient form to describe long-range dependence; see Baillie (1996) and Robinson (2003) for reviews. While the unit root literature has suggested alternative procedures to deal with non-Gaussian errors (see, among others, Lucas, 1993; Campbell and Dufour, 1995; Phillips, 1995; Herce, 1996; Breitung and Gourieroux, 1997; Rothenberg and Stock, 1997; Hasan and Koenker, 1997; Wright 2000; Ling and McAleer, 2004; Koenker and Xiao 2004; and Galvao, 2009), there have been little attempts to develop robust tests for fractional integration. To the best of our knowledge, only the tests in Delgado and Velasco (2005), based on signed residuals, and Hassler, Rodrigues and Rubia (2016), based on quantile regression (QR), ensure a form of robustness in the estimation of the fractional parameter in a general context characterized by either stationary or nonstationary dynamics with errors drawn from a heavy-tailed distribution.[1]

In this paper, we propose a class of tests for fractional integration in the time domain under M estimation; see Huber (1981) and Amemiya (1985) for a review of this topic. This framework is fairly general and encompasses different estimation techniques includ-

[1] Beran (1994) proposes an approximated maximum likelihood estimator based on the autoregressive representation of a stationary ARFIMA model which belongs to a class of M-estimators; see also Agostinelli and Bisaglia (2010). Li and Li (2008) discuss the asymptotic properties of LAD estimators of stationary ARFIMA models in a Laplace quasi-maximum likelihood estimation setting.

ing, among others, LS, QR and maximum likelihood (ML) estimation. The main interest is on the class of weighting functions that reduce the influence of large observations and, therefore, lead to robust properties against heavy-tailed distributions. This analysis extends the tests for fractional integration in Breitung and Hassler (2002), Demetrescu, Kuzin and Hassler (2008) and Hassler *et al.* (2009) to a non-gaussian framework. We discuss the asymptotic properties of the new tests under fairly general conditions on the data generating process, and for different estimation strategies based on direct optimization of the M loss-function and on iterated k-step and reweighted LS numeric algorithms. As in the LS context, the null asymptotic distribution of these tests is shown to be standard normal and independent of the value of the fractional integration parameter. Monte Carlo experimentation shows that M tests for fractional integration exhibit empirical size close to the nominal level in finite samples and enhanced power in relation to alternative procedures, such as LS and even QR, when innovations are drawn from heavy-tailed distributions.

The empirical section addresses the long-run dynamics of different volatility measures of stock market indices in developed and emerging markets. In particular, we consider daily observations of log absolute returns and log price range estimates (the log-transform of the spread between the highest and lowest log-asset prices over the day) in the period 2000-2016. Absolute returns and related transformations are well-known proxies of volatility. Similarly, the high-low range is a highly efficient proxy of volatility as it builds on the entire intraday price path rather than on closing prices; see, among others, Alizadeh, Brandt and Diebold (2002) and references therein. Both measures exhibit the distinctive pattern of long-range persistence in the autocorrelation function that characterizes fractional integration, but they are drawn from statistical distributions with different properties. Whereas log absolute returns are highly non-normal, the distribution of log-range estimates is approximately Gaussian; see Alizadeh *et al.* (2002). Consequently, the use of different estimation techniques with complementary properties in terms of efficiency/robustness given the distribution of the data seems naturally motivated in this context.

Given these series, we provide formal insight on the long-term dynamics of volatility by constructing confidence intervals of the fractional parameter using a large battery of alternative estimation principles. These include regression-based tests for long-memory

based on LS, QR and M estimation, the sign test in Delgado and Velasco (2005), and the frequency-domain local Whittle estimator in Shimotsu and Phillips (2005). Consistent with previous evidence, the overall results from this analysis pinpoint that market return volatility is driven by fractional integration. The analysis of log-range volatility estimates systematically indicates a stronger degree of persistence than that based on log absolute returns. Because the former is widely considered as a more efficient volatility proxy (see, *e.g.*, Andersen and Bollerslev 1998; Alizadeh *et al.* 2002; Brandt and Diebold 2006), the related evidence is more reliable. In this analysis, M estimation produces estimates which are not markedly different from those based on LS, but which nevertheless tend to exhibit smaller amplitude (*i.e.*, smaller parameter uncertainty). This evidence, which completely agrees with the experimental results reported in the Monte Carlo section, suggests that M-based testing could improve the empirical efficiency of LS-based inference, with relative gains that depend on the extent of non-normality of the data. In practice, these gains come at the expense of little incremental computational cost because iterated algorithms used in M estimation do not require numeric optimization or complex operations. M-based inference, therefore, represents a valuable alternative (or complement) since it provides significant refinements over LS even in a quasi-Gaussian context.

The remainder of the paper is organized as follows. Section 2 outlines the general fractional integration context analyzed in the paper. Section 3 details the asymptotic behavior of the test statistics. Section 4 reports Monte Carlo simulation results on the small-sample performance of the tests. Section 5 discusses the empirical application. Finally, Section 6 summarizes and concludes. All proofs of the main theoretical statements are collected in a technical appendix.

In what follows, '$\Rightarrow$' and '$\xrightarrow{p}$' denote weak convergence and convergence in probability, respectively, of a sequence of random elements when the sample length is allowed to diverge. The terms $o_p(1)$ and $O_p(1)$ represent a sequence of random numbers converging to zero in probability and bounded in probability, respectively.

2 Testing for fractional integration

2.1 The data generating process

Assume that the observable time series, $\{y_t\}_{t=1}^{T}$, is generated as,

$$(1-L)^{d+\theta} y_t = \varepsilon_t I(t>1) \tag{1}$$

where L denotes the lag operator, $d+\theta$ is a real value usually referred to as the long-memory or fractional integration parameter, $I(\cdot)$ denotes the indicator function, and $\{\varepsilon_t\}$ is a covariance stationary and invertible noise process.

According to this specification, $\{y_t\}$ in (1) is generally said to be a (Type-II) fractionally integrated process of order $d+\theta$, which shall be referred to as $FI(d+\theta)$ in the sequel. The $\{y_t\}$ variable is driven by a unit root process when $d+\theta=1$, and by a weakly-stationary process when $d+\theta=0$. Non-integer values in the range $(0,1)$ give rise to long-range dependence characterized by hyperbolically-decaying impulse response functions (Hassler and Kokoska, 2010), offering an intermediate case between the characteristic exponential decay of short memory and the infinite persistence of unit-root processes. In contrast to most of the existing tests, model (1) does not require that $d+\theta$ lies in the $(-0,5, 0.5)$ interval in which $\{y_t\}$ is stationary and invertible. This outstanding property provides a considerable degree of generality in our analysis.

For a real-valued d, our main purpose is to address whether $\{y_t\}$ is $FI(d)$ or, equivalently, to test the null hypothesis $\mathrm{H}_0 : \theta = 0$. This hypothesis is tested against the two-sided alternative $\mathrm{H}_1 : \theta \neq 0$, noting that one-sided alternatives are also a straightforward possibility in this context. The following assumption lays out the properties of $\{\varepsilon_t\}$ in (1) and completes the characterization of the data generating process (DGP) considered in this paper.

Assumption 1. *The error process $\{\varepsilon_t\}$ in (1) is characterized as $\mathcal{A}(L)\varepsilon_t = e_t$, with $\mathcal{A}(L) = 1 - \sum_{j=1}^{p} a_j L^j$ having all roots outside the unit root circle, and $\{e_t, \mathcal{F}_t\}$ is a strictly stationary and ergodic Martingale Difference Sequence (MDS) with $E(e_t|\mathcal{F}_{t-1}) = 0$, $E(e_t^2|\mathcal{F}_{t-1}) = \sigma^2$, and $E(|e_t|^{2+\epsilon}) < \infty$ for some $\epsilon > 2/3$, where $\mathcal{F}_t = \{e_s : s \leq t\}$ denotes the σ-field generated by $\{e_s, s \leq t\}$.*

Remark 1. Assumption 1 allows for AR(p) short-run dynamics. When the innovations $\{e_t\}$ are i.i.d., (1) is often referred to as an ARFIMA$(p, d+\theta, 0)$ process. For large values of p, this can be seen as a truncated approximation of an ARFIMA$(p^*, d+\theta, q^*)$ model with finite $p^* \geq 0$ and $q^* > 0$. Assumption 1 allows $\{e_t\}$ to be a conditionally homoskedastic MDS, a generalization of the i.i.d. setting usually considered in the fractional integration literature; see, e.g., Hualde and Robinson (2007). In this literature it is also customary to assume L_4-bounded innovations, but a natural motivation for M estimators is that they can downweigh the influence of innovations drawn from a heavy-tailed distribution for which lower-order moments may not exist. Assumption 1 only requires innovations to be $L_{2+2/3}$-bounded.

2.2 Regression-based tests for fractional integration

Given the $\{y_t\}$ series and the real value d, define the stochastic process

$$\varepsilon_{t,d} := (1-L)_+^d \, y_t = \sum_{j=0}^{t-1} \lambda_j(d) \, y_{t-j}, \tag{2}$$

and

$$x^*_{t-1,d} := \sum_{j=1}^{t-1} j^{-1} \varepsilon_{t-j,d}, \; t = 2, ..., T \tag{3}$$

with $\{\lambda_j(d)\}_{j=0}^{t-1}$ characterized by the truncated series of polynomial coefficients in the binomial expansion $(1-L)^d := \sum_{j=0}^{\infty} \lambda_j(d) \, L^j$, namely,

$$\lambda_0(d) := 1, \text{ and } \lambda_j(d) := \frac{j-1-d}{j} \lambda_{j-1}(d) \; , \; j \geq 1. \tag{4}$$

As discussed in Breitung and Hassler (2002) and Demeterscu et al. (2008), testing the null hypothesis that d is the order of integration of $\{y_t\}$ in (1), or $\mathrm{H}_0 : \theta = 0$, is equivalent to testing $\mathrm{H}_0 : \phi = 0$ in the LS auxiliary regression

$$\varepsilon_{t,d} = \phi x^*_{t-1,d} + \sum_{j=1}^{p} \pi_j \varepsilon_{t-j,d} + v_t, \tag{5}$$

because this characterization holds exactly with $\phi := 0$, $\pi_j := a_j$, and $v_t := e_t$ under $\mathrm{H}_0 : \theta = 0$ and Assumption 1. This is the result of the application of the Lagrange Multiplier (LM) principle; see also Tanaka (1999). Under local alternatives of the form $\mathrm{H}_1 : \theta = c/\sqrt{T}$ with a fixed $c \neq 0$, it can be shown that $\phi = c/\sqrt{T} + O(T^{-1})$ and that the regression disturbances $\{v_t\}$ have a fractionally integrated noise component. As

a result, the heterogenous behavior of ϕ and the different stochastic properties of the random disturbance provide a sound statistical basis to identify the order of fractional integration in $\{y_t\}$. Despite the apparent theoretical simplicity of this framework, the fact that $x^*_{t-1,d}$ converges in mean square sense to $x^{**}_{t-1,d} := \sum_{j=1}^{\infty} j^{-1}\varepsilon_{t-j,d}$ under the null hypothesis and Assumption 1, with $\left\{x^{**}_{t-1,d}\right\}$ being a stationary linear process with non-absolutely summable coefficients, is a source of major technical difficulties for the asymptotic analysis in this context.

Demetrescu *et al.* (2008) and Hassler *et al.* (2009) derive the asymptotic theory of the fractional integration tests under LS estimation of the set of parameters $\boldsymbol{\kappa} := (\phi, \pi_1, ..., \pi_p)'$, showing $\sqrt{T}$-consistency and asymptotic normality under fairly general conditions. As a result, $\mathrm{H}_0 : \phi = 0$ can be tested by means of a standard t-ratio, or measurable transformations such as its squares. If Assumption 1 is strengthened to require $\varepsilon_t \sim iid\mathcal{N}(0, \sigma^2)$, the specific harmonic weighting upon which $\left\{x^*_{t-1,d}\right\}$ is constructed in (3) ensures efficient testing, and the squared t-statistic for $\mathrm{H}_0 : \phi = 0$ is asymptotically equivalent to the LM test for $\mathrm{H}_0 : \theta = 0$ under ML; see also Robinson (1994) and Tanaka (1999).

If Assumption 1 holds with non-Gaussian innovations, the regression-based approach still ensures asymptotically correct nominal size, independently of the underlying distribution of $\{e_t\}$. However, LS estimation is no longer efficient, and alternative procedures may render more powerful tests. Hassler *et al.* (2016) discuss the asymptotic theory for QR estimators, showing that median-based tests can largely outperform LS-based inference when innovations are driven by heavy-tailed distributions. Similarly, LS is known to be highly sensitive to extreme values, which may lead to parameter bias and wrong inference even in large samples. These considerations provide a natural motivation for developing robust tests based on alternative estimation principles, such as M estimation.

3 M estimation

3.1 Theoretical setup

We now proceed to the asymptotic analysis of M-based tests for fractional integration. To this end, consider the following auxiliary regression:

$$\varepsilon_{t,d} = \alpha + \phi x^*_{t-1,d} + \sum_{j=1}^{p} \pi_j \varepsilon_{t-j,d} + u_t,\ t = p+1, ..., T, \tag{6}$$

or, in vector notation,

$$\varepsilon_{t,d} = \boldsymbol{\beta}' \mathbf{x}^*_{t-1,d} + u_t \tag{7}$$

with $\boldsymbol{\beta} := (\alpha, \phi, \pi_1, ..., \pi_p)'$, $\mathbf{x}^*_{t-1,d} := (1, x^*_{t-1,d}, \varepsilon_{t-1,d}, ..., \varepsilon_{t-p,d})'$ and $\{u_t\}$ denoting a random disturbance. Although the similitudes with (5) are straightforward, there are meaningful differences. In the LS context, the restriction $\alpha = 0$ follows directly from model (1) and, hence, there is no need for an intercept. In contrast, in M regressions it is necessary to include this additional term because it can generally be different from zero. Specifically, its the theoretical value, denoted α_ρ, is determined by sample-dependent features of the data and the choice of the $\rho(\cdot)$ function that characterizes M optimization.[2] Consequently, under the null hypothesis, it follows readily that $u_t := e_t - \alpha_\rho$, and hence $\text{Var}(u_t) = \sigma^2$.

The M estimator of $\boldsymbol{\beta}$ in (6), denoted $\widehat{\boldsymbol{\beta}}_M$, can generally be defined as the solution of the optimization problem:

$$\min_{\boldsymbol{\beta} \in \Theta} Q^*_T(\boldsymbol{\beta}) := \min_{\boldsymbol{\beta} \in \Theta}\ T^{-1} \sum_{t=p+1}^{T} \rho\left(\frac{\varepsilon_{t,d} - \boldsymbol{\beta}' \mathbf{x}^*_{t-1,d}}{\widehat{\sigma}}\right) \tag{8}$$

where $\rho(\cdot)$ is a measurable function, $\widehat{\sigma}$ is a preliminary consistent estimate of the scale of the residuals, and $\Theta \in \mathbb{R}^{p+2}$ denotes the parameter space. When $\theta = 0$, then $\phi = 0$ holds true in (6), independently of the choice of the $\rho(\cdot)$ function, and therefore the null hypothesis that $\{y_t\}$ is $FI(d)$ can be addressed by testing the restriction $\text{H}_0 : \phi = 0$ on the solution of (8), with the validity of this procedure relying once more on the LM principle under quasi-ML estimation; see Appendix A for a discussion. The particular choice of the

[2] This term correspond to the M-measure of location of the error term; see, for instance, Huber (1981). Note, for instance, that in the QR setting in Hassler *et al.* (2016), the theoretical value of the intercept corresponds to the τ-th conditional quantile of $\{e_t\}$, which is dictated by the distribution of innovations and the (arbitrary) choice of the quantile in the estimation.

$\rho(\cdot)$ function is in general driven by efficiency, robustness, or computational issues and characterizes the properties of the resultant M estimator. For instance, $\rho(r/\widehat{\sigma}) = r^2$ leads to the LS estimation in Demetrescu *et al.* (2008). Also, $\rho(r/\widehat{\sigma}) = r(\tau - I(r < 0))$, with $\tau \in (0,1)$, leads to the QR estimator in Hassler *et al.* (2016), and $\rho(r/\widehat{\sigma}) = -\ln f_\varepsilon(r)$, with $f_\varepsilon(r)$ denoting a differentiable density function, leads to the ML estimator.

3.2 Asymptotic distribution of the M estimator

In this section, we characterize the existence, consistency, and first-order asymptotic distribution of the solution of $\min_{\boldsymbol{\beta}\in\Theta} Q_T^*(\boldsymbol{\beta})$ for a class of $\rho(\cdot)$ functions. This is the basis to construct tests for fractional integration. To this end, we first introduce further notation and additional conditions that, together with Assumption 1, conform the set of sufficient conditions in our analysis. Thus, define the variables,

$$u_{t,s}^*(\boldsymbol{\beta}) := \frac{\varepsilon_{t,d} - \boldsymbol{\beta}'\mathbf{x}_{t-1,d}^*}{s}; \text{ and } u_{t,s}^{**}(\boldsymbol{\beta}) := \frac{\varepsilon_{t,d} - \boldsymbol{\beta}'\mathbf{x}_{t-1,d}^{**}}{s} \tag{9}$$

for all $\boldsymbol{\beta} \in \Theta$, where $s > 0$ is a generic scale factor and $\mathbf{x}_{t-1,d}^{**} := \left(1, x_{t-1,d}^{**}, \varepsilon_{t-1,d}, ..., \varepsilon_{t-p,d}\right)'$, with $x_{t-1,d}^{**} := \sum_{j=1}^{\infty} j^{-1}\varepsilon_{t-j,d}$. Similarly, given the theoretical value α_ρ, formally defined in Assumption 2 below, define $\widetilde{e}_{t,s} := (e_t - \alpha_\rho)/s$, noting that under $\mathrm{H}_0 : \phi = 0$, $u_{t,s}^{**}(\boldsymbol{\beta}) := \widetilde{e}_{t,s}$.

Assumption 2. $\rho : \mathbb{R} \to \mathbb{R}$ *is a measurable function satisfying the following conditions:*
i) $\rho(r)$ *is twice differentiable;*
ii) $\psi(r) := \partial\rho(r)/\partial r$ *is bounded;*
iii) $\psi'(r) := \partial\psi(r)/\partial r$ *is first-order Lipschitz continuous and bounded;*
iv) $E(\psi(\widetilde{e}_{t,\sigma})|\mathcal{F}_{t-1}) = 0$, $E(\psi'(\widetilde{e}_{t,\sigma})|\mathcal{F}_{t-1}) > 0$ *almost surely, where* α_ρ *is the unique, real-valued solution of* $\min_{c\in\Theta_\alpha} E\left(\rho\left(\frac{e_t - c}{\sigma}\right)|\mathcal{F}_{t-1}\right)$;
v) $E(\psi'(\widetilde{e}_{t,\sigma})\, e_t|\mathcal{F}_{t-1})$ *and* $E(\psi'(\widetilde{e}_{t,\sigma})|\mathcal{F}_{t-1})$ *are constant.*

Assumption 3. *The scale estimator* $\widehat{\sigma}$ *satisfies* $\widehat{\sigma} - \sigma = o_p\left(T^{-1/4}\right)$.

Assumption 4. $\boldsymbol{\beta}_0 \in \Theta$, *where* $\Theta = \Theta_\alpha \times \Theta_\kappa$ *is a compact subset of* $\mathbb{R}^{p+2}$, *with* $\boldsymbol{\beta}_0 := (\alpha_\rho, \boldsymbol{\kappa}_0')'$ *and* $\boldsymbol{\kappa}_0 := (0, a_1, ..., a_p)'$ *denoting the true value of* $\boldsymbol{\beta}$ *when* $\mathrm{H}_0 : \theta = 0$ *and the previous assumptions hold true.*

Remark 2. Assumption 2 is slightly more general than related conditions in the extant literature; see, for instance, Lucas (1995). In this literature, it is often assumed

that $\rho(r)$ is bounded, but this restriction is not essential and can be dispensed since $E\left(\left|\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)\right|\right) < \infty$, here implied under Assumption 1. Condition *i)* requires twice differentiability, which can be weakened by simply requiring Lipschitz-continuity without affecting consistency, but *i)* plays a role in deriving the limiting distributions; see Appendix B for details. The boundedness and smoothness conditions in *ii)* and *iii)* aim to reduce the influence of large observations. Condition *iv)* ensures that $\boldsymbol{\beta}_0$ is the unique solution of $\min_{\boldsymbol{\beta}\in\Theta} E\left(\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)\right)$ under the null hypothesis. Condition *v)* is a technical restriction that ensures the asymptotic negligibility of certain remaining terms. It holds true, for instance, when $\{e_t\}$ is an i.i.d. process, but it may hold under more general conditions as well. Assumption 3 requires $\widehat{\sigma}$ to be consistent at a rate greater than $T^{1/4}$, which could be obtained from the residuals of any preliminary $\sqrt{T}$-consistent estimate of $\boldsymbol{\beta}$; see Theorem 3 below. Finally, Assumption 4 is standard in this framework.

The following Theorems characterize the existence, consistency, and asymptotic null distribution of $\widehat{\boldsymbol{\beta}}_M$ under $\mathrm{H}_0 : \theta = 0$ and Assumptions 1 to 4. Detailed proofs of all these theoretical statements are provided in Appendix B.

Theorem 1. *Let* $\{y_t\}_{t=1}^T$ *be a sample generated according to* (1). *Consider the optimization problem* $\min_{\boldsymbol{\beta}\in\Theta} Q_T^*(\boldsymbol{\beta})$ *as characterized in* (8), *with* $\{\varepsilon_{t,d}\}$ *and* $\left\{x_{t-1,d}^*\right\}$ *generated as in (2) and (3), respectively. Under the null hypothesis,* $\mathrm{H}_0 : \theta = 0$, *and Assumptions 1 to 4, there exists a random vector* $\widehat{\boldsymbol{\beta}}_M$ *which solves* $\min_{\boldsymbol{\beta}\in\Theta} Q_T^*(\boldsymbol{\beta})$ *such that,*

$$\widehat{\boldsymbol{\beta}}_M \xrightarrow{p} \boldsymbol{\beta}_0 \tag{10}$$

with $\boldsymbol{\beta}_0 := (\alpha_\rho, 0, a_1, ..., a_p)'$ *denoting the vector of true parameters.*

Theorem 2. *For constants* K *and* ζ *such that* $K > 0$ *and* $3/8 < \zeta < 1/2$, *define* $\Phi_T := \{\boldsymbol{\beta} \in \Theta : T^\zeta \, ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq K\}$ *and let* $\widetilde{\boldsymbol{\beta}}_M$ *be the solution of* $\min_{\boldsymbol{\beta}\in\Phi_T} Q_T^*(\boldsymbol{\beta})$. *Then, under the null hypothesis,* $\mathrm{H}_0 : \theta = 0$, *and Assumptions 1-4, it follows that:*

$$\sqrt{T}\left(\widetilde{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_0\right) \Rightarrow \mathcal{N}(0, \boldsymbol{\Omega}_{\boldsymbol{\beta}}) \tag{11}$$

where $\boldsymbol{\Omega}_{\boldsymbol{\beta}} := \sigma^2 \mathbf{A}_{\boldsymbol{\beta}}^{-1} \mathbf{B}_{\boldsymbol{\beta}} \mathbf{A}_{\boldsymbol{\beta}}^{-1}$, $\mathbf{A}_{\boldsymbol{\beta}} := E\left(\psi'\left(\widetilde{e}_{t,\sigma}\right) \mathbf{x}_{t-1,d}^{**} \mathbf{x}_{t-1,d}^{\prime **}\right)$, $\mathbf{B}_{\boldsymbol{\beta}} := E\left(\psi^2\left(\widetilde{e}_{t,\sigma}\right) \mathbf{x}_{t-1,d}^{**} \mathbf{x}_{t-1,d}^{\prime **}\right)$.

Remark 3. When the ρ function is strictly convex such that ψ is monotonically increasing, $Q_T^*(\boldsymbol{\beta})$ is a strictly convex function with a unique minimum attainable in Θ, so $\widehat{\boldsymbol{\beta}}_M$ exists and is unique. Unfortunately, strict convexity is not compatible with the class of ρ-functions typically used in related literature and, hence, we do not impose this restriction. As a result, $\min_{\boldsymbol{\beta}\in\Theta} Q_T^*(\boldsymbol{\beta})$ may present multiple solutions corresponding to local minima, a well-known practical concern in M estimation; see Amemiya (1985). Theorem 1 ensures the existence of a 'correct' solution which converges in probability to the true parameter vector. In order to characterize the asymptotic distribution of this solution, Theorem 2 considers the compact ball $\Phi_T \subset \Theta$ in a local neighborhood of $\boldsymbol{\beta}_0$. The bounds on ζ are chosen to ensure that any $\sqrt{T}$-consistent preliminary estimator of $\boldsymbol{\beta}$ belongs to Φ_T with probability one. Since $Q_T^*(\boldsymbol{\beta})$ is continuous on Θ and Φ_T is compact, the objective function takes its minimum value in Φ_T, which ensures the existence of a local solution, denoted $\widetilde{\boldsymbol{\beta}}_M$, which necessarily corresponds to the global minimum in large samples. Consequently, Theorem 2 characterizes the asymptotic behavior of the consistent solution in Theorem 1, showing that $\widehat{\boldsymbol{\beta}}_M$ is $\sqrt{T}$-consistent and asymptotically normal distributed with zero mean and covariance matrix $\boldsymbol{\Omega}_{\boldsymbol{\beta}}$.

Remark 4. Consistency holds if twice differentiability in Assumption 2i) is replaced by Lipschitz continuity, a more general condition; see Lemma A1 in Appendix B. This allows us to extend consistency to the QR context as a corollary of Theorem 2. QR is characterized by $\rho(r/\widehat{\sigma}) = r(\tau - I(r < 0))$, $\tau \in (0,1)$, *i.e.,* a Lipschitz-continuous function not differentiable at the origin. Hence, the QR test in Hassler *et al.* (2016) –derived under i.i.d. innovations– is shown to generate consistent estimates under the more general conditions considered here. Since Hassler *et al.* (2016) require $L_{2+\epsilon}$-bounded innovations for some $\epsilon > 0$, and Assumption 1 requires $\epsilon \geq 2/3$, this property may seem to come at the cost of a slight strengthening of the moment condition. Neverthelss, this requirement is due to scaling the residuals under ρ in Assumption 2 which, unlike the check function used for QR, is not homogenous. Consequently, for the specific purpose of showing consistency in QR, Assumption 1 could be weakened to simply require $\epsilon > 0$ as in Hassler *et al.* (2016); see Appendix B.

3.3 Asymptotic distribution of iterated estimators

Alternatively to the direct optimization of $Q_T^*(\boldsymbol{\beta})$, a consistent M estimator of $\boldsymbol{\beta}$ can be obtained using iterated numerical methods. These build on a preliminary $\sqrt{T}$-consistent estimate, say $\widehat{\boldsymbol{\beta}}_{(0)}$, which can be obtained from the optimization of a strictly convex function. This preliminary estimate is then iterated in a numeric algorithm without engaging in further optimization, obtaining an estimator which can be shown to be asymptotically distributed as $\widehat{\boldsymbol{\beta}}_M$; see, for example, Kreiss (1985), Welsh and Ronchetti (2002), and Ling and Li (2003).

In this section, we first propose a suitable methodology to determine $\widehat{\boldsymbol{\beta}}_{(0)}$, and then characterize the asymptotic null distribution of two alternative estimators building on iterated algorithms, namely, the k-iterated Newton-Raphson and the Iterated Reweighted Least Squares (IRLS) algorithms.

Theorem 3. *Let $\widehat{\boldsymbol{\kappa}}_{(0)}$ be a $\sqrt{T}$-consistent estimate of $\boldsymbol{\kappa} := (\phi, \pi_1, ..., \pi_p)'$ under Assumption 1. Denote $\widehat{e}_{(0)t} := \varepsilon_{t,d} - \widehat{\boldsymbol{\kappa}}'_{(0)}\mathbf{z}^*_{t-1,d}$ and let $\widehat{\sigma}^2_{(0)}$ be the sample variance of $\widehat{e}_{(0)t}$. Furthermore, let $\widehat{\alpha}_{(0)} := \arg\min_{c\in\Theta_\alpha} \sum_{t=p+1}^{\infty} \rho\left(\frac{\widehat{e}_{(0)t}-c}{\widehat{\sigma}_{(0)}}\right)$ and $\widehat{\boldsymbol{\beta}}_{(0)} := \left(\widehat{\alpha}_{(0)}, \widehat{\boldsymbol{\kappa}}'_{(0)}\right)'$. Under Assumptions 1 to 4 and $\mathrm{H}_0 : \theta = 0$, it then follows that $\widehat{\boldsymbol{\beta}}_{(0)} = \boldsymbol{\beta}_0 + O_p\left(T^{-1/2}\right)$ and $\widehat{\sigma}_{(0)} - \sigma = o_p\left(T^{-1/4}\right)$.*

Remark 5. The simplest method to construct a $\sqrt{T}$-consistent estimate of $\boldsymbol{\kappa}$ under Assumption 1 is LS, but other methods are possible as well. Given the resulting residuals, it is straightforward to construct a consistent estimate of σ satisfying Assumption 3. The simplest alternative is the standard deviation of residuals. Other alternatives building on absolute residuals, such as transformation of the mean absolute deviation, may render this property as well. Finally, given the regression residuals $\left\{\widehat{e}_{(0)t}\right\}$, a preliminary estimate of α_ρ arises by solving the M equation. Optimization at this stage only involves a single parameter and, hence, grid-search methods are highly effective in ensuring convergence to the global minimum. As a result, $\widehat{\boldsymbol{\beta}}_{(0)} := \left(\widehat{\alpha}_{(0)}, \widehat{\boldsymbol{\kappa}}'_{(0)}\right)'$ can be seen as a two-stage consistent preliminary estimator of $\boldsymbol{\beta}_0$.

3.3.1 Iterated Newton-Raphson estimators

Given $\widehat{\boldsymbol{\beta}}_{(0)}$ and $\widehat{\sigma}_{(0)}$, the one-step M estimator of $\boldsymbol{\beta}$ based on the Newton-Raphson algorithm, denoted $\widehat{\boldsymbol{\beta}}_{NR}$, is determined as

$$\widehat{\boldsymbol{\beta}}_{NR} = \widehat{\boldsymbol{\beta}}_{(0)} - \left[\frac{1}{\widehat{\sigma}_{(0)}} \sum_{t=p+1}^{T} \psi' \left(u^*_{t,\widehat{\sigma}_{(0)}} (\boldsymbol{\beta}) \right) \mathbf{x}^*_{t-1,d} \mathbf{x}^{*\prime}_{t-1,d} \right]^{-1}_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_{(0)}} \left[\sum_{t=p+1}^{T} \psi \left(u^*_{t,\widehat{\sigma}_{(0)}} (\boldsymbol{\beta}) \right) \mathbf{x}^*_{t-1,d} \right]_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_{(0)}} \tag{12}$$

see Lehmann (1983, Theorems 3.1 and 4.2 of Chapter 6). This procedure can be iterated a finite k number of times, leading to the k-step Newton-Raphson M estimator of $\boldsymbol{\beta}$. In practice, the most common choice is $k = 1$, since a single iteration starting from a $\sqrt{T}$-consistent pre-estimate suffices to ensure the asymptotic properties of M estimators. The following result formally proves this statement in our context and characterizes the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{NR}$ under the set of assumptions considered.

Theorem 4. *Let $\widehat{\boldsymbol{\beta}}_{(0)}$ be a preliminary estimator of $\boldsymbol{\beta}$ such that $\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)} - \boldsymbol{\beta}_0\right) = O_p(1)$ and $\widehat{\sigma}_{(0)} - \sigma = o_p\left(T^{-1/4}\right)$ under $H_0 : \theta = 0$ and Assumptions 1 to 4. Denote $\widehat{\boldsymbol{\beta}}_{NR}$ as the M estimator of $\boldsymbol{\beta}$ based on a one-step iteration of the Newton-Raphson algorithm as defined in* (12). *Then, under the set of conditions considered,*

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{NR} - \boldsymbol{\beta}_0\right) \Rightarrow \mathcal{N}(0, \boldsymbol{\Omega}_{\boldsymbol{\beta}}). \tag{13}$$

Remark 6. Theorem 4 shows that iterated estimators from the Newton-Raphson algorithm, based on a preliminary $\sqrt{T}$-consistent estimate, have the same asymptotic null distribution as the consistent solution from $\min_{\boldsymbol{\beta}\in\Theta} Q^*_T(\boldsymbol{\beta})$. The null asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{NR}$ is not affected by the distribution of the preliminary estimate $\widehat{\boldsymbol{\beta}}_{(0)}$ and, remarkably, even a single iteration suffices to produce an estimator which is asymptotically equivalent to that obtained from the direct numerical optimization of $Q^*_T(\boldsymbol{\beta})$. Furthermore, because a single iteration produces a $\sqrt{T}$-consistent estimator of $\boldsymbol{\beta}$, Theorem 4 applies trivially on any of the subsequent iterations, thereby characterizing the distribution of k-iterated Newton-Raphson estimators for any $k \geq 1$ when building on the estimates of the previous iteration.

3.3.2 Iterated Reweighted Least Squares (IRLS)

The k-step Newton-Raphson algorithm involves the computation of the Hessian. Alternative methods which do not rely upon this estimation may result more attractive. Among these alternatives, the IRLS estimator is probably the most common numerical technique implemented in practice. This procedure exploits the analogy between the first-order condition of the optimization problem and the equation vector that characterizes a simple Weighted Least Squares (WLS) problem. In particular, if we denote $\omega(r_t) := \psi(r_t)/r_t$, setting $\omega(0) := 0$, the first-order condition equation system $\sum_{t=p+1}^{T} \psi\left(u^*_{t,\widehat{\sigma}}(\boldsymbol{\beta})\right) \mathbf{x}^*_{t-1,d} = 0$ that characterizes M estimators can be rewritten as $\sum_{t=p+1}^{T} \omega\left(u^*_{t,\widehat{\sigma}}(\boldsymbol{\beta})\right) u^*_{t,\widehat{\sigma}}(\boldsymbol{\beta}) \mathbf{x}^*_{t-1,d} = 0$. As a result, $\widehat{\boldsymbol{\beta}}_M := \min_{\boldsymbol{\beta}\in\Theta} Q^*_T(\boldsymbol{\beta})$ admits an implicit WLS-type representation of the form:

$$\left(\sum_{t=p+1}^{T} \omega\left(u^*_{t,\widehat{\sigma}}(\boldsymbol{\beta})\right) \mathbf{x}^*_{t-1,d}\mathbf{x}'^*_{t-1,d}\right)^{-1} \sum_{t=p+1}^{T} \omega\left(u^*_{t,\widehat{\sigma}}(\boldsymbol{\beta})\right) \mathbf{x}^*_{t-1,d}\varepsilon_{t,d}. \tag{14}$$

This property suggests an iterated algorithm to approximate the solution in the optimization problem by recursive methods using (14). In particular, starting from a preliminary estimate $\widehat{\boldsymbol{\beta}}_{(0)}$, one obtains the scaled residuals $u^*_{t,\widehat{\sigma}}\left(\widehat{\boldsymbol{\beta}}_{(0)}\right)$ and the corresponding weights $\omega_t\left(u^*_{t,\widehat{\sigma}}\left(\widehat{\boldsymbol{\beta}}_{(0)}\right)\right)$ and, hence, a new estimator, say $\widehat{\boldsymbol{\beta}}_{IRLS,(1)}$, by direct computation of (14). This procedure is then repeated a number of times until convergence, leading to the IRLS M estimator of $\boldsymbol{\beta}$, which we shall denote as $\widehat{\boldsymbol{\beta}}_{IRLS}$. Theorem 5 below characterizes the asymptotic distribution of this estimator upon additional regularity conditions.

Theorem 5. *Let $\widehat{\boldsymbol{\beta}}_{(0)}$ be a preliminary estimate of $\boldsymbol{\beta}$ satisfying Theorem 3, and denote $\widehat{\boldsymbol{\beta}}_{IRLS}$ as the M estimator of $\boldsymbol{\beta}$ based on k iterations of the IRLS algorithm. Under the null hypothesis, Assumptions 1 to 4, and when i) $\psi(r)$ is odd such that $\psi(0) = 0$ and $\psi''(0)$ is finite, ii) $\omega(r)$ is Lipschitz continuous, and iii) the eigenvalues of $\mathbf{D}_{\boldsymbol{\beta}} := \mathbf{I} - \mathbf{C}^{-1}_{\boldsymbol{\beta}}\mathbf{A}_{\boldsymbol{\beta}}$ are smaller than unity, with $\mathbf{C}_{\boldsymbol{\beta}} := \text{plim}_{T\to\infty} T^{-1}\sum_{t=p+1}^{T} \omega_t(\widetilde{e}_{t,\sigma})\mathbf{x}^{**}_{t-1,d}\mathbf{x}'^{**}_{t-1,d}$, it follows that:*

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{IRLS} - \boldsymbol{\beta}_0\right) \Rightarrow \mathcal{N}(0, \boldsymbol{\Omega}_{\boldsymbol{\beta}})$$

when k is allowed to diverge.

Remark 7. The condition that ψ is odd is standard in this literature; *e.g.*, Welsh and Ronchetti (2002). In contrast to the Newton-Raphson algorithm, the IRLS method generally yields a numerical approximation of the true global solution. When $\mathbf{C}_{\boldsymbol{\beta}} = \mathbf{A}_{\boldsymbol{\beta}}$, the term $\mathbf{D}_{\boldsymbol{\beta}} = 0$ and the approximation of the asymptotic distribution of the test is exact for any finite number of iterations. This may be the case, for instance, if ρ has a constant, non-trivial third derivative. More generally, $\mathbf{C}_{\boldsymbol{\beta}}$ may differ from $\mathbf{A}_{\boldsymbol{\beta}}$, but if the difference is small enough because the eigenvalues of $\mathbf{D}_{\boldsymbol{\beta}}$ are smaller than unity, the asymptotic null distribution of $\widehat{\boldsymbol{\beta}}_{IRLS}$ is the same as $\widehat{\boldsymbol{\beta}}_M$ when the number of iterations is allowed to diverge. It can be seen from the proof (see Appendix B) that the k-step IRLS estimator, as T diverges, has a different limiting distribution for each k. In this context, condition *iii*) in Theorem 5 guarantees numerical convergence of this sequence of approximations for a finite, large enough value of k. This is the case, for instance, when using Huber's function.

3.4 Testing for fractional integration

Building on any of the previous estimates of $\boldsymbol{\beta}$ and on a consistent estimate of the asymptotic covariance matrix, a test for the order of fractional integration of $\{y_t\}$ can readily be implemented through a standard t-statistic. The asymptotic distribution of this test is formally stated in the following Theorem.

Theorem 6. *Let* $\widehat{\boldsymbol{\beta}}_{\mathcal{M}} \in \mathcal{S}$, $\mathcal{S} := \left\{\widehat{\boldsymbol{\beta}}_M, \widehat{\boldsymbol{\beta}}_{NR}, \widehat{\boldsymbol{\beta}}_{IRLS,}\right\}$, *be the M estimator of* $\boldsymbol{\beta}$ *such that* $\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{\mathcal{M}} - \boldsymbol{\beta}_0\right) = O_p(1)$ *holds true under the respective conditions outlined in Theorems 2, 4 and 5. Given* $\widehat{u}_t := \varepsilon_{t,d} - \widehat{\boldsymbol{\beta}}'_{\mathcal{M}} x^*_{t-1,d}$, *define* $\widehat{\sigma}^2_{\mathcal{M}} := T^{-1}\sum_{t=p+1}^{T}(\widehat{u}_t - \overline{u})^2$, *and the matrices* $\mathbf{A}_{\boldsymbol{\beta}T} := T^{-1}\sum_{t=p+1}^{T}\psi'\left(\frac{\widehat{u}_t}{\widehat{\sigma}_{\mathcal{M}}}\right)\mathbf{x}^*_{t-1,d}\mathbf{x}'^*_{t-1,d}$, $\mathbf{B}_{\boldsymbol{\beta}T} := T^{-1}\sum_{t=p+1}^{T}\psi^2\left(\frac{\widehat{u}_t}{\widehat{\sigma}_{\mathcal{M}}}\right)\mathbf{x}^*_{t-1,d}\mathbf{x}'^*_{t-1,d}$, *and* $\boldsymbol{\Omega}_{\boldsymbol{\beta}T} := \widehat{\sigma}^2_{\mathcal{M}}\ \mathbf{A}^{-1}_{\boldsymbol{\beta}T}\ \mathbf{B}_{\boldsymbol{\beta}T}\ \mathbf{A}^{-1}_{\boldsymbol{\beta}T}$, *it then follows that* $\widehat{\sigma}^2_{\mathcal{M}} \xrightarrow{p} \sigma^2$, $\mathbf{A}_{\boldsymbol{\beta}T} \xrightarrow{p} \mathbf{A}_{\boldsymbol{\beta}}$, $\mathbf{B}_{\boldsymbol{\beta}T} \xrightarrow{p} \mathbf{B}_{\boldsymbol{\beta}}$, *and, consequently,*

$$\boldsymbol{\Omega}_{\boldsymbol{\beta}T} \xrightarrow{p} \boldsymbol{\Omega}_{\boldsymbol{\beta}}.$$

Hence, under the null hypothesis, $\mathrm{H}_0 : \theta = 0$, *and the remaining assumptions considered*,

$$t_{\mathcal{M}} := \frac{\widehat{\phi}_{\mathcal{M}}}{\sqrt{\widehat{\omega}_{22}/T}} \Rightarrow \mathcal{N}(0,1)$$

where $\widehat{\phi}_{\mathcal{M}}$ *and* $\widehat{\omega}_{22}$ *denote the second element in* $\widehat{\boldsymbol{\beta}}_{\mathcal{M}}$ *and in the diagonal of* $\boldsymbol{\Omega}_{\boldsymbol{\beta}T}$, *respectively.*

Remark 8. Theorem 6 states the null distribution of the t-statistic for $\mathrm{H}_0 : \theta = 0$. It also gives a theoretical basis for the construction of confidence intervals that include the true value of the fractional parameter with $100\,(1-\lambda)\,\%$ asymptotic coverage by inverting the non-rejection region of $t_{\mathcal{M}}$; see Hassler *et al.* (2016). More specifically, let $t_{\mathcal{M}}(l)$ denote the value of $t_{\mathcal{M}}$ when testing $\mathrm{H}_0 : y_t \sim FI\,(l)$ for an arbitrary l. For a closed interval Ψ in $\mathbb{R}$, define $\mathcal{D}_\lambda = \{s \in \Psi : \Pr\,[\mathcal{Z} \leq |t_{\mathcal{M}}(s)|] \leq 1-\lambda\}$ with $\lambda \in (0,1)$, and $\mathcal{Z}$ the standard normal variate, *i.e.*, the subset of Ψ for which the null hypothesis cannot be rejected at the λ significance level. From Theorem 6, it follows that if $\mathcal{D}_\lambda$ is in the interior of Θ, then the probability of the true order of integration being within $\mathcal{D}_\lambda$ is at least $(1-\lambda)$. Thus, a confidence interval can be constructed through a grid-search process in Ψ, which is computationally feasible because the fractional parameter tipically lies in the interval $(0,1)$.

Finally, the following theorem characterizes the asymptotic distribution of $t_{\mathcal{M}}$ under sequences of local alternatives in a $\sqrt{T}$ neighbourhood of the null hypothesis and, hence, completes the theoretical discussion. It is shown that M-based tests for fractional integration exhibit non-trivial power against such alternatives, so departures from the null hypothesis will be detected with increasing probability.

Theorem 7. *Consider a sequence of local alternatives of the form* $\mathrm{H}_1 : \theta = c/\sqrt{T}$ *for some finite* $c \neq 0$. *Under the assumptions of Theorem 6, it then follows under the alternative hypothesis that:*

$$t_{\mathcal{M}} \Rightarrow \mathcal{N}\left(\frac{c}{\sqrt{\omega_{22}}},\, 1\right)$$

where ω_{22} *is the second element in the diagonal of* $\mathbf{\Omega}_{\boldsymbol{\beta}}$. *Consequently,* $\Pr\big(|t_{\mathcal{M}}| > z_{1-\lambda/2}\big)$ *is an increasing sequence on* $|c|$, *with* $z_{1-\lambda/2}$ *denoting the two-sided critical value of the standard normal distribution for the* $\lambda\%$ *nominal size level.*

4 Monte Carlo analysis

In this section, we analyze the finite-sample properties of the M tests for fractional integration by means of Monte Carlo simulation. We consider two alternative ρ-functions widely used in related literature, namely, the so-called Huber and Bisquare (or Biweight)

ρ-functions defined, respectively, as $\rho(r) = 0.5r^2 I(|r| \leq c_H) + \left[c_H|r| - \frac{k^2}{2}\right] I(|r| > k)$ with $c_H = 1.345$, and $\rho(r) = \left[\frac{c_B^2}{6}\left\{1 - \left[1 - \left(\frac{r}{k}\right)^2\right]^3\right\}\right] I(|r| \leq c_B) + \frac{c_B^2}{6} I(|r| > c_B)$ with $c_B = 4.685$. Since IRLS is the most common approach used in practice, we report results based on this algorithm, noting that results based on Newton-Raphson estimation are similar. In the empirical implementation of the IRLS algorithm, we first use LS to estimate a preliminary value of the slope parameters in the augmented regression, and then infer the scatter of the residuals using the Mean Absolute Deviation, as usual in this literature. We iterate until convergence setting max $k = 100$.

The t-statistics from the Huber and Bisquare ρ-functions are denoted as t_M^H and t_M^B, respectively. For benchmarking purposes, we consider the LS-based test statistics with standard errors computed for i.i.d. innovations as in Breitung and Hassler (2000), denoted t_{LS}, and using Eicker-White's correction against heteroskedasticity as in Demetrescu *et al.* (2008), denoted t_{LS}^{HC}. In addition, we compute the QR t-test in Hassler *et al.* (2016) from the QR estimation of (6) at the 50th quantile, denoted t_{QR}, using Powell's robust standard errors with a Gaussian kernel and bandwidth parameter $0.3 \times \min\{\widehat{\sigma}_u, IQR_u\} \times T^{-1/5}$, where $\widehat{\sigma}_u$ and IQR_u denote the sample standard deviation and the sample interquartile range of the residuals of the regression, respectively. As discussed previously, both LS and QR can be seen as particular cases of the generalized M estimation framework, exhibiting different properties because of the different choices of the weighting function implemented. LS-based inference is expected to provide more powerful results under normality, whereas QR-based inference can lead to enhanced power in relation to LS under departures from normality, as shown experimentally in Hassler *et al.* (2016). Finally, we also consider the sign test in Delgado and Velasco (2005), denoted DV. This test is based on the same harmonic weighting structure that characterizes the regression-based tests discussed in this paper, but has the outstanding property of being formally valid even if $E(e_t^2) = \infty$. On the other hand, it requires the median of e_t to be zero, i.e., requires symmetric errors, which in practice may result excessively restrictive, but which holds true in our experimental analysis. The DV test statistic is asymptotically distributed as a standard normal distribution under the null hypothesis and, given $S_{t,d} := sign(\varepsilon_{t,d})$, can be computed as

$$DV = \sqrt{\frac{6}{\pi^2 T}} \sum_{j=1}^{T-1} \frac{1}{j} \sum_{t=j+1}^{T} S_{t,d} S_{t-j,d} \tag{15}$$

4.1 Short-run dynamics

In the first experiment, we consider a DGP given by $(1-a_1L)(1-L)^{1+\theta}y_t = e_t,\ t = 1, ..., T$, where $a_1 \in \{0, 0.5\}$, $\{e_t\}$ are i.i.d. innovations drawn from a Student-t distribution with $v \in \{2, 3, 1000\}$ degrees of freedom, and $T \in \{250, 500\}$. The case $v = 1000$ corresponds closely to the Gaussian distribution, whereas $v \in \{2, 3\}$ are characterized by heavy-tailed distributions, having infinite variance when $v = 2$. As in Breitung and Hassler (2002) and Hassler *et al.* (2016), under the null hypothesis we test for a unit root, namely, $\mathrm{H}_0 : y_t \sim FI(1)$, noting that the true order of integration is given by $d_0 = 1 + \theta$ with $\theta \in \{-0.3, ..., -0.1, 0, 0.1, ..., 0.3\}$. We then compute the different test stastistics against a two-sided alternative at the 5% significance level and analyze the average rejection frequencies given $5,000$ replications of the experiment. The case $\theta = 0$ determines the empirical size of the tests, while values $\theta \neq 0$ characterize the finite-sample power behaviour. Finally, the autoregressive coefficients $a_1 = 0$ and $a_1 = 0.5$ allow us to analyze the performance of the tests under errors driven by i.i.d. innovations and short-run dependence, respectively. In the latter case, LS- and M-based tests are computed from auxiliary regressions augmented with one lag of the dependent variable, while the DV test is computed on the residuals of a first-order autoregression.

[Insert Table 1 around here]

Table 1 reports the rejection frequencies in the i.i.d. experiment ($a_1 = 0$). Under the null hypothesis, all regression-based tests exhibit approximately correct size with rejection frequencies close to 5%. LS-based estimation tends to yield more stable results than QR in small samples, since the latter requires numerical optimization, but these differences tend to disappear quickly as the sample size increases. For $v = 2$ and $T = 500$, t_{LS} and t_{QR} suffer size departures in this experiment, which is not surprising because errors have infinite variance and this possibility is ruled out in their theoretical derivations. In contrast, the remaining tests, including t_{LS}^{HC}, exhibit good size performance in this context, even though only DV is theoretically ensured to exhibit correct nominal size (asymptotically) when errors have infinite variance.

Under the alternative hypotheses, the Gaussian environment $v = 1000$ provides the conditions for the optimality of t_{LS}, which generally outperforms the alternative tests.

The differences are fairly small under M estimation and tend to be more substantial for the QR and DV tests. When $v = 2$ or $v = 3$, LS-based tests tend to become conservative in relation to QR- and M-based tests. In this context, t_{LS} is no longer efficient, and the Monte Carlo analysis confirms that M-estimation leads to large gains in relative power. For instance, for $T = 250$, $v = 3$ and $\theta = -0.1$, the power of t_{LS} and t_{LS}^{HC} is 54.00% and 61.22%, respectively, whereas that of t_M^H and t_M^B is 73.62% and 72.96%, respectively. The differences in power between the two IRLS tests are small and tend to disappear as the sample length increases. The QR test exhibits enhanced properties under heavy-tailed distributions, as reported in Hassler *et al.* (2016), but the power of this test tends to be dominated by M testing. For instance, for $T = 250$, $v = 3$ and $\theta = -0.1$, t_{QR} has power of 64.10%. The DV test displays considerably smaller power in comparison to the other tests, which is not surprising because sign tests are fairly robust, but known to exhibit reduced power in small samples.

The ability to reject the false null increases with the sample length. In the case of M-based tests, this is formally expected from Theorem 7. For instance, for $T = 500$, $v = 3$ and $\theta = -0.1$, the power of t_{LS} and t_{LS}^{HC} is 82.24% and 83.64%, showing sizeable increments. Similarly, the power of the M tests, t_M^H and t_M^B, is 96.04% and 95.74%, respectively, while the power of t_{QR} is 90.98%. All these tests exhibit good power, and the differences tend to disappear in large samples, but M-based tests clearly outperform the other alternatives under heavy-tailed innovations. Interestingly, power exhibits an asymmetric pattern such that the relative gains tend to be much larger when $\theta < 0$, a feature noted in previous literature (see e.g. Hassler *et al.*, 2016). This pattern is data-dependent and, for instance, tends to disappear as v approaches 2, for which power exhibits a more symmetric behavior.

[Insert Table 2 around here]

Table 2 reports the rejection frequencies when the DGP is driven by weakly-dependent errors with $a_1 = 0.5$. Under the null hypothesis all tests, and particularly the DV test, display finite sample size departures as a consequence of augmentation. These departures are more evident for $T = 250$, and with Gaussian errors. The size distortion caused by augmentation is a small-sample feature and, it is almost completely eliminated when the sample length increases to $T = 500$. For the DV test, the overall distortion still remains sizeable with $T = 500$. Finally, under the alternative hypotheses, all tests display

significant power reductions in relation to the i.i.d. context, a well-known feature caused by augmentation. Nevertheless, power increases as $|\theta|$ and/or the sample size increases, showing that these distortions are a finite-sample result. In this context, the power of all tests is characterized by a strong asymmetric pattern such that when $\theta < 0$ alternatives are easier detected. As in the i.i.d. case, the power of the M- and QR-based tests largely improves in relation to the LS-based alternatives as the degree of excess kurtosis increases, particularly. when $\theta < 0$.

In summary, the overall picture that emerges from this experiment supports the asymptotic theory discussed in Theorems 6 and 7, showing that M-based tests are well-suited in finite-samples and exhibit approximately correct empirical size. Short-run dependence can be handled successfully through augmentation, particularly in large samples, and M estimation can provide improved performance over LS- and even QR-based alternatives when the data is driven by non-Gaussian distributions.

4.2 MDS with time-varying volatility

The second experiment addresses the empirical size and power of M tests when errors exhibit time-varying dependence in conditional volatility. Assumption 1 does not formally allow for this possibility, but our interest is motivated by the empirical relevance of this pattern in financial variables sampled at high frequencies. As in the previous experiment, data are generated from $(1-L)^{1+\theta}y_t = \varepsilon_t$, where $\varepsilon_t := \sigma_t\eta_t$, σ_t^2 characterizes the GARCH dynamics, and η_t are i.i.d. innovations drawn from a Student-t distribution with $v \in \{2, 3, 1000\}$ degrees of freedom. In the experiment, we consider two GARCH processes, namely, $\sigma_t^2 = 0.01 + 0.10\varepsilon_{t-1}^2 + 0.60\sigma_{t-1}^2$ (GARCH-A) and $\sigma_t^2 = 0.01 + 0.05\varepsilon_{t-1}^2 + 0.90\sigma_{t-1}^2$(GARCH-B), characterizing different degrees of persistence in volatility. Since time-varying volatility is a distinctive pattern of high-frequency data, the sample length is typically large and often spans thousands of obsevations (see, for instance, the empirical analysis in Section 5). In order to acknowledge this empirical feature, we set $T = 1000$ in this experiment. As in the previous experiment, LS- and M-based tests are computed from augmented regressions with one lag of the dependent variable.

[Insert Table 3 around here]

Table 3 presents the rejection frequencies when the volatility dynamics is generated by the GARCH-A and GARCH B processes. The picture that emerges from this analysis is remarkably similar to that discussed previously. Under the null hypothesis, empirical sizes are close to the nominal level in all cases, with the degree of persistence in volatility playing little role. For M-based estimators, size ranges between 5.28% and 6.26%. Under the alternative hypothesis, the empirical rejection frequencies of all tests are very similar regardless of whether GARCH-A or GARCH-B is considered. Remarkably, and as also observed in Tables 1 and 2, departures from normalilty lead the M-tests tend display improved power performance, even in relation to QR-based tests.

5 Empirical analysis: volatility of financial assets

Volatility modeling has taken a predominant position in financial econometrics. In this section, we analyze the long-term dynamics of two alternative daily volatility measures. On the one hand, we consider log absolute returns. Functions of absolute returns, such as log-power transformations, are the most common proxy of the conditional variability of financial assets and exhibit autocorrelations with hyperbolic rates of decay (e.g., Bollerslev and Wright 2000) and a considerable degree of non-normality; see, e.g., Brand and Jones (2006) and references therein. On the other hand, we consider the log transform of the high-low range price range, a simple, yet highly effective empirical proxy of volatility. This measure displays the characteristic pattern of strong persistence of volatility measures, but in contrast to log absolute returns, its distribution is close to be normal. This framework poses an interesting scenario to empirically compare the performance of M tests in relation to LS.

We compute both measures on daily prices of stock market indices in developed and emerging markets over the period 01/01/2000 to 31/12/2016. These include indices for the U.S. (SP500 and NASDAQ Composite), France (CAC40), Germany (DAX30), Japan (NIKKEI250), Spain (IBEX35), U.K. (FTSE100), Brazil (BOVESPA), Hong Kong (HANG SENG), Argentina (MERVAL), and Mexico (MXX).[3] For both volatility measures, Table 4 reports standard descriptive statistics (mean, standard deviation, skewness,

[3]The data used to compute these measures (maximum, minimum and closing prices) are available from commercial data providers and can be obtained freely from Yahoo Finance.

and kurtosis) as well as the Ljung-Box Q-test statistic for absence of serial correlation up to the first 100 lags, and the Jarque-Bera test statistic for normality. The main features of this analysis shall be commented in greater detail below.

[Insert Table 4 around here]

Given these series, we construct 95% confidence intervals for the fractional parameter by inverting the non-rejection regions of the following test statistics: the LS-based test in Demetrescu *et al.* (2008) computed with robust standard errors to (conditional) heteroskedasticity, denoted t_{LS}^{HC}; the QR test in Hassler *et al.* (2016) computed at the median with robust standard errors, as described in the Monte Carlo section, denoted t_{QR}; the Kolmogorov-Smirnov joint QR test to simultaneously address the null hypothesis at the quantiles $[0.1, 0.9]$, as described in Hassler *et al.* (2016), denoted KS; and, finally, the M-based tests computed from IRLS and one-step Newton-Raphson algorithms, denoted IRLS and NR, respectively, with Huber and Bisquare weighting ρ-functions, denoted with subscripts H and B, respectively (e.g., IRLS_{H} and IRLS_{B}). These algorithms are iterated starting from the LS estimation, setting the initial value of the intercept equal to zero and allowing the iterative procedure to freely determine all the parameters involved.

In the implementation of these tests, we follow two common methodological approaches. First, in order to account for a likely non-zero deterministic mean in the level of the volatility measures, we use the demeaning process described in Robinson (1994); see also Demetrescu *et al.* (2008) and Hassler *et al.* (2016). In particular, model (1) can be generalized by setting $y_t = \mu + (1-L)^{-d-\theta}\,\varepsilon_t I(t>1)$ with $\mu \neq 0$. Under $\text{H}_0 : \theta = 0$, $(1-L)_+^d\, y_t = \mu\,(1-L)_+^d + \varepsilon_t I(t>1)$, so μ can be estimated consistently under Assumption 1 from the regression of $\Delta_+^d y_t := \sum_{j=0}^{t-1} \lambda_j(d)\, y_{t-j}$ on $b_{t,d} := \sum_{j=0}^{t-1} \lambda_j(d)$, $t = 2, ..., T$, with $\{\lambda_j(d)\}$ as defined in (4). The residuals of this regression correspond to $\{\varepsilon_t\}$ and, therefore, it suffices to redefine (2) as $\varepsilon_{t,d} := \Delta_+^d y_t - \widehat{\mu} b_{t,d}$, with $\widehat{\mu}$ denoting the estimate of μ, to remove the effects of the deterministic mean. We determine $\widehat{\mu}$ from a LS regression in all cases, noting that alternative estimation methods (e.g., M or QR estimation) may be valid as well. Secondly, the auxiliary regressions are augmented based on Schwert's rule, i.e., $p := \left\lfloor 4\,(T/100)^{1/4} \right\rfloor$; see Demetrescu *et al.* (2008) for a discussion on the convenience of this procedure.

Finally, together with these tests, we implement the DV sign test and the Exact Local Whittle (ELW) estimator in Shimotsu and Phillips (2005). In order to account

for short-run dependence in the DV test, this test is computed on the residuals of an AR(p). The 95% confidence intervals are then constructed by inverting the empirical non-rejection regions of this test. The confidence intervals from the ELW are constructed from point-estimates based on estimation with bandwidth parameter $[T^{0.6}]$, using asymptotic standard errors, and building on the asymptotic normality of this estimator.

5.1 Log absolute returns

For illustrative purposes, Figure 1 displays the time-series dynamics, the sample histogram (confronted with the theoretical normal distribution), and the sample autocorrelation function (with asymptotic 95% confidance bands) of the log absolute-valued returns of the SP500 index. As expected, log absolute returns exhibit a pattern of slow-decaying autocorrelations that suggests long-range persistence. Furthermore, as discussed previously, these series are highly non-normal owing to the occurrence of extreme returns that cause excess kurtosis and skewness, even after applying the log-transform. For instance, the log absolute return of the SP500 index has sample skewness and kurtosis of -1.09 and 5.27, respectively, so normality is strongly rejected according to the Jarque-Bera test owing to extreme observations in the left tail of the distribution. The remaining series are characterized by similar characteristics; see Table 4, Panel A.

[Insert Figure 1 around here]

Table 5 reports the 95% confidence interval estimates for the fractional parameter from the different testing procedures. In this analysis, the DV test always rejects the null hypothesis, failing to provide reasonable estimates, so we do not report results for this statistic. This evidence should not be surprising in view that log absolute returns are strongly left-skewed and that the DV test builds on the critical assumption that the underlying distribution is symmetric. As a result, the test is largely biased towards overrejection, with rejection signalling that (at least) one of the key assumptions that define the DGP under the null is not supported by the data.

[Insert Table 5 around here]

In contrast, the results from the alternative time- and frequency-domain tests suggest

strong evidence of fractional integration, with the hypotheses of $FI(0)$ or $FI(1)$ dynamics being largely rejected in all cases. Excluding MERVAL, all confidence intervals include $d = 0.4$, considered as the "characteristic" value of the long-memory parameter in empirical studies involving daily transformations of squared returns or realized volatility; see, for instance, Bollerslev and Wright (2000) and Andersen *et al.* (2001, 2003). For summarizing purposes, the column labelled *Intersection* in Table 5 shows the set of values for which the null hypothesis cannot be rejected by any of the regression-based tests, *i.e.*, a "core" set of values for which there is methodological agreement. This region is essentially formed by values smaller than the cut-off level $d < 0.5$, essentially suggesting that log absolute returns are driven by a stationary long-memory process. The results from local Whittle estimation agree with this evidence.

Particularizing on M estimates, there are minor differences attending to the choice of the weighting function or the iterative algorithm. Confidence intervals are not markedly different from those based on LS estimation, which is not surprising since the empirical analysis involves a fairly large sample and LS-based inference is consistent and should not be affected critically by the distribution of the data provided regularity conditions. As discussed previously, however, LS is not efficient in a non-Gaussian context, and it is worth noting that the amplitudes of the M confidence intervals tend to be smaller than their LS counterparts in all cases. As shown in the Monte Carlo section, this evidence is consistent with relative gains in power performance. Similarly, M confidence intervals are smaller than their QR counterpart, which, again, agrees with the finite-sample performance exhibitted by M and QR tests under heavy-tailed innovations in the Monte Carlo analysis.

In order to give a sense of the relative size of these reductions, the last two columns in Table 5 report the average relative change in the amplitude of the confidence intervals when moving from LS to M estimation with IRLS and NR algorithms, respectively.[4] Relative changes are determined as $RC := (A_{\mathcal{M}} - A_{LS})/A_{LS}$, with $A_{\mathcal{M}}$ and A_{LS} denoting the amplitude of the M- and LS-based confidence interval, respectively. For ease of presentation, we report the average values of RC based on the estimates from the two different weighting functions. For instance, for the log absolute return of the SP500, the

[4]Note that this is an intuitive way to appraise the differences in the estimation. We do not pursue a formal, statistical analysis to determine if, for instance, reductions are statistically significant.

LS confidence interval is $[0.37, 0.53]$, whereas the IRLS_{H} and IRLS_{B} yield confidence intervals of $[0.38, 0.49]$ and $[0.37, 0.49]$, respectively. Hence, IRLS estimation leads to an average relative change (ARC) in the amplitude of the LS interval of -28.13%. Table 5 shows that the ARC tend to be larger for IRLS and somewhat more conservative for NR, with values ranging, respectively, from -25.00% and -15.63% (MERVAL) to -50.00% and -38.24% (NASDAQ). The overall cross-sectionally averaged values are -37.66% and -24.16%, respectively.

5.2 Log-price range volatility

Under the assumptions that the asset price P_s follows a driftless geometric Brownian motion, Parkinson (1980) shows that the variance estimator $\widehat{\sigma}_t^2 := \kappa_1 (H_t - L_t)^2$ is about five times more efficient than the squared return over the same interval, with $\kappa_1 := \frac{1}{4 \ln 2}$ and $H_t := \max_{s \in [t-\Delta, t]} \ln P_s$ and $L_t := \min_{s \in [t-\Delta, t]} \ln P_s$ denoting the log-transform of the high and low prices over the interval $[t - \Delta, t]$; see also Andersen and Bollerslev (1998). This property holds independently of the discrete amplitude Δ (Martens and van Dijk, 2007), but the typical interval considered in literature is the trading day. Alizadeh *et al.* (2002) discuss the statistical properties of the daily log-range volatility estimator $\mathcal{L}_t := \ln (H_t - L_t)$, arguing that this is approximately distributed as a normal distribution. Since $\ln \widehat{\sigma}_t = 0.5 \ln \kappa_1 + \mathcal{L}_t$, $\mathcal{L}_t$ is a noisy linear proxy of the log-volatility process of returns; see also Brand and Jones (2006). Consequently, the evidence based on this variable can be compared directly to that obtained for the log absolute-valued returns.

[Insert Figure 2 around here]

Paralleling the analysis on log absolute returns, Figure 2 shows the dynamics of the log-range volatility of the SP500 index and the related histogram and sample autocorrelation function. The most striking differences are that persistence is now much more evident, with autocorrelations exhibiting a stronger pattern of dependence (*i.e.*, a greater degree of fractional integration), and that the distribution of the log-range volatility estimator is close to be normal in most cases; see Table 4, Panel B. For instance, the Jarque-Bera test is unable to reject normality for the variability of the IBEX index. Nevertheless, Panel B in Table 4 still shows significant departures in terms of skewness and excess kurtosis, leading to rejections of normality in most cases. For instance, kurtosis in the

log-range of NIKKEI is nearly 5.2, not very different from the value reported for the log absolute return. In a context characterized by nearly-Gaussianity and large sample sizes, the performance of LS- and M-based tests is expected to be similar. Neverthess, M tests may still produce more efficient results, reflected in narrower confidence intervals, when departures from normality are larger.

[Insert Table 6 around here]

Table 6 reports the confidence interval estimates for the log-range series. Consistent with the results reported for log absolute returns, all regression-based tests and the ELW estimator strongly reject $FI(0)$ or $FI(1)$ dynamics. On the other hand, the DV test fails to reject *FI(0)* dynamics in most cases because the confidence intervals are so wide that they include the origin. For similar reasons, the DV cannot reject that the volatility of the HANG SENG index is driven by a unit root process. As shown in the Monte Carlo section, the DV test may suffer important finite-sample undersizing effects when dealing with short-run dependence and, like other sign-based tests, tends to exhibit low power. Finite-sample underrejections, therefore, seem a likely explanation to understand the unusual performance of this test in relation to the alternartive approaches.

The conclusions drawn from the results of the other test lead to consistent picture. In contrast to the results reported in the previous section, confidence intervals now tend to include values above and below the cut-off limit of 0.5. Consequently, there is more uncertainty about the non-stationary of many of these series and, in some cases, such as DAX, stationarity is rejected by all tests. In our view, this evidence is reasonable because the total sample analyzed comprises the 2007-2009 financial crisis, a period of considerable distress in financial markets in which market volatility increased to unprecedented levels. This is a form of nonstationarity which could be accommodated by a fractional integrated model with a long-memory coefficient close to or larger than 0.5, as reflected in the estimates.

Particularizing on the outcomes from M estimation, once again the results are not critically affected by the choices of the weighting ρ-function nor the iterative algorithm implemented. For instance, for the volatility of SP500, IRLS_{H} and IRLS_{B} yield 95% confidence intervals of $[0.49, 0.66]$ and $[0.48, 0.66]$, respectively, while NR_{H} and NR_{B} yield $[0.47, 0.68]$ and [0.47.0.67]. As expected in view of the quasi-Gaussian nature of log-ranges, M estimates are not very different from LS estimates; e.g., the LS confidence interval in

the SP500 case is $[0.50, 0.69]$. The ARC, shown in the last columns of Table 6, suggest more moderate changes which tend to be larger for the IRLS algorithm, ranging from -3.57% (NASDAQ) to -31.03% (BOVESPA). The overall cross-sectional mean value of ARC for IRLS is -18.09%, which is about 50% smaller than the corresponding value in log absolute returs. The difficulties to improve LS estimates in this context are more evident under the NR algorithm. The ARC exhibits noisier behavior around zero, ranging from -15.52% (BOVESPA) to 14.29% (NASDAQ). The overall cross-sectional value is 0.1%, suggesting that, on average, the parameter uncertainty embedded in this procedure is similar to that in LS.

Nevertheless, this analysis shows that M estimation may yield sizeable reductions in the amplitude of the confidence intervals. In particular, the largest values of the ARC when implementing the IRLS and NR algorithms correspond to BOVESPA, with relative changes of -31.03% and -15.52%, respectively, and NIKKEI, with relative gains of -28.57% and -14.29%, respectively. In fact, for these series, M estimation leads to smaller confidence intervals estimaton than any of the other procedures considered. As reported in Table 4, Panel B, it is precisely these series that present the largest combined skewness-kurtosis departures of normality, leading to the largest values of the Jarque-Bera test statistics in the sample. This evidence, therefore, suggests that M estimators can still take advantage of non-normality and succesfully refine the outcomes from LS estimation.

5.3 Discussion

Whereas the results based on both volatility measures essentially agree that the dynamics of this latent process is driven by fractional integration, the estimates for the log absolute returns tend to be significantly smaller than those based on the log-range series, suggesting smaller persistence. These differences may partially be related to differences in the distribution of the data, but they more likely stem from the negative influence of measurement errors on log absolute returns, causing a well-known problem of attenuation bias when inferring the order of fractional integration.

To see this, note that discrete returns can generally be written as $r_t = \mu_t + \sigma_t \eta_t$, where μ_t and σ_t denote the conditional mean and volatility of the process, and η_t is a short-run innovation component, typically assumed to be i.i.d. in theoretical models. At the daily frequency, μ_t is typically small, so $r_t \simeq \sigma_t \eta_t$ and hence $\ln |r_t| \simeq \ln \sigma_t + \ln |\eta_t|$. As a

result, innovation to returns introduce random measurement errors in log absolute returns. Intuitively, this term weakens the pattern of correlation in $\{\ln|r_t|\}$ because the short-run component $\ln|\eta_t|$ makes this series noisier. This causes seemingly smaller persistence (an effect that is evident through the differences in the autocorrelation functions shown in Figure 3 and 4) and downward-biased estimates of the long-memory coefficient; see, for instance, Haldrup and Nielsen (2007).[5] As discussed by Bollerslev and Wright (2000), the problem can be related to temporal aggregation. Even at a relatively high sampling frequency such as daily, aggregation induces downward biases in the estimates based on the squared, log-squared or absolute returns because the variability of the term $\ln|\eta_t|$ largely increases with temporal aggregation.

The log-range volatility estimator relies on a different form of aggregation, as it essentially builds on the spread of the high and low log-prices over the session. Whereas this difference may still be contaminated with measurement errors, the variability of the long-range volatility estimate is much smaller than that of log absolute returns, which implies that the log-volatility process is more accurately captured, or, in other words, that measurement errors have a much weaker influence.[6] Andersen and Bollerslev (1998) argue that the daily range has approximately the same informational content as sampling intradaily returns every 4 hours; see also Brandt and Diebold (2006). Higher efficiency, comparable to realized variance estimates over such intraday intervals, must necessarily reduce the downward bias in long-memory estimation caused by aggregation, following the arguments in Andersen and Bollerslev (1998). The evidence reported in Table 5 seems entirely consistent with this statement. Consequently, the results from the range-based

[5] This effect has been documented for log-periodogram based estimators, but is extensible to regression-based tests because the properties of the tests are critically determined by the stochastic properties of the regressor. Since mesurament errors would feed into the right-hand side variables, the pattern of correlatio that serves as a basis to identify long-memory is distorted, leading to similar biases as in the case of log-periodogram estimation.

[6] Under the assumption that log stock price follows a martingale process with constant variance, Alizadeh *et al.* (2002) showed that the log range obeys approximately a normal distribution with mean of $0.43 + \ln\sigma_t$ and a variance of 0.29^2, whereas the log absolute return has a mean of $-0.64 + \ln\sigma_t$, and a variance of 1.11^2. As remarked by Brand and Jones (2006), both measures are linear proxies of the log-volatility process (with the same loading of one), but the variability of the latter is about four times larger. Measurement errors, consequently, have a smaller impact on the variability of the variable and the range is a much more informative proxy of the true log-volatility process.

estimation are, in our view, more reliable and likely to capture the long-term behavior of the volatility process.

6 Conclusion

This paper has discussed the theory under M estimation for the class of regression-based tests of fractional integration put forward in Breitung and Hassler (2002), Demetrescu *et al.* (2008) and Hassler *et al.* (2009). These tests are derived under the Lagrange Multiplier principle of a Gaussian score function, which ensures efficiency in the appropriate context of normality. Because normality is not a crucial assumption in order to characterize the asymptotic distribution of these tests, they still are valid when the series of interest display non-normal features. In the empirically relevant context in which variables are not normal distributed, however, alternative estimation procedures, such as M estimation, may exhibit better properties. This paper has discussed the asymptotic null distribution for M tests for fractional integration, showing that they retain the most appealing properties of LS tests, namely, asymptotic normality, which holds independently of the degree of fractional integration, and power to detect (local) departures. Monte Carlo analysis confirms that these tests can be more appropriate under non-Gaussian innovations, largely improving the power exhibitted by LS tests.

The empirical analysis has addressed the existence of long-memory dynamics in some well-known proxies of volatility, characterized by strong persistence in their autocorrelation function, but with very different distributional properties. Log absolute returns is a highly non-normal proxy of volatility, whereas log-range volatility is very well approximated by a normal distribution. The evidence put forward for the latter is more reliable, because log-range is a more efficient proxy of volatililty. In a nearly-Gaussian context, LS-based estimation is expected to produce reliable and (approximately) efficient results, and the evidence obtained from alternative methods essentially matches that of LS estimation. Interestingly, the empirical evidence in our paper suggests that M-based estimators may still be able to introduce finite-sample refinements, leading to more accurate estimates of the long-memory parameter in the form of confidence intervals with smaller amplitude. While these refinements may generally be conservative in a quasi-Gaussian context, in practice they do not imply an incremental computational burden, since iterative meth-

ods used in M estimation essentially refine the preliminary LS-based estimation without involving further optimization or complex operations. Consequently, M estimation seems to provide a valuable complementary alternative to LS.

References

[1] Agostinelli, C. and L. Bisaglia. 2010.ARFIMA Processes and Outliers: a Weighted Likelihood Approach. *Journal of Applied Statistics* 37: 1569-1584.

[2] Alizadeh, S., Brandt, M.W., and F.X. Diebold. 2002. Range-Based Estimation of Stochastic Volatility Models. *Journal of Finance* 57: 1047-1091.

[3] Andersen, T.G. and T. Bollerslev. 1998. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. International Economic Review 39: 885-905.

[4] Andersen, T.G.; Bollerslev, T.; Diebold, F.X. and P. Labys. 2001. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96: 42-55.

[5] Andersen, T.G.; Bollerslev, T.; Diebold, F.X. and P. Labys. 2003. Modeling and Forecasting Realized Volatility. *Econometrica* 71: 529-626.

[6] Baillie, R.T. 1996. Long Memory Processes and Fractional Integration in Econometrics. *Journal of Econometrics* 73: 6-59.

[7] Beran, J. 1994. On a class of M-estimators for Gaussian Long-Memory Models. *Biometrika* 81: 755-766.

[8] Bollerslev, T. and J. Wright. 2000. Semiparametric Estimation of Long-Memory Volatility Dependencies: The role of High-Frequency Data. *Journal of Econometrics* 98: 81-106.

[9] Brandt, M.W., and F.X. Diebold. 2006. A No-Arbitrage Approach to Range-based Estimation of Return Covariances and Correlations. *Journal of Business* 79: 61–74.

[10] Brandt, M.W., and C.S. Jones. 2006. Forecasting With Range-Based EGARCH Models. *Journal of Business and Economic Statistics* 24:470-486.

[11] Breitung, J., and U. Hassler. 2002. Inference on the Cointegration Rank in Fractionally Integrated Processes. *Journal of Econometrics* 110: 167-185.

[12] Delgado, M.A., and C. Velasco. 2005. Sign Tests for Long-Memory Time Series. *Journal of Econometrics* 128: 215-251.

[13] Demetrescu, M., U. Hassler, and V. Kuzin. 2011. Pitfalls of Post-Model-Selection Testing: Experimental Quantification. *Empirical Economics* 40: 359-372.

[14] Demetrescu, M., V. Kuzin, and U. Hassler. 2008. Long Memory Testing in the Time Domain. *Econometric Theory* 24: 176-215.

[15] Georgiev, I.; P.M.M. Rodrigues and A.M.R. Taylor. 2017. Unit Root Tests and Heavy-Tailed Innovations. *Journal of Time Series Analysis*, in press.

[16] Haldrup, N., and M.Ø. Nielsen. 2007. Estimation of Fractional Integration in the Presence of Data Noise. *Computational Statistics & Data Analysis* 51: 3100-3114.

[17] Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel. 1986. *Robust Statistics: the Approach based on Influence Functions*. (Wiley, New York).

[18] Hassler, U., and P. Kokoszka. 2010. Impulse Responses of Fractionally Integrated Processes with Long Memory. *Econometric Theory* 26: 1855-1861.

[19] Hassler, U., P.M.M. Rodrigues, and A. Rubia. 2009. Testing for the General Unit Root Hypothesis in the Time Domain. *Econometric Theory* 25: 1793-1828.

[20] Huber, P.J. 1981. *Robust Statistics*. (Wiley, New York).

[21] Koul, H.L. and D. Surgailis. 1997. Asymptotic Expansion of M-Estimators with Long-Memory Errors. *The Annals of Statistics* 25: 818-850.

[22] Kew, H., and D. Harris. 2009. Heteroskedasticity-Robust Testing for a Fractional Unit Root. *Econometric Theory* 25: 1734-1753.

[23] Kreiss, J.P. 1985. A note on M-estimation in stationary ARMA processes. Statistical Decision 3, 317-336.

[24] Lehmann, E.L. 1983. *Theory of Point Estimation*. (Wiley, New York).

[25] Li, G. and W.K. Li. 2008. Least Absolute Deviation Estimation for Fractionally Integrated Autoregressive Moving Average Time Series Models with Conditional Heteroscedasticity. *Biometrika* 95: 399-414.

[26] Parkinson, M. 1980. The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of Business* 53: 61-65.

[27] Robinson, P.M. 2003. Long-Memory Time Series. In: Robinson, P.M. (Ed.), *Time Series With Long Memory*. Oxford University Press, Oxford, pp. 4-32.

[28] Robinson, P.M. 1994. Efficient Tests of Nonstationary Hypotheses. *Journal of the American Statistical Association* 89: 1420-1437.

[29] Rothenberg, T.J., and J.H. Stock. 1997. Inference in a Nearly Integrated Autoregressive Model With Nonnormal Innovations. *Journal of Econometrics* 80: 269-286.

[30] Shimotsu, K., and P.C.B. Phillips. 2005. Exact Local Whittle Estimation of Fractional Integration. *Annals of Statistics* 33: 1890-1933.

[31] Tanaka, K. 1999. The Nonstationary Fractional Unit Root. *Econometric Theory* 15: 549-582.

[32] Tolvi, J. 2003. Long Memory and Outliers in Stock Market Returns. *Applied Financial Economics*:13: 495-502.

[33] Welsh A.H., and E. Ronchetti. 2002. A journey in single steps: robust one-step M-estimation in linear regression. *Journal of Statistical Planning and Inference*, 103, 287–310.

[34] Wooldridge, J.M. 1994. Estimation and inference for dependent processes, in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics vol. IV*, (North-Holland, Amsterdam). P. 2639-2738.

Appendix A: M estimation and the LM principle

For ease of exposition, assume that Assumption 1 holds with $\varepsilon_t \sim iid\left(0,\sigma^2\right)$, having a differentiable, continuos probability density function $f_\varepsilon\left(r\right)$. Let $\Delta_+^\theta := \left(1-L\right)_+^\theta$, noting that $\varepsilon_t = \Delta_+^\theta \varepsilon_{t,d}$. Then, for a parameter α_ρ and a constant $s > 0$, define $\widetilde{\varepsilon}_{t,d,s} := \left(\varepsilon_{t,d} - \alpha_\rho\right)/s$ such that

$$\widetilde{\varepsilon}_{t,d,s} = \Delta_+^\theta \frac{\varepsilon_{t,d}}{s} - \frac{\alpha_\rho}{s}.$$

The log-likelihood considered fo the estimation of $\boldsymbol{\lambda} := \left(\alpha_\rho, \theta\right)'$ conditional on $\left\{\varepsilon_{t,d}; s\right\}$ can be written as $\ell\left(\boldsymbol{\lambda}\right) := \sum_{t=1}^T \ln f_\varepsilon\left(\widetilde{\varepsilon}_{t,d,s}\right)$. The score vector is given by:

$$\left[\begin{array}{l} \frac{\ell(\boldsymbol{\lambda})}{\partial \alpha_\rho} = -\frac{1}{s}\sum_{t=1}^T \left(\frac{\partial \ln f_\varepsilon(r)}{\partial r}\right)\Big|_{r=\widetilde{\varepsilon}_{t,d,s}} \\ \frac{\ell(\boldsymbol{\lambda})}{\partial \theta} = \sum_{t=1}^T \left(\frac{\partial \ln f_\varepsilon(r)}{\partial r}\right)\Big|_{r=\widetilde{\varepsilon}_{t,d,s}} \Delta_+^\theta \ln\left(1-L\right)\frac{\varepsilon_{t,d}}{s} \end{array}\right]$$

Using the series expansion of the logarithm

$$-\ln\left(1-L\right) = L + \frac{L^2}{2} + \frac{L^3}{3}\ldots,$$

and setting $\psi_{ML}(r) := -\partial \ln f_\varepsilon(r)/\partial r$, the score vector can equivalently be written as:

$$\frac{1}{s}\left(-\sum_{t=1}^T \psi_{ML}(\widetilde{\varepsilon}_{t,d,s}), \sum_{t=1}^T \psi_{ML}\left(\widetilde{\varepsilon}_{t,d,s}\right) x_{t-1,d}^*\right)'.$$

Under $\mathrm{H}_0 : \theta = 0$, we have $\varepsilon_{t,d} := \varepsilon_t$ and $\widetilde{\varepsilon}_{t,d,\widehat{\sigma}} = \widetilde{\varepsilon}_{t,\sigma} + o_p\left(1\right)$ under Assumption 3, with $\widetilde{\varepsilon}_{t,\sigma} := \left(\varepsilon_t - \alpha_\rho\right)/\sigma$. Then, the ML first-order condition $\ell\left(\boldsymbol{\lambda}\right)/\partial\boldsymbol{\lambda}' = 0$ with $s = \widehat{\sigma}$ leads to the vector equation sytem:

$$\sum_{t=1}^T \psi_{ML}\left(\frac{\varepsilon_t - \alpha_\rho}{\widehat{\sigma}}\right)\mathbf{x}_{t-1,d}^* = 0; \quad \mathbf{x}_{t-1,d}^* := \left(1, x_{t-1,d}^*\right)'; \tag{16}$$

which, as discussed in Section 3.3.1, would characterize M estimation in the regression of $\varepsilon_{t,d}$ on $x_{t-1,d}^*$ and a constant. More generally, consider an arbitrary ψ-function in (16), and let $\widehat{\boldsymbol{\beta}}_M$ be the solution of $\sum_{t=1}^T \psi\left(u_{t,\widehat{\sigma}}^*\left(\boldsymbol{\beta}\right)\right)\mathbf{x}_{t-1,d}^* = 0$, with $u_{t,\widehat{\sigma}}^*\left(\boldsymbol{\beta}\right) = \left(\varepsilon_{t,d} - \boldsymbol{\beta}'\mathbf{x}_{t-1,d}^*\right)/\widehat{\sigma}$. When $\psi = \psi_{ML}$, $\widehat{\boldsymbol{\beta}}_M$ matches the ML solution under $\mathrm{H}_0 : \theta = 0$ and subsequent inference has optimal asymptotic properties under regularity conditions, but this typically requires knowledge of f_ε. On the other hand, under suitable conditions that do not impose a parametric assumption on f_ε, we typically have $\psi \neq \psi_{ML}$, but $\widehat{\boldsymbol{\beta}}_M$ will still retain consistency and asymptotic normality, so it is possible to design tests for $\mathrm{H}_0 : \theta = 0$ with non-trivial power, as formally proven in Appendix B. The case of dependent errors can be handled similarly, using p-th order augmentation such that $\mathbf{x}_{t-1,d}^* := \left(1, x_{t-1,d}^*, \varepsilon_{t-1,d}, ..., \varepsilon_{t-p,d}\right)'$.

Appendix B: Technical Proofs

In what follows, recall that $x_{t-1,d}^{**} := \sum_{j=0}^{\infty} j^{-1}\varepsilon_{t-j,d}$, $\mathbf{x}_{t-1,d}^{**} = \left(1, x_{t-1,d}^{**}, \varepsilon_{t-1,d}, ..., \varepsilon_{t-p,d}\right)'$, $u_{t,s}^{**}(\boldsymbol{\beta}) := \left(\varepsilon_{t,d} - \boldsymbol{\beta}'\mathbf{x}_{t-1,d}^{**}\right)/s$, $u_{t,s}^{*}(\boldsymbol{\beta}) := \left(\varepsilon_{t,d} - \boldsymbol{\beta}'\mathbf{x}_{t-1,d}^{*}\right)/s$ for any $\boldsymbol{\beta} \in \boldsymbol{\Theta}$ and for any finite $s > 0$, and that $\widetilde{e}_{t,s} := (e_t - \alpha_\rho)/s$.

Lemma A1. *Under the null hypothesis,* $\mathrm{H}_0 : \theta = 0$, *Assumptions 1, 3 and 4, and a first-order Lipschitz continuous* ρ*-function, it follows that* $\arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}} Q_T^{**}(\boldsymbol{\beta}) \xrightarrow{p} \boldsymbol{\beta}_0$, *where* $Q_T^{**}(\boldsymbol{\beta}) := T^{-1}\sum_{t=p+1}^{T} \rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)$, $\beta_0 := (\alpha_\rho, 0, a_1, ..., a_p)'$, *and* $\alpha_\rho := \arg\min_{c\in\boldsymbol{\Theta}_\alpha} E\left(\rho\left(\frac{e_t - c}{\sigma}\right)\right)$.

Proof of Lemma A.1.

The proof takes two steps. In the first step (Step 1) we show that, under $\mathrm{H}_0 : \theta = 0$ and the assumptions considered, the random function $Q_T^{**}(\boldsymbol{\beta})$ converges uniformly in probability to $E\left(\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)\right)$ on Θ. In the second step (Step 2), we show that $E\left(\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)\right)$ is uniquely minimized at $\boldsymbol{\beta}_0$ on Θ under these restrictions. Because Θ is compact and the continuity of ρ ensures that the measurable functions $Q_T^{**}(\boldsymbol{\beta})$ and $E\left(\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)\right)$ are continuous in $\boldsymbol{\beta}$, conditions (A) to (C) in Thm. 4.1.1 in Amemiya (1985) are verified and, hence, $\arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}} Q_T^{**}(\boldsymbol{\beta}) \xrightarrow{p} \arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}} E\left(\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)\right)$.

Proof of Step 1.

Under $\mathrm{H}_0 : \theta = 0$ and Assumption 1, $\left(\varepsilon_{t,d}, \mathbf{x}_{t-1,d}'^{**}\right)'$ is a strictly stationary and ergodic vector, and so are the measurable transformations $u_{t,\sigma}^{**}(\boldsymbol{\beta})$ and $\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)$ for all $\boldsymbol{\beta} \in \Theta$ and $\sigma > 0$. Since $E|\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)| < \infty$, the Ergodic Theorem (ET) ensures that $Q_T^{**}(\boldsymbol{\beta}) \xrightarrow{a.s} E\left(\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)\right)$ for any fixed $\boldsymbol{\beta} \in \Theta$, and hence $Q_T^{**}(\boldsymbol{\beta})$ converges pointwise to $E\left(\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)\right)$ in Θ. A sufficient condition for uniform convergence in probability (Newey 1989; Andrews 1989; and Pötscher and Prucha 1994) is that $Q_T^{**}(\boldsymbol{\beta})$ be stochastically equicontinuous, *i.e.*, if for any $\epsilon > 0$,

$$\lim_{\delta\to 0}\lim_{T\to\infty} \sup_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2:\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\|<\delta} \Pr\left(|Q_T^{**}(\boldsymbol{\beta}_1) - Q_T^{**}(\boldsymbol{\beta}_2)| > \epsilon\right) = 0$$

which follows if

$$\lim_{\delta\to 0}\lim_{T\to\infty} \sup_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2:\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\|<\delta} |Q_T^{**}(\boldsymbol{\beta}_1) - Q_T^{**}(\boldsymbol{\beta}_2)| \xrightarrow{p} 0. \tag{17}$$

Recall that $f(x) : \mathbb{R} \to \mathbb{R}$ is said to be first-order Lipschitz continuous if $|f(c_1) - f(c_2)| \leq$

$C|c_1 - c_2|$ for some finite $C > 0$ and all $c_1, c_2 \in \mathbb{R}$. Therefore,

$$\begin{aligned}\sup_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2:\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\|<\delta} |Q_T^{**}(\boldsymbol{\beta}_1) - Q_T^{**}(\boldsymbol{\beta}_2)| &\leq T^{-1}\sum_{t=p+1}^{T} |\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta}_1)\right) - \rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta}_2)\right)| \\ &\leq \frac{C\delta}{\sigma}\left(T^{-1}\sum_{t=p+1}^{T}\left\|\mathbf{x}_{t-1,d}^{**}\right\|\right) = o_p(1)\end{aligned}$$

as $T \to \infty$ and $\delta \to 0$, irrespectively of $\boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_2$, noting that $T^{-1}\sum_{t=p+1}^{T}\left\|\mathbf{x}_{t-1,d}^{**}\right\| = O_p(1)$ from Chebyshev's inequality because $E||\mathbf{x}_{t-1,d}^{*}||^2 < \infty$ under Assumption 1. Consequently, $\sup_{\boldsymbol{\beta}\in\Theta}\left|Q_T^{**}(\boldsymbol{\beta}) - E\left[\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)\right]\right| = o_p(1)$ as required.

Proof of Step 2.

Consider the partitions $\boldsymbol{\beta} = (\alpha, \boldsymbol{\kappa}')'$, $\boldsymbol{\beta}_0 = (\alpha_p, \boldsymbol{\kappa}_0')'$, and $\mathbf{x}_{t-1,d}^{**} = \left(1, \boldsymbol{z}_{t-1,d}'^{**}\right)'$, with these terms defined implicitly. Under the null hypothesis and Assumptions 1 to 3, the characterization $\varepsilon_{t,d} = \boldsymbol{\kappa}_0'\boldsymbol{z}_{t-1,d}^{**} + e_t$ holds true and we have

$$\begin{aligned}&\arg\min_{\boldsymbol{\beta}\in\Theta} E\left(\rho\left(\frac{\varepsilon_{t,d} - \boldsymbol{\beta}'\mathbf{x}_{t-1,d}^{**}}{\sigma}\right)\right) \\ &= \arg\min_{\boldsymbol{\beta}\in\Theta} E\left(\rho\left(\frac{e_t}{\sigma} - (\boldsymbol{\kappa}-\boldsymbol{\kappa}_0)'\frac{\boldsymbol{z}_{t-1,d}^{**}}{\sigma} - \frac{\alpha}{\sigma}\right)\right) \\ &= \arg\min_{\boldsymbol{\beta}\in\Theta} E\left(\rho\left(\frac{e_t - \alpha_p}{\sigma} - \left[(\boldsymbol{\kappa}-\boldsymbol{\kappa}_0)'\frac{\boldsymbol{z}_{t-1,d}^{**}}{\sigma} + \frac{(\alpha-\alpha_p)}{\sigma}\right]\right)\right) \\ &= \arg\min_{\boldsymbol{\beta}\in\Theta} E\left(\rho\left(\frac{e_t - \alpha_p}{\sigma} - (\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\frac{\mathbf{x}_{t-1,d}^{**}}{\sigma}\right)\right). \qquad (18)\end{aligned}$$

The orthogonality restriction $E\left(\psi\left(\frac{e_t-\alpha_\rho}{\sigma}\right)\middle|\mathcal{F}_{t-1}\right) = 0$ and the condition $\mathrm{E}\left(\psi'\left(\frac{e_t-\alpha_\rho}{\sigma}\right)\middle|\mathcal{F}_{t-1}\right) > 0$ in Assumption 2 *iv*) imply that $\min_\lambda E\left(\rho\left(\frac{e_t-\alpha_p}{\sigma} - \lambda\right)\middle|\lambda\right)$ for adapted λ is uniquely solved at $\lambda = 0$. Hence the Law of Iterated Expectations (LIE) show that $E\left[\rho\left(u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)\right]$ attains its global minimum at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. ■

Lemma A2. *Let* $R_T(\boldsymbol{\beta}) := T^{-1}\sum_{t=p+1}^{T}\psi(\xi_t)\left(u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta}) - u_{t,\sigma}^{**}(\boldsymbol{\beta})\right)$, *with* ξ_t *lying between* $u_{t,\sigma}^{**}(\boldsymbol{\beta})$ *and* $u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta})$ *for each* t. *Under the null hypothesis* $\mathrm{H}_0 : \theta = 0$, *and Assumptions 1 to 4 holding true,* $R_T(\boldsymbol{\beta}) \xrightarrow{p} 0$ *uniformly on* Θ.

Proof of Lemma A2.

For any arbitrary, finite constant $K > 0$,

$$\begin{aligned}\sup_{||\boldsymbol{\beta}||\leq K}|R_T(\boldsymbol{\beta})| &\leq \sup_{||\boldsymbol{\beta}||\leq K}\left|\frac{1}{T}\sum_{t=p+1}^{T}\psi(\xi_t)\ \varepsilon_{t,d}\ b_\sigma\right| + \sup_{||\boldsymbol{\beta}||\leq K}\left|\frac{1}{T}\sum_{t=p+1}^{T}\psi(\xi_t)\ \boldsymbol{\beta}'\mathbf{x}_{t-1,d}^{*}\ b_\sigma\right| \\ &\quad + \sup_{||\boldsymbol{\beta}||\leq K}\left|\frac{1}{\sigma T}\sum_{t=p+1}^{T}\psi(\xi_t)\,\boldsymbol{\beta}'\left(\mathbf{x}_{t-1,d}^{**} - \mathbf{x}_{t-1,d}^{*}\right)\right| \qquad (19)\end{aligned}$$

where $R_T(\boldsymbol{\beta}) := T^{-1}\sum_{t=p+1}^{T}\psi(\xi_t)\left(u^*_{t,\widehat{\sigma}}(\boldsymbol{\beta}) - u^{**}_{t,\sigma}(\boldsymbol{\beta})\right)$ and $b_\sigma := \widehat{\sigma}^{-1} - \sigma^{-1}$. Hence, we prove the required result by showing that all terms on the right-hand side of (19) are $o_p(1)$. For the first term, note that $\psi(r)$ is bounded and $|b_\sigma| = o_p(1)$ under Assumptions 2 and 3, respectively, so there exists some $C > 0$ such that $\sup|\psi(r)| \leq C$. Setting $M := \max[C, K]$, it follows that under $\mathrm{H}_0 : \theta = 0$ and Assumptions 1 to 3,

$$\sup_{||\boldsymbol{\beta}||\leq K}\left|\frac{1}{T}\sum_{t=p+1}^{T}\psi(\xi_t)\ \varepsilon_{t,d}\ b_\sigma\right| \leq M\ |b_\sigma|\left(T^{-1}\sum_{t=p+1}^{T}|\varepsilon_{t,d}|\right) = o_p(1)$$

For the second term, it follows similarly that

$$\sup_{||\boldsymbol{\beta}||\leq K}\left|\frac{1}{T}\sum_{t=p+1}^{T}\psi(\xi_t)\ \boldsymbol{\beta}'\mathbf{x}^*_{t-1,d}\ b_\sigma\right| \leq\ M^2\ |b_\sigma|\left(\frac{1}{T}\sum_{t=p+1}^{T}\left\|\mathbf{x}^*_{t-1,d}\right\|\right) = o_p(1)$$

because $E\left(\varepsilon^2_{t,d}\right) < \infty$ and $E||\mathbf{x}^*_{t-1,d}||^2 < \infty$ for all $\boldsymbol{\beta} \in \Theta$, so $T^{-1}\sum_{t=p+1}^{T}\left\|\mathbf{x}^*_{t-1,d}\right\| = O_p(1)$ from Chebyshev's inequality. Finally,

$$\sup_{||\boldsymbol{\beta}||\leq K}\left|\frac{1}{\sigma T}\sum_{t=p+1}^{T}\psi(\xi_t)\,\boldsymbol{\beta}'\left(\mathbf{x}^{**}_{t-1,d} - \mathbf{x}^*_{t-1,d}\right)\right| \leq \frac{M^2}{T}\sum_{t=p+1}^{T}\left\|\mathbf{x}^{**}_{t-1,d} - \mathbf{x}^*_{t-1,d}\right\|,$$

since under $\mathrm{H}_0 : \theta = 0$ and Assumption 1 (see e.g. Demetrescu *et al.*, 2008) $E\left\|\mathbf{x}^{**}_{t-1,d} - \mathbf{x}^*_{t-1,d}\right\| = \left\|\sum_{j=t}^{\infty} j^{-1}\varepsilon_{t-j,d}\right\| = O\left(1/\sqrt{t}\right)$, this implies that $T^{-1}\sum_{t=p+1}^{T}\left\|\boldsymbol{x}^{**}_{t-1,d} - \boldsymbol{x}^*_{t-1,d}\right\| = O_p\left(T^{-1/2}\right)$. Consequently, for any finite, arbitrary $K > 0$, we have that $\sup_{||\boldsymbol{\beta}||\leq K}|R_T(\boldsymbol{\beta})| = o_p(1)$, and thus $\sup_{\boldsymbol{\beta}\in\Theta}|R_T(\boldsymbol{\beta})| = o_p(1)$ under Assumption 4. ■

Lemma A3. *Let* $\mathbf{A}_{\boldsymbol{\beta}} := E\left(\psi'(\widetilde{e}_{t,\sigma})\,\boldsymbol{x}^{**}_{t-1,d}\boldsymbol{x}'^{**}_{t-1,d}\right)$, *and* $\mathbf{B}_{\boldsymbol{\beta}} := E\left(\psi^2(\widetilde{e}_{t,\sigma})\,\boldsymbol{x}^{**}_{t-1,d}\boldsymbol{x}'^{**}_{t-1,d}\right)$. *Then, under the null hypothesis* $\mathrm{H}_0 : \theta = 0$ *and Assumptions 1-4, the following results hold:*

i) $||\mathbf{A}_{\boldsymbol{\beta}}||$ *is bounded, bounded away from zero, and* $det(\mathbf{A}_{\boldsymbol{\beta}}) > 0$;

ii) $T^{-1}\sum_{t=p+1}^{T}\psi'(\widetilde{e}_{t,\sigma})\,\mathbf{x}^{**}_{t-1,d}\mathbf{x}'^{**}_{t-1,d} \xrightarrow{p} \mathbf{A}_{\boldsymbol{\beta}}$;

iii) $T^{-1/2}\sum_{t=p+1}^{T}\psi(\widetilde{e}_{t,\sigma})\,\mathbf{x}^{**}_{t-1,d} \Rightarrow \mathcal{N}(0, \mathbf{B}_{\boldsymbol{\beta}})$.

Proof of Lemma A3.

To prove *i*), we need an upper bound on the expectation, which follows from the finite variance of $\mathbf{x}^{**}_{t-1,d}$ and the fact that ψ' is bounded under Assumption 1 *ii*) and *iii*). Next, let $\mathbf{a} \in \mathbb{R}^{p+2} \neq \mathbf{0}$ and note that

$$\mathbf{a}'E\left(\psi'(\widetilde{e}_{t,\sigma})\,\mathbf{x}^{**}_{t-1,d}\mathbf{x}'^{**}_{t-1,d}\right)\mathbf{a} = E\left(\left(\mathbf{a}'\mathbf{x}^{**}_{t-1,d}\right)^2 E\left(\psi'(\widetilde{e}_{t,\sigma})|\,\mathcal{F}_{t-1}\right)\right).$$

Since $E\left(\psi'\left(\widetilde{e}_{t,\sigma}\right)\middle|\mathcal{F}_{t-1}\right) > 0$ a.s., the expectation on the right-hand side is positive because $\mathbf{a}'\mathbf{x}^{**}_{t-1,d}$ is nondegenerate, so $\mathbf{A}_{\boldsymbol{\beta}}$ is positive definite. Part *ii)* follows from the ET because $\left\{\psi'\left(\widetilde{e}_{t,\sigma}\right), \mathbf{x}'^{**}_{t-1,d}\right\}$ is strictly stationary and ergodic with finite expectation. For *iii)*, consider $\boldsymbol{\varsigma}_t := \psi\left(\widetilde{e}_{t,\sigma}\right)\mathbf{x}^{**}_{t-1,d}$ and note that $E\left(\boldsymbol{\varsigma}_t|\mathcal{F}_{t-1}\right) = 0$ and $E\left(|\boldsymbol{\varsigma}_t|\right) < \infty$ under $\mathrm{H}_0 : \theta = 0$ and Assumptions 1 and 2, so that $\{\boldsymbol{\varsigma}_t, \mathcal{F}_t\}$ is a strictly stationary, ergodic L_2-bounded MDS vector with $E\left(\boldsymbol{\varsigma}_t\boldsymbol{\varsigma}_t'\right) = \mathbf{B}_{\boldsymbol{\beta}}$, where it can be shown that $\mathbf{B}_{\boldsymbol{\beta}}$ is bounded and bounded away from zero using similar arguments as in *i)* above. The required result then follows from the Central Limit Theorem (CLT) for MDS (Davidson, 1994, Theorem 24.3) and the Cramér-Wold device. ■

Lemma A4. *Under the null hypothesis* $\mathrm{H}_0 : \theta = 0$ *and Assumptions 1-4 it follows that,*
i) $T^{-1/2}\sum_{t=p+1}^{T}\mathbf{x}^{**}_{t-1,d} = O_p\left(\sqrt{\ln T}\right)$ *and* $T^{-1/2}\sum_{t=p+1}^{T}\mathbf{x}^{*}_{t-1,d} = O_p\left(\sqrt{\ln T}\right)$;
ii) $T^{-1/2}\sum_{t=p+1}^{T}\psi\left(\widetilde{e}_{t,\widehat{\sigma}}\right)\mathbf{x}^{*}_{t-1,d} \Rightarrow \mathcal{N}\left(0, \mathbf{B}_{\boldsymbol{\beta}}\right)$;
iii) $\sup_{p+1\leq t\leq T}||\mathbf{x}^{*}_{t-1,d}|| = o_p\left(T^{3/8}\ln T\right)$.

Proof of Lemma A4.

Part *i)* holds because $\mathbf{x}^{**}_{t-1,d}$ is a linear process with Wold coefficient matrices satisfying $\|\mathbf{B}_j\| \leq Cj^{-1}$ and, hence, the autocovariances satisfy $\|\boldsymbol{\Gamma}_h\| \leq Ch^{-1}$, which implies that $Var\left(T^{-1}\sum_{t=p+1}^{T}\mathbf{x}^{**}_{t-1,d}\right) = O\left(\frac{\log T}{T}\right)$ and, hence, $T^{-1/2}\sum_{t=p+1}^{T}\mathbf{x}^{**}_{t-1,d} = O_p\left(\sqrt{\log T}\right)$ from Markov's Theorem. The same argument applies to $\mathbf{x}^{*}_{t-1,d}$. To show *ii)*, note that

$$\begin{aligned}
T^{-1/2}\sum_{t=p+1}^{T}\psi\left(\widetilde{e}_{t,\widehat{\sigma}}\right)\mathbf{x}^{*}_{t-1,d} &= T^{-1/2}\sum_{t=p+1}^{T}\psi\left(\widetilde{e}_{t,\sigma}\right)\mathbf{x}^{**}_{t-1,d} \\
&\quad + \left\{T^{-1/2}\sum_{t=p+1}^{T}\psi\left(\widetilde{e}_{t,\sigma}\right)\left(\mathbf{x}^{*}_{t-1,d} - \mathbf{x}^{**}_{t-1,d}\right)\right\} \\
&\quad + \left\{T^{-1/2}\sum_{t=p+1}^{T}\left(\psi\left(\widetilde{e}_{t,\widehat{\sigma}}\right) - \psi\left(\widetilde{e}_{t,\sigma}\right)\right)\mathbf{x}^{**}_{t-1,d}\right\} \\
&= T^{-1/2}\sum_{t=p+1}^{T}\psi\left(\widetilde{e}_{t,\sigma}\right)\mathbf{x}^{**}_{t-1,d} + \{\mathbf{R}_{1T}\} + \{\mathbf{R}_{2T}\}, \text{ say,}
\end{aligned}$$

so the required result holds from Lemma A3 *iii)* if $\mathbf{R}_{1T} = o_p(1)$ and $\mathbf{R}_{2T} = o_p(1)$. Since $\psi\left(\widetilde{e}_{t,\sigma}\right)$ is a bounded MDS under Assumption 2, $\mathbf{R}_{1T} = o_p(1)$ because $\psi^2\left(\widetilde{e}_{t,\sigma}\right)\left\|\boldsymbol{x}^{**}_{t-1,d} - \boldsymbol{x}^{*}_{t-1,d}\right\|^2 \xrightarrow{p} 0$ as $t \to \infty$. For the second term, $\mathbf{R}_{2T}$, recall that $b_\sigma := \widehat{\sigma}^{-1} - \sigma^{-1}$ and use the mean-value

expansion $\psi(\widetilde{e}_{t,\widehat{\sigma}}) - \psi(\widetilde{e}_{t,\sigma}) = \psi'(\xi_t)(e_t - \alpha_\rho) b_\sigma$, with ξ_t between $\widetilde{e}_{t,\widehat{\sigma}}$ and $\widetilde{e}_{t,\sigma}$, to write

$$\begin{aligned}\mathbf{R}_{2T} &= \left\{\frac{b_\sigma}{\sqrt{T}}\sum_{t=p+1}^{T}\psi'(\widetilde{e}_{t,\sigma})(e_t-\alpha_\rho)\mathbf{x}^*_{t-1,d}\right\} + \left\{\frac{b_\sigma}{\sqrt{T}}\sum_{t=p+1}^{T}(\psi'(\xi_t)-\psi'(\widetilde{e}_{t,\sigma}))(e_t-\alpha_\rho)\mathbf{x}^*_{t-1,d}\right\} \\ &= \{\mathbf{R}_{21T}\} + \{\mathbf{R}_{22T}\}, \text{ say.}\end{aligned}$$

After some straightforward algebraic manipulations,

$$\begin{aligned}\mathbf{R}_{21T} &= \frac{b_\sigma}{T^{1/2}}\sum_{t=p+1}^{T}\left[\psi'(\widetilde{e}_{t,\sigma})e_t - E\left(\psi'(\widetilde{e}_{t,\sigma})e_t|\mathcal{F}_{t-1}\right)\right]\mathbf{x}^*_{t-1,d} \\ &\quad -\alpha_\rho\frac{b_\sigma}{T^{1/2}}\sum_{t=p+1}^{T}\left[\psi'(\widetilde{e}_{t,\sigma}) - E\left(\psi'(\widetilde{e}_{t,\sigma})|\mathcal{F}_{t-1}\right)\right]\mathbf{x}^*_{t-1,d} \\ &\quad +b_\sigma\left(T^{-1/2}\sum_{t=p+1}^{T}\mathbf{x}^*_{t-1,d}\left[E\left(\psi'(\widetilde{e}_{t,\sigma})e_t|\mathcal{F}_{t-1}\right) - \alpha_\rho E\left(\psi'(\widetilde{e}_{t,\sigma})|\mathcal{F}_{t-1}\right)\right]\right)\end{aligned}$$

noting that $\vartheta_{t,\sigma} := \psi'(\widetilde{e}_{t,\sigma})e_t - E\left(\psi'(\widetilde{e}_{t,\sigma})e_t|\mathcal{F}_{t-1}\right)$ is a MDS. Then, under Assumptions 1 and 2,

$$E\left(\left\|\vartheta_{t,\sigma}\mathbf{x}^*_{t-1,d}\right\|^{1+\tilde{\epsilon}/2}\right) \leq \sqrt{E\left(\left|\vartheta_{t,\sigma}\right|^{8/3+\tilde{\epsilon}}\right)E\left(\left\|\mathbf{x}^*_{t-1,d}\right\|^{8/3+\tilde{\epsilon}}\right)},$$

from the Cauchy-Schwarz inequality, setting $2/3 < \tilde{\epsilon} < \epsilon$ with ϵ in Assumption 1, with $E\left(\left|\vartheta_{t,\sigma}\right|^{8/3+\tilde{\epsilon}}\right) < \infty$ because $\psi'(\cdot)$ is bounded and $\tilde{\epsilon} < \epsilon$, and $E\left(\left\|\mathbf{x}^*_{t-1,d}\right\|^{8/3+\tilde{\epsilon}}\right) = O(\log t)$, as discussed in greater detail in the proof of item *iii)* of this Lemma below. The strong LLN for MDS (Davidson, Theorem 20.11) would allows to show that

$$T^{-3/4}\sum_{t=p+1}^{T}\left[\psi'(\widetilde{e}_{t,\sigma})e_t - E\left(\psi'(\widetilde{e}_{t,\sigma})e_t|\mathcal{F}_{t-1}\right)\right]\mathbf{x}^*_{t-1,d} \xrightarrow{a.s.} 0.$$

The sufficient condition is that $\sum_{t\geq 1}\frac{E\left(\left\|[\psi'(\widetilde{e}_{t,\sigma})e_t - E(\psi'(\widetilde{e}_{t,\sigma})e_t|\mathcal{F}_{t-1})]\mathbf{x}^*_{t-1,d}\right\|^p\right)}{t^{3p/4}} < \infty$ be finite for some $1 \leq p \leq 2$. For $p > 4/3$, the series is indeed converging since $E\left(\left\|\vartheta_{t,\sigma}\mathbf{x}^*_{t-1,d}\right\|^{1+\tilde{\epsilon}/2}\right) = O(\log t)$ for some $\tilde{\epsilon} > 2/3$, and the first term vanishes given that $T^{1/4}b_\sigma = o_p(1)$ according to Assumption 3. The second summand is dealt with analogously, while, for the third summand, we note that the conditional expectations are constant under Assumption 2 and $T^{-1/2}\sum_{t=p+1}^{T}\mathbf{x}^*_{t-1,d} = O_p\left(\sqrt{\log T}\right)$ from *i)* of this Lemma, so the result follows with $b_\sigma = o_p\left(T^{-1/4}\right)$.

Finally, for $\mathbf{R}_{22T}$ we use the Lipschitz property of ψ' together with $|\xi_t - \widetilde{e}_{t,\sigma}| \leq |e_t - \alpha_\rho||b_\sigma|$ to bound

$$|\psi'(\xi_t) - \psi'(\widetilde{e}_{t,\sigma})| \leq C|e_t - \alpha_\rho|\ |b_\sigma|$$

for some $C > 0$ such that

$$\begin{aligned}
||\mathbf{R}_{22T}|| &\leq \frac{C|b_\sigma|^2}{\sqrt{T}} \sum_{t=p+1}^{T} |e_t - \alpha_\rho|^2 \, ||\mathbf{x}^*_{t-1,d}|| \\
&\leq \frac{C|b_\sigma|^2}{\sqrt{T}} \sum_{t=p+1}^{T} e_t^2 ||\mathbf{x}^*_{t-1,d}|| + \frac{C|b_\sigma|^2\alpha_\rho^2}{\sqrt{T}} \sum_{t=p+1}^{T} ||\mathbf{x}^{**}_{t-1,d}|| + \frac{2C|b_\sigma|^2|\alpha_\rho|}{\sqrt{T}} \sum_{t=p+1}^{T} |e_t| \, ||\mathbf{x}^*_{t-1,d}|| \\
&= o_p(1)
\end{aligned}$$

from Assumption 3, noting that $E\left(\left|\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} e_t^2 \left\|\boldsymbol{x}^{**}_{t-1,d}\right\|\right|\right) \leq \sqrt{T}E\left(e_t^2\right) E\left(\left\|\boldsymbol{x}^{**}_{t-1,d}\right\|\right)$ under conditional homoskedasticity from Assumption 1, so consistency of $\widehat{\sigma}$ at any rate higher than $T^{1/4}$ collapses the first term. Clearly, the same condition carries over to the other two components as well.

Moving on to *iii)*, we use Minkowski's norm inequality and the linear representation of $\mathbf{x}^*_{t,d}$ to conclude that

$$\left\|\mathbf{x}^*_{t,d}\right\|_{8/3} \leq \sum_{j=0}^{t-1} \|\mathbf{B}_j\| \, \|e_t\|_{8/3} = O(\ln T)$$

since the Wold coefficient matrices satisfy $\|\mathbf{B}_j\| = O\left(j^{-1}\right)$. Hence, $E\left(\left\|\frac{1}{\ln T}\mathbf{x}^*_{t,d}\right\|^{8/3}\right)$ is uniformly bounded, and thus

$$\sup_{p+1\leq t\leq T} \left\|\frac{1}{\ln T}\mathbf{x}^*_{t,d}\right\| = O_p\left(T^{3/8}\right)$$

as required for the result.■

Lemma A5. *Assume that ξ_t lies on the line segment between $u^*_{t,\widehat{\sigma}}(\boldsymbol{\beta}_1)$ and $u^*_{t,\widehat{\sigma}}(\boldsymbol{\beta}_0)$, with $\boldsymbol{\beta}_1 \in \Phi_T$, $\Phi_T := \{\boldsymbol{\beta} \in \Theta : T^\zeta ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq K\}$, for any finite, arbitrary $K > 0$ and $3/8 < \zeta < 1/2$. Under $\mathrm{H}_0 : \theta = 0$ and Assumptions 1 to 4,*

$$\sup_{\boldsymbol{\beta}_1\in\Phi_T} \left\|T^{-1}\sum_{t=p+1}^{T} \psi'(\xi_t)\, \mathbf{x}^*_{t-1,d}\mathbf{x}'^*_{t-1,d} - \mathbf{A}_{\boldsymbol{\beta}}\right\| = o_p(1).$$

Proof of Lemma A5. Recall that $\widetilde{e}_{t,\sigma} := \sigma^{-1}(e_t - \alpha_\rho)$. Then,

$$\begin{aligned}
\frac{1}{T}\sum_{t=p+1}^{T} \psi'(\xi_t)\, \mathbf{x}^*_{t-1,d}\mathbf{x}'^*_{t-1,d} &= \left\{\frac{1}{T}\sum_{t=p+1}^{T} \psi'(\widetilde{e}_{t,\sigma})\, \mathbf{x}^{**}_{t-1,d}\mathbf{x}'^{**}_{t-1,d}\right\} \\
&\quad + \left\{\frac{1}{T}\sum_{t=p+1}^{T} \psi'(\widetilde{e}_{t,\sigma})\left(\mathbf{x}^*_{t-1,d}\mathbf{x}'^*_{t-1,d} - \mathbf{x}^{**}_{t-1,d}\mathbf{x}'^{**}_{t-1,d}\right)\right\} \\
&\quad + \left\{\frac{1}{T}\sum_{t=p+1}^{T} \left(\psi'(\xi_t) - \psi'(\widetilde{e}_{t,\sigma})\right)\mathbf{x}^*_{t-1,d}\mathbf{x}'^*_{t-1,d}\right\} \\
&= \left\{\mathbf{A}^{**}_{\boldsymbol{\beta}T}\right\} + \left\{\mathbf{R}_{1T}\right\} + \left\{\mathbf{R}_{2T}(\xi_t)\right\}, \text{ say.}
\end{aligned}$$

Neither $\mathbf{A}_{\boldsymbol{\beta}T}^{**}$ nor $\mathbf{R}_{1T}$ depend on $\boldsymbol{\beta}_1$, so

$$\sup_{\boldsymbol{\beta}_1\in\Phi_T}\left\|T^{-1}\sum_{t=p+1}^{T}\psi'\left(\xi_t\right)\mathbf{x}_{t-1,d}^{*}\mathbf{x}_{t-1,d}^{\prime *}-\mathbf{A}_{\boldsymbol{\beta}}\right\|\leq\left\|\mathbf{A}_{\boldsymbol{\beta}T}^{**}-\mathbf{A}_{\boldsymbol{\beta}}\right\|+\left\|\mathbf{R}_{1T}\right\|+\sup_{\boldsymbol{\beta}_1\in\Phi_T}\left\|\mathbf{R}_{2T}\left(\xi_t\right)\right\|$$

and the required result holds by showing the asymptotic negligibility of the three terms on the right-hand side. Lemma A3ii) ensures that $\mathbf{A}_{\boldsymbol{\beta}T}^{**}\xrightarrow{p}\mathbf{A}_{\boldsymbol{\beta}}$, so $\left\|\mathbf{A}_{\boldsymbol{\beta}T}^{**}-\mathbf{A}_{\boldsymbol{\beta}}\right\|=o_p\left(1\right)$. For the second term, note that

$$\begin{aligned}\left\|\mathbf{R}_{1T}\right\| &\leq \sup_{1\leq t\leq T}\left|\psi'\left(\widetilde{e}_{t,\sigma}\right)\right|\frac{1}{T}\sum_{t=p+1}^{T}\left\|\mathbf{x}_{t-1,d}^{*}\mathbf{x}_{t-1,d}^{\prime *}-\mathbf{x}_{t-1,d}^{**}\mathbf{x}_{t-1,d}^{\prime **}\right\|=o_p\left(T^{-1/8}\right)\\ &= o_p\left(1\right)\end{aligned}$$

because $\sup_{\leq t\leq T}\left|\psi'\left(\widetilde{e}_{t,\sigma}\right)\right|=o_p\left(T^{3/8}\right)$ since ψ', being Lipschitz, has at most linear tails at infinity so $\psi'\left(\widetilde{e}_{t,\sigma}\right)$ is uniformly $L_{8/3}$-bounded, and, as discussed in the proof of Lemma A2, $T^{-1}\sum_{t=p+1}^{T}\left\|\mathbf{x}_{t-1,d}^{*}\mathbf{x}_{t-1,d}^{\prime *}-\mathbf{x}_{t-1,d}^{**}\mathbf{x}_{t-1,d}^{\prime **}\right\|=O_p\left(T^{-1/2}\right)$. For the last term, note that

$$\left\|\mathbf{R}_{2T}\left(\xi_t\right)\right\|\leq\frac{1}{T}\sum_{t=p+1}^{T}\left|\psi'\left(\xi_t\right)-\psi'\left(\widetilde{e}_{t,\sigma}\right)\right|\left\|\boldsymbol{x}_{t-1,d}^{*}\right\|^2.$$

The Lipschitz condition on ψ' indicates that $\left|\psi'\left(\xi_t\right)-\psi'\left(\widetilde{e}_{t,\sigma}\right)\right|\leq C\left|\xi_t-\widetilde{e}_{t,\sigma}\right|$, where, for each t, we may represent ξ_t as $\xi_t=\frac{\lambda_t}{\widehat{\sigma}}\left(\varepsilon_{t,d}-\boldsymbol{\beta}_1'\mathbf{x}_{t-1,d}^{*}\right)+\frac{\left(1-\lambda_t\right)}{\widehat{\sigma}}\left(\varepsilon_{t,d}-\boldsymbol{\beta}_0'\mathbf{x}_{t-1,d}^{*}\right)$ for some $\lambda_t\in\left[0,1\right]$, and $1\leq t\leq T$. Since $\varepsilon_{t,d}=\boldsymbol{\beta}_0'\mathbf{x}_{t-1,d}^{*}+\left(e_t-\alpha_\rho\right)$, and recalling that $b_\sigma:=\widehat{\sigma}^{-1}-\sigma^{-1}$, it follows that

$$\begin{aligned}\left|\xi_t-\widetilde{e}_{t,\sigma}\right| &= \left|\frac{\varepsilon_{t,d}}{\widehat{\sigma}}-\frac{\lambda_t}{\widehat{\sigma}}\boldsymbol{\beta}_1'\mathbf{x}_{t-1,d}^{*}-\frac{\left(1-\lambda_t\right)}{\widehat{\sigma}}\boldsymbol{\beta}_0'\mathbf{x}_{t-1,d}^{*}-\widetilde{e}_{t,\sigma}\right|\\ &= \left|-\frac{\lambda_t}{\widehat{\sigma}}\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0\right)\mathbf{x}_{t-1,d}^{*}+\left(e_t-\alpha_\rho\right)b_\sigma\right|\\ &\leq \frac{1}{\widehat{\sigma}}\left\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_0\right\|\left\|\boldsymbol{x}_{t-1,d}^{*}\right\|+\ \left|b_\sigma\right|\left(\left|e_t\right|+\left|\alpha_\rho\right|\right)\end{aligned}$$

so setting $M=\max\left[C,K\right]$ leads to

$$\sup_{\boldsymbol{\beta}_1\in\Phi_T}\left\|\mathbf{R}_{2T}\left(\xi_t\right)\right\|\leq\frac{M^2}{\widehat{\sigma}}\left(\frac{1}{T^{1+\zeta}}\sum_{t=p+1}^{T}\left\|\mathbf{x}_{t-1,d}^{*}\right\|^3\right)+\left|b_\sigma\right|\frac{M}{T}\sum_{t=p+1}^{T}\left(\left|e_t\right|+\left|\alpha_\rho\right|\right)\left\|\mathbf{x}_{t-1,d}^{*}\right\|^2.$$

The first term on the right-hand side of this expression vanishes because $\widehat{\sigma}\xrightarrow{p}\sigma$ and

$$\frac{1}{T^{1+\zeta}}\sum_{t=p+1}^{T}\left\|\mathbf{x}_{t-1,d}^{*}\right\|^3\leq\frac{1}{T^{\zeta}}\sup_{1\leq t\leq T}\left\|\mathbf{x}_{t-1,d}^{*}\right\|\left(T^{-1}\sum_{t=p+1}^{T}\left\|\mathbf{x}_{t-1,d}^{*}\right\|^2\right)=o_p\left(1\right)$$

given that $\sup_{1\leq t\leq T}\left\|\mathbf{x}_{t-1,d}^{*}\right\|=o_p\left(T^{\zeta}\right)$ which follows from Lemma A4iii) and $\zeta>3/8$, and $T^{-1}\sum_{t=p+1}^{T}\left\|\mathbf{x}_{t-1,d}^{*}\right\|^2=O_p\left(1\right)$.

For the second term,

$$T^{-1}\sum_{t=p+1}^{T}|e_t|\left\|\mathbf{x}_{t-1,d}^{*}\right\|^2 \leq \sup_{1\leq t\leq T}|e_t|^{1/3}\frac{1}{T}\sum_{t=p+1}^{T}|e_t|^{2/3}\left\|\mathbf{x}_{t-1,d}^{*}\right\|^2$$

where $\sup_{1\leq t\leq T}|e_t|^{1/3}=O_p\left(\sqrt[3]{T^{3/8}}\right)$ given the uniform $L_{8/3}$-boundedness, while Hölder's inequality gives that $E\left(|e_t|^{2/3}\left\|\mathbf{x}_{t-1,d}^{*}\right\|^2\right) \leq \sqrt[4]{E\left(|e_t|^{8/3}\right)}\sqrt[4/3]{E\left(\left\|\mathbf{x}_{t-1,d}^{*}\right\|^{8/3}\right)}$. Thus, since $\alpha_\rho=O(1)$ under Assumption 4, it follows that

$$|b_\sigma|\sup_{1\leq t\leq T}|e_t|^{1/3}\left(\frac{1}{T}\sum_{t=p+1}^{T}|e_t|^{2/3}\left\|\mathbf{x}_{t-1,d}^{*}\right\|^2\right)=o_p(1)$$

and

$$|b_\sigma|\left(\frac{1}{T}\sum_{t=p+1}^{T}\left\|\mathbf{x}_{t-1,d}^{*}\right\|^2\right)|\alpha_\rho|=o_p(1)$$

for any $\widehat{\sigma}$ fulfilling $\widehat{\sigma}-\sigma=o_p\left(T^{-1/4}\right)$, as required in Assumption 3, so $\sup_{\boldsymbol{\beta}_1\in\Phi_T}\|\mathbf{R}_{3T}(\xi_t)\|=o_p(1)$. ■

Lemma A6. *For any $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\in\Theta$, let*

$$\boldsymbol{\pi}_T(\boldsymbol{\beta}_1\boldsymbol{\beta}_2):=\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\left[\psi\left(u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta}_1)\right)-\psi\left(u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta}_2)\right)\right]\mathbf{x}_{t-1,d}^{*}.$$

Then,

$$\boldsymbol{\pi}_T(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=\sqrt{T}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)\left[\frac{1}{\widehat{\sigma}T}\sum_{t=2}^{T}\psi'(\xi_t)\mathbf{x}_{t-1,d}^{*}\mathbf{x}_{t-1,d}^{\prime *}\right]$$

with ξ_t between $u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta})$ and $u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta}_0)$ for each t.

Proof of Lemma A6.

Since

$$\boldsymbol{\pi}_T(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\left[\psi\left(u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta}_1)\right)-\psi\left(u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta}_2)\right)\right]\mathbf{x}_{t-1,d}^{*}$$

and mean-value expanding $\psi\left(u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta}_1)\right)$ about $u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta}_2)$ yields

$$\sqrt{T}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\left[\frac{1}{\widehat{\sigma}T}\sum_{t=p+1}^{T}\psi'(\xi_t)\mathbf{x}_{t-1,d}^{*}\mathbf{x}_{t-1,d}^{\prime *}\right]$$

with ξ_t between $u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta}_1)$ and $u_{t,\widehat{\sigma}}^{*}(\boldsymbol{\beta}_2)$. ■

Proof of Theorem 1.

Noting that conditions i) and ii) in Assumption 2 imply that the ρ-function is Lipschitz, the proof follows directly from Lemmas A1, A2 and the Asymptotic Equivalence Lemma (AEL). Using the Mean Value Theorem (MVT), we can write $Q_T^*(\boldsymbol{\beta}) = Q_T^{**}(\boldsymbol{\beta}) + R_T(\boldsymbol{\beta})$, with $Q_T^{**}(\boldsymbol{\beta})$ from Lemma A1 and $R_T(\boldsymbol{\beta})$ from Lemma A2. Consequently, under $\mathrm{H}_0 : \theta = 0$ and the set of assumptions considered, Lemma A2 ensures that

$$\sup_{\boldsymbol{\beta}\in\boldsymbol{\Theta}} ||Q_T^*(\boldsymbol{\beta}) - Q_T^{**}(\boldsymbol{\beta})|| = o_p(1)$$

so that $Q_T^*(\boldsymbol{\beta}) = Q_T^{**}(\boldsymbol{\beta}) + o_p(1)$ uniformly on Θ. The required result then follows from Lemma A1 and the AEL. ■

Proof of Theorem 2.

Let $\nabla Q_T^*(\boldsymbol{\beta})$ be the gradient of $Q_T^*(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, i.e.,

$$\nabla Q_T^*(\boldsymbol{\beta}) := -T^{-1}\sum_{t=p+1}^{T}\psi\left(u_{t,\widehat{\sigma}}^*(\boldsymbol{\beta})\right)\frac{\boldsymbol{x}_{t-1,d}^*}{\widehat{\sigma}} \tag{20}$$

and note that, since the ρ-function is twice continuous differentiable, the mean-value expansion of $\psi\left(u_{t,\widehat{\sigma}}^*(\boldsymbol{\beta}_0)\right)$ about $u_{t,\widehat{\sigma}}^*\left(\widehat{\boldsymbol{\beta}}_M\right)$ yields,

$$\nabla Q_T^*(\boldsymbol{\beta}_0) = \nabla Q_T^*\left(\widehat{\boldsymbol{\beta}}_M\right) - \left(T^{-1}\sum_{t=p+1}^{T}\psi'(\xi_t)\frac{\boldsymbol{x}_{t-1,d}^*\boldsymbol{x}_{t-1,d}'^*}{\widehat{\sigma}^2}\right)\left(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_0\right),$$

with ξ_t lying between $u_{t,\widehat{\sigma}}^*(\boldsymbol{\beta}_0)$ and $u_{t,\widehat{\sigma}}^*\left(\widehat{\boldsymbol{\beta}}_M\right)$ for each t. From the continuity of the objective function and compactness of Φ_T, then $\widehat{\boldsymbol{\beta}}_M \in \Phi_T$, and since Lemma A5 makes clear that for any $\boldsymbol{\beta} \in \Phi_T$

$$\sup_{\boldsymbol{\beta}\in\Phi_T}\left\|T^{-1}\sum_{t=p+1}^{T}\psi'(\xi_t)\,\boldsymbol{x}_{t-1,d}^*\boldsymbol{x}_{t-1,d}'^* - \mathbf{A}_{\boldsymbol{\beta}}\right\| = o_p(1)$$

where $\mathbf{A}_{\boldsymbol{\beta}}$ is invertible by Lemma A3 i), the following representation holds:

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_0\right) = -\sigma\mathbf{A}_{\boldsymbol{\beta}}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(\frac{e_t-\alpha_\rho}{\widehat{\sigma}}\right)\boldsymbol{x}_{t-1,d}^*\right] + o_p(1)$$

noting that $\nabla Q_T\left(\widehat{\boldsymbol{\beta}}_M\right) = 0$ and $\widehat{\sigma} \xrightarrow{p} \sigma$. Then, from Lemma A4$ii$) it follows that,

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(\frac{e_t-\alpha_\rho}{\widehat{\sigma}}\right)\boldsymbol{x}_{t-1,d}^* \Rightarrow \mathcal{N}(0,\mathbf{B}_{\boldsymbol{\beta}})$$

and, consequently,

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_0\right) \Rightarrow \mathcal{N}\left(0;\sigma^2\mathbf{A}_{\boldsymbol{\beta}}^{-1}\mathbf{B}_{\boldsymbol{\beta}}\mathbf{A}_{\boldsymbol{\beta}}^{-1}\right),$$

as required.■

Proof of Theorem 3.

Note that, given the differentiability of the ρ-function with bounded derivative, $\sup_{\alpha\in\Theta_a} E\left(|\rho\left(\frac{e_t-\alpha}{\sigma}\right)|\right) < \infty$, with Θ_a being a closed interval in $\mathbb{R}$. Define now $\widetilde{e}_{t,\sigma} := \frac{e_t-\alpha}{\sigma}$, $P_T(\alpha) := T^{-1}\sum_{t=p+1}^{T}\rho\left(\frac{e_t-\alpha}{\sigma}\right)$, and $\widehat{P}_T(\alpha) := T^{-1}\sum_{t=p+1}^{T}\rho\left(\frac{\widehat{e}_{(0)t}-\alpha}{\widehat{\sigma}_{(0)}}\right)$. Using the MVT, we may characterize

$$\widehat{P}_T(\alpha) = P_T(\alpha) + \left\{\frac{1}{\widehat{\sigma}_{(0)}T}\sum_{t=p+1}^{T}\psi\left(\xi_{t,1}\right)\left(e_t-\widehat{e}_{(0)t}\right) + \frac{b_\sigma}{T}\sum_{t=p+1}^{T}\psi\left(\xi_{t,2}\right)\left(e_t-\alpha\right)\right\}$$

for all $\alpha\in\Theta_a$, with $b_\sigma := \widehat{\sigma}_{(0)}^{-1}-\sigma^{-1}$ and $\xi_{t,1}$ and $\xi_{t,2}$ lying in the line connecting $\widehat{e}_{(0)t}$ and $e_{(0)t}$, and $\widehat{\sigma}_{(0)}$ and σ, respectively. Then, for any $K>0$ and finite $C>0$, consistency of $\widehat{\boldsymbol{\kappa}}_{(0)}$ and $\widehat{\sigma}_{(0)}$ together with Assumptions 1 and A2ii) imply that,

$$\begin{aligned}\sup_{|\alpha|\leq K}\left|\widehat{P}_T(\alpha)-P_T(\alpha)\right| &\leq \frac{C}{\widehat{\sigma}_{(0)}}||\widehat{\boldsymbol{\kappa}}_{(0)}-\boldsymbol{\kappa}||\left(T^{-1}\sum_{t=p+1}^{T}||\mathbf{x}^*_{t-1,d}||\right)+C|b_\sigma|\left(K+T^{-1}\sum_{t=p+1}^{T}|e_t|\right)\\ &= o_p(1)\end{aligned}$$

and, hence, $\sup_{\alpha\in\Theta_a}\left|\widehat{P}_T(\alpha)-P_T(\alpha)\right| = o_p(1)$ for a bounded Θ_a. Paralleling the proof of Lemma A1, it can be shown that $P_T(\alpha)$ is stochastically equicontinuous under Assumption 1 because the ρ-function is Lipschitz under Assumption 2i) and ii) and, therefore, $\widehat{P}_T(\alpha)$ converges uniformly in probability to $E(\rho(\widetilde{e}_{t,\sigma}))$ in Θ_a by the AEL. Since conditions (A) to (C) in Thm. 4.1.1 in Amemiya (1985) are verified, there exists a solution to $\widehat{\alpha}_{(0)} := \arg\min_{\alpha\in\Theta_a}\widehat{P}_T(\alpha)$ for which $\widehat{\alpha}_{(0)}\xrightarrow{p}\arg\min_{\alpha\in\Theta_\alpha}E\left(\rho\left(\frac{e_t-\alpha}{\sigma}\right)\right)$. Moreover, given that α_p is assumed to minimize the conditional risk $E\left(\rho\left(\frac{e_t-\alpha}{\sigma}\right)|\mathcal{F}_{t-1}\right)$, we have that

$$E\left(\rho\left(\frac{e_t-\alpha}{\sigma}\right)|\mathcal{F}_{t-1}\right) > E\left(\rho\left(\frac{e_t-\alpha_p}{\sigma}\right)|\mathcal{F}_{t-1}\right)\quad\forall\alpha\neq\alpha_p;$$

and the LIE concludes that α_p minimizes the unconditional risk as well. Furthermore, because $\widehat{\alpha}_{(0)}$ is an M estimator of α_ρ, it can be shown under the assumptions considered that $\sqrt{T}\left(\widehat{\alpha}_{(0)}-\alpha_p\right)$ follows a limiting normal distribution; the arguments are the same as in Theorem 2 above and, therefore, we omit the details.

Finally, we show that the scale estimator fulfills the desired convergence rate. Standard regression algebra shows that, under the null,

$$\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=p+1}^{T}\widehat{e}^2_{(0)t} = \frac{1}{T}\sum_{t=1}^{T}e_t^2 + O_p\left(T^{-1}\right)$$

for $\sqrt{T}$-consistent $\widehat{\boldsymbol{\kappa}}$, so the convergence rate of the left-hand side hinges on the convergence rate of $\frac{1}{T}\sum_{t=1}^{T} e_t^2$. Examine to this end

$$\frac{1}{T}\sum_{t=1}^{T}\left(e_t^2-\sigma^2\right)=T^{-1/4}\left(\frac{1}{T^{3/4}}\sum_{t=1}^{T}\left(e_t^2-\sigma^2\right)\right)$$

where $\sigma^2 := E\left(e_t^2\right)$ and note that $E\left(\left|e_t^2-\sigma^2\right|^{1+\epsilon/2}\right)<\infty$ when $E\left(\left|e_t\right|^{2+\epsilon}\right)<\infty$. Given that $\{e_t^2-\sigma^2,\mathcal{F}_t\}$ is a MDS under Assumption 1, we can invoque the strong law of large number to show that $T^{-3/4}\sum_{t=1}^{T}\left(e_t^2-\sigma^2\right)\overset{a.s.}{\to}0$ (Davidson 2002, Theorem 20.11). The sufficient condition for the application of the strong law is that $\sum_{t\geq 1}\frac{E\left(\left|e_t^2-\sigma^2\right|^p\right)}{t^{3p/4}}<\infty$, which is clearly fulfilled for any $4/3<p\leq 2$, since $1+\epsilon/2>4/3$ by assumption. Since σ^2 is bounded away from zero and $\widehat{\sigma}^2$ consistent, the convergence rate of $\widehat{\sigma}^2$ carries over to $\widehat{\sigma}$ and, hence, $\widehat{\sigma}=\sigma+o_p\left(T^{-1/4}\right)$, as required.■

Proof of Theorem 4.

Without loss of generality, we set $k=1$, because any posterior iteration would build on an estimator with the same asymptotic properties as that in the previous iteration. Computing a one-step Newton-Raphson iteration produces $\widehat{\boldsymbol{\beta}}_{NR,(1)}=\widehat{\boldsymbol{\beta}}_{(0)}-\mathbf{s}_T\left(\widehat{\boldsymbol{\beta}}_{(0)}\right)$ with

$$\mathbf{s}_T\left(\widehat{\boldsymbol{\beta}}_{(0)}\right):=\left[\frac{1}{\widehat{\sigma}T}\sum_{t=p+1}^{T}\psi'\left(u_{t,\widehat{\sigma}}^{*}\left(\boldsymbol{\beta}\right)\right)\mathbf{x}_{t-1,d}^{*}\mathbf{x}_{t-1,d}^{\prime *}\right]_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_{(0)}}^{-1}\left[\frac{1}{T}\sum_{t=p+1}^{T}\psi\left(u_{t,\widehat{\sigma}}^{*}\left(\boldsymbol{\beta}\right)\right)\mathbf{x}_{t-1,d}^{*}\right]_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_{(0)}}$$

and, therefore,

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{M,(1)}-\boldsymbol{\beta}_0\right)=\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)}-\boldsymbol{\beta}_0\right)-\sqrt{T}\mathbf{s}_T\left(\widehat{\boldsymbol{\beta}}_{(0)}\right).$$

Since $\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)}-\boldsymbol{\beta}_0\right)=O_p\left(1\right)$, we may use Lemma A5 to write

$$\begin{aligned}\sqrt{T}\mathbf{s}_T\left(\widehat{\boldsymbol{\beta}}_{(0)}\right) &= \sigma\mathbf{A}_{\boldsymbol{\beta}}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\widehat{\sigma}}^{*}\left(\widehat{\boldsymbol{\beta}}_{(0)}\right)\right)\mathbf{x}_{t-1,d}^{*}\right]+o_p\left(1\right)\\ &= \sigma\mathbf{A}_{\boldsymbol{\beta}}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\widehat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\mathbf{x}_{t-1,d}^{*}+\boldsymbol{\pi}_T\left(\widehat{\boldsymbol{\beta}}_{(0)},\boldsymbol{\beta}_0\right)\right]+o_p\left(1\right)\end{aligned}$$

for

$$\boldsymbol{\pi}_T\left(\widehat{\boldsymbol{\beta}}_{(0)},\boldsymbol{\beta}_0\right):=\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\widehat{\sigma}}^{*}\left(\widehat{\boldsymbol{\beta}}_{(0)}\right)\right)\mathbf{x}_{t-1,d}^{*}-\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\widehat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\mathbf{x}_{t-1,d}^{*}$$

but since

$$\boldsymbol{\pi}_T\left(\widehat{\boldsymbol{\beta}}_{(0)},\boldsymbol{\beta}_0\right)=\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)}-\boldsymbol{\beta}_0\right)\left[\frac{1}{\widehat{\sigma}T}\sum_{t=2}^{T}\psi'\left(\xi_t\right)\mathbf{x}_{t-1,d}^{*}\mathbf{x}_{t-1,d}^{\prime *}\right]$$

we have from Lemma A6

$$\sqrt{T}\mathbf{s}_T\left(\widehat{\boldsymbol{\beta}}_{(0)}\right) = \sigma\mathbf{A}_{\boldsymbol{\beta}}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\hat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\mathbf{x}_{t-1,d}^{*}\right] + \sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)}-\boldsymbol{\beta}_0\right) + o_p\left(1\right)$$

and hence

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{M,(1)} - \boldsymbol{\beta}_0\right) = \sigma\mathbf{A}_{\boldsymbol{\beta}}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\hat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\mathbf{x}_{t-1,d}^{*}\right] + o_p\left(1\right)$$

from which we can conclude that,

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{M,(1)} - \boldsymbol{\beta}_0\right) \Rightarrow \mathcal{N}\left(0, \sigma^2\mathbf{A}_{\boldsymbol{\beta}}^{-1}\mathbf{B}_{\boldsymbol{\beta}}\mathbf{A}_{\boldsymbol{\beta}}^{-1}\right).$$

from Lemma A4*ii*).■

Proof of Theorem 5.

First, extend $\frac{\psi(r)}{r}$ continuously on $\mathbb{R}$ by setting it to equal $\psi'(0)$ at $r = 0$. Note then that $\frac{\psi(r)}{r}$ is Lipschitz since its derivative $\frac{\psi'}{r} - \frac{\psi}{r^2}$ is bounded on $\mathbb{R}$ with ψ' being itself Lipschitz, and thus having at most linear behavior at $r \to \pm\infty$, and ψ being bounded, and also noting that $\frac{\psi'}{r} - \frac{\psi}{r^2} \to \psi''(0)$ as $r \to 0$.

With $\frac{\psi(r)}{r}$ Lipschitz, like ψ', we may use the arguments from the proof of Lemma A5 to conclude that

$$\frac{1}{T}\sum_{t=p+1}^{T}\left[\frac{\psi\left(u_{t,\hat{\sigma}}^{*}\left(\widehat{\boldsymbol{\beta}}_{(0)}\right)\right)}{u_{t,\hat{\sigma}}^{*}\left(\widehat{\boldsymbol{\beta}}_{(0)}\right)}\right]\boldsymbol{x}_{t-1,d}^{*}\boldsymbol{x}_{t-1,d}^{\prime *} \xrightarrow{p} E\left(\left[\frac{\psi\left(\tilde{e}_{t,\sigma}\right)}{\tilde{e}_{t,\sigma}}\right]\boldsymbol{x}_{t-1,d}^{*}\boldsymbol{x}_{t-1,d}^{\prime *}\right) := \mathbf{C}_{\boldsymbol{\beta}} > 0. \quad (21)$$

Let us now examine the first iteration, $k = 1$. Given the convergence in (21), it is easily shown that

$$\begin{aligned}\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{IRLS,(1)} - \boldsymbol{\beta}_0\right) &= \sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)} - \boldsymbol{\beta}_0\right) - \sigma\mathbf{C}_{\boldsymbol{\beta}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\hat{\sigma}}^{*}\left(\widehat{\boldsymbol{\beta}}_{(0)}\right)\right)\boldsymbol{x}_{t-1,d}^{*}\right) + o_p\left(1\right)\\ &= \sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)} - \boldsymbol{\beta}_0\right) - \sigma\mathbf{C}_{\boldsymbol{\beta}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\hat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\boldsymbol{x}_{t-1,d}^{*} + \pi_T\right) + o_p\left(1\right).\end{aligned}$$

But

$$\pi_T = \left(\frac{1}{\widehat{\sigma}T}\sum_{t=p+1}^{T}\psi'\left(\xi_t\right)\boldsymbol{x}_{t-1,d}^{*}\boldsymbol{x}_{t-1,d}^{\prime *}\right)\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)} - \boldsymbol{\beta}_0\right)$$

such that for $\mathbf{D}_{\boldsymbol{\beta}} := \mathbf{I} - \mathbf{C}_{\boldsymbol{\beta}}^{-1}\mathbf{A}_{\boldsymbol{\beta}}$, we have

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{IRLS,(1)} - \boldsymbol{\beta}_0\right) = \mathbf{D}_{\boldsymbol{\beta}}\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)} - \boldsymbol{\beta}_0\right) - \sigma\mathbf{C}_{\boldsymbol{\beta}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\hat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\boldsymbol{x}_{t-1,d}^{*}\right) + o_p\left(1\right).$$

For $k = 2$, we obtain analogously that

$$\begin{aligned}
&\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{IRLS,(2)} - \boldsymbol{\beta}_0\right) \\
&\quad = \mathbf{D}_{\boldsymbol{\beta}}\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{IRLS,(1)} - \boldsymbol{\beta}_0\right) - \sigma\mathbf{C}_{\boldsymbol{\beta}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\hat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\boldsymbol{x}_{t-1,d}^{*}\right) + o_p\left(1\right) \\
&\quad = \mathbf{D}_{\boldsymbol{\beta}}^{2}\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)} - \boldsymbol{\beta}_0\right) - \sigma\mathbf{D}_{\boldsymbol{\beta}}\mathbf{C}_{\boldsymbol{\beta}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\hat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\boldsymbol{x}_{t-1,d}^{*}\right) \\
&\qquad -\sigma\mathbf{C}_{\boldsymbol{\beta}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\hat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\boldsymbol{x}_{t-1,d}^{*}\right);
\end{aligned}$$

and iterating k times leads to

$$\begin{aligned}
\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{IRLS,(k)} - \boldsymbol{\beta}_0\right) \;&=\; \mathbf{D}_{\boldsymbol{\beta}}^{k}\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{(0)} - \boldsymbol{\beta}_0\right) \\
&\quad - \left(\sum_{j=0}^{k}\mathbf{D}_{\boldsymbol{\beta}}^{j}\right)\sigma\mathbf{C}_{\boldsymbol{\beta}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\hat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\boldsymbol{x}_{t-1,d}^{*}\right) + o_p\left(1\right).
\end{aligned}$$

Since $\mathbf{I} - \mathbf{C}_{\boldsymbol{\beta}}^{-1}\mathbf{A}_{\boldsymbol{\beta}}$ has eigenvalues smaller than unity, we have that $\left(\mathbf{D}_{\boldsymbol{\beta}}\right)^{k} \to 0$ and $\sum_{j=0}^{k}\left(\mathbf{D}_{\boldsymbol{\beta}}\right)^{j} \to \left(\mathbf{C}_{\boldsymbol{\beta}}^{-1}\mathbf{A}_{\boldsymbol{\beta}}\right)^{-1}$ as $k \to \infty$, such that

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{IRLS,(k)} - \boldsymbol{\beta}_0\right) \to_{k\to\infty} -\sigma\mathbf{A}_{\boldsymbol{\beta}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(u_{t,\hat{\sigma}}^{*}\left(\boldsymbol{\beta}_0\right)\right)\boldsymbol{x}_{t-1,d}^{*}\right) + o_p\left(1\right)$$

which possesses the required weak limit as $T \to \infty$. ■

Proof of Theorem 6.

Note that Theorem 2 and Lemma A5 ensure that $\mathbf{A}_{\boldsymbol{\beta}T} \xrightarrow{p} \mathbf{A}_{\boldsymbol{\beta}}$, and that the same arguments may be employed to show that $\mathbf{B}_{\boldsymbol{\beta}T} \xrightarrow{p} \mathbf{B}_{\boldsymbol{\beta}}$ since $\psi^2(r)$ is easily shown to be Lipschitz. The result then follows from Theorems 2, 4 and 5. ■

Proof of Theorem 7.

We give the details for the test based on $\widehat{\boldsymbol{\beta}}_M$; the arguments are essentially the same for the iterated estimators. First, since $\theta = c/\sqrt{T}$, the fractional difference $\Delta^{\theta} := (1-L)^{\theta}$ may be linearized (cf. Tanaka, 1999),

$$\Delta^{\theta} - 1 \approx \theta\left(L + \frac{L^2}{2} + \frac{L^3}{3} + \ldots\right) + O(\theta^2)$$

leading to

$$\varepsilon_{t,d} = \Delta^{-\theta}\varepsilon_{t,d+\theta} = \varepsilon_{t,d+\theta} + \frac{c}{\sqrt{T}}x_{t-1,d+\theta}^{*} + O_p\left(T^{-1}\right)$$

with $x^*_{t-1,d+\theta} = \sum_{j=1}^{t-1} j^{-1}\varepsilon_{t-j,d+\theta} = \sum_{j=1}^{t-1} j^{-1}\varepsilon_{t-j}$. At the same time,

$$\varepsilon_{t,d+\theta} = \varepsilon_t = e_t + \sum_{j=1}^{p} \pi_j \varepsilon_{t-j} = e_t + \boldsymbol{\beta}_0' \boldsymbol{x}^*_{t-1,d+\theta},$$

hence

$$\begin{aligned} \varepsilon_{t,d} &= e_t + \sum_{j=1}^{p} \pi_j \varepsilon_{t-j,d+\theta} + \frac{c}{\sqrt{T}} x^*_{t-1,d+\theta} \\ &= 0 x^*_{t-1,d} + \sum_{j=1}^{p} \pi_j \varepsilon_{t-j,d} + \left(e_t + \frac{c}{\sqrt{T}} x^*_{t-1,d+\theta} - \frac{c}{\sqrt{T}} \sum_{j=1}^{p} \pi_j x^*_{t-j,d+\theta} \right); \end{aligned}$$

the DGP under the local alternative is

$$\varepsilon_{t,d} = \boldsymbol{\beta}_0' \boldsymbol{x}^*_{t-1,d} + e_{t,\theta}$$

where

$$e_{t,\theta} = e_t + \frac{c}{\sqrt{T}} x^*_{t-1,d+\theta} - \frac{c}{\sqrt{T}} \sum_{j=1}^{p} \pi_j x^*_{t-j,d+\theta} + O_p\left(\frac{1}{T}\right) \tag{22}$$

and

$$x^*_{t-1,d} = \Delta^{-\theta} x^*_{t-1,d+\theta} = x^*_{t-1,d+\theta} + \frac{c}{\sqrt{T}} \sum_{j=0}^{t-2} j^{-1} x^*_{t-1-j,d+\theta}$$

has analogous properties to $x^*_{t-1,d+\theta}$; in particular, $E\left(\left\|\boldsymbol{x}^*_{t-1,d}\right\|^2\right) < C \ \forall t$ (this is because $x^*_{t-1-j,d+\theta}$ is uniformly L_2-bounded and $T^{-1/2} \ln T \to 0$).

Now, since $\theta = c/\sqrt{T}$ is a deviation in a $1/\sqrt{T}$ neighbourhood of the null which is of the same order of magnitude as the estimation error, it is straightforward to show that $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$, $\mathbf{A}_{\boldsymbol{\beta}T} \xrightarrow{p} \mathbf{A}_{\boldsymbol{\beta}}$ and $\mathbf{B}_{\boldsymbol{\beta}T} \xrightarrow{p} \mathbf{B}_{\boldsymbol{\beta}}$ like in the proof of Theorem 6. To discuss the limiting behavior of the estimator $\widehat{\phi}_M$, we adapt the proof of Theorem 2 accordingly. The key step is to establish the limiting behavior of

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \psi\left(\frac{\varepsilon_{t,d} - \boldsymbol{\beta}_0' \boldsymbol{x}^*_{t-1,d}}{\widehat{\sigma}} \right) \frac{\boldsymbol{x}^*_{t-1,d}}{\widehat{\sigma}}.$$

Unlike the case $\theta = 0$, the limiting distribution of this normalized (pseudo-)score is not centered at zero. We resort to the mean value theorem for ψ with ξ_t between $\frac{\varepsilon_{t,d} - \boldsymbol{\beta}_0' \boldsymbol{x}^*_{t-1,d}}{\widehat{\sigma}}$

and the expansion points $\frac{e_t-\alpha_\rho}{\widehat{\sigma}}$ to obtain

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(\frac{\varepsilon_{t,d}-\boldsymbol{\beta}_0'\boldsymbol{x}_{t-1,d}^*}{\widehat{\sigma}}\right)\frac{\boldsymbol{x}_{t-1,d}^*}{\widehat{\sigma}}\\
&\quad=\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(\frac{e_t-\alpha_\rho}{\widehat{\sigma}}\right)\frac{\boldsymbol{x}_{t-1,d}^*}{\widehat{\sigma}}+\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi'\left(\xi_t\right)\left(\frac{\varepsilon_{t,d}-\boldsymbol{\beta}_0'\boldsymbol{x}_{t-1,d}^*}{\widehat{\sigma}}-\frac{e_t-\alpha_\rho}{\widehat{\sigma}}\right)\frac{\boldsymbol{x}_{t-1,d}^*}{\widehat{\sigma}}\\
&\quad=\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(\frac{e_t-\alpha_\rho}{\widehat{\sigma}}\right)\frac{\boldsymbol{x}_{t-1,d+\theta}^*}{\widehat{\sigma}}+\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(\frac{e_t-\alpha_\rho}{\widehat{\sigma}}\right)\left(\frac{\boldsymbol{x}_{t-1,d}^*}{\widehat{\sigma}}-\frac{\boldsymbol{x}_{t-1,d+\theta}^*}{\widehat{\sigma}}\right)\\
&\qquad+\frac{1}{\widehat{\sigma}^2\sqrt{T}}\sum_{t=p+1}^{T}\psi'\left(\xi_t\right)\left(e_{t,\theta}-e_t\right)\boldsymbol{x}_{t-1,d}^* \qquad (23)
\end{aligned}$$

with $e_{t,\theta}=e_t+\frac{c}{\sqrt{T}}x_{t-1,d+\theta}^*-\frac{c}{\sqrt{T}}\sum_{j=1}^{p}\pi_j x_{t-j,d+\theta}^*$. Using the same arguments as under the null hypothesis and exploiting the fact that now $\theta=c/\sqrt{T}$, it follows that

$$\frac{1}{\widehat{\sigma}}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(\frac{e_t-\alpha_\rho}{\widehat{\sigma}}\right)\boldsymbol{x}_{t-1,d+\theta}^*\right)\Rightarrow\mathcal{N}\left(0;\frac{1}{\sigma^2}\mathbf{B}_{\boldsymbol{\beta}}\right);$$

we also notice that

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\psi\left(\frac{e_t-\alpha_\rho}{\widehat{\sigma}}\right)\left(\frac{\boldsymbol{x}_{t-1,d}^*}{\widehat{\sigma}}-\frac{\boldsymbol{x}_{t-1,d+\theta}^*}{\widehat{\sigma}}\right)\xrightarrow{p}0$$

(see the discussion in the proof of Lemma A4*ii)*). By replacing $e_{t,\theta}-e_t$ by $\frac{c}{\sqrt{T}}x_{t-1,d+\theta}^*-\frac{c}{\sqrt{T}}\sum_{j=1}^{p}\pi_j x_{t-j,d+\theta}^*$, we obtain for the third summand in (23) that,

$$\boldsymbol{\Delta}_T=\frac{1}{\widehat{\sigma}^2\sqrt{T}}\sum_{t=p+1}^{T}\psi'\left(\xi_t\right)\left(\frac{c}{\sqrt{T}}x_{t-1,d+\theta}^*-\frac{c}{\sqrt{T}}\sum_{j=1}^{p}\pi_j x_{t-j,d+\theta}^*\right)\boldsymbol{x}_{t-1,d}^*.$$

Denote now $\boldsymbol{\delta}_c:=c\left(0,1,-\pi_1,\ldots,-\pi_p\right)'$ and recall that $x_{t-1,d+\theta}^*=x_{t-1,d}^*-\frac{c}{\sqrt{T}}\sum_{j=0}^{t-2}j^{-1}x_{t-1-j,d+\theta}^*$ so

$$\begin{aligned}
&\frac{1}{\widehat{\sigma}^2\sqrt{T}}\sum_{t=p+1}^{T}\psi'\left(\xi_t\right)\left(\frac{c}{\sqrt{T}}x_{t-1,d+\theta}^*-\frac{c}{\sqrt{T}}\sum_{j=1}^{p}\pi_j x_{t-j,d+\theta}^*\right)\boldsymbol{x}_{t-1,d}^*\\
&\quad=\frac{1}{\widehat{\sigma}^2T}\sum_{t=p+1}^{T}\psi'\left(\xi_t\right)\boldsymbol{x}_{t-1,d}^*\left(\boldsymbol{x}_{t-1,d}'^*\boldsymbol{\delta}_c-\frac{c^2}{\sqrt{T}}\sum_{j=0}^{t-2}j^{-1}x_{t-1-j,d+\theta}^*\right)\\
&\quad=\frac{1}{\widehat{\sigma}^2T}\sum_{t=p+1}^{T}\psi'\left(\xi_t\right)\boldsymbol{x}_{t-1,d}^*\boldsymbol{x}_{t-1,d}'^*\boldsymbol{\delta}_c+o_p\left(1\right)
\end{aligned}$$

since

$$E\left(\left\|\psi'\left(\xi_t\right)\boldsymbol{x}_{t-1,d}^*\sum_{j=0}^{t-2}j^{-1}x_{t-1-j,d+\theta}^*\right\|\right)\leq C\sum_{j=0}^{t-2}j^{-1}\sqrt{E\left(\left\|\boldsymbol{x}_{t-1,d}^*\right\|^2\right)E\left(\left|x_{t-1-j,d+\theta}^*\right|^2\right)}=O\left(\ln T\right)$$

and thus

$$\frac{c^2}{\widehat{\sigma}^2 T^{3/2}} \sum_{t=p+1}^{T} \psi'(\xi_t)\, \boldsymbol{x}^*_{t-1,d} \left(\sum_{j=0}^{t-2} j^{-1} x^*_{t-1-j,d+\theta} \right) = O_p\left(\frac{\ln T}{\sqrt{T}} \right).$$

Hence, using Lemma A5 again,

$$\boldsymbol{\Delta}_T \xrightarrow{p} \frac{1}{\sigma^2} \mathbf{A}_{\boldsymbol{\beta}} \boldsymbol{\delta}_c$$

Consequently, under the conditions considered,

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_0 \right) \Rightarrow \mathcal{N}\left(\boldsymbol{\delta}_c; \boldsymbol{\Omega}_{\boldsymbol{\beta}} \right)$$

from which the required result follows directly. ■

Table 1: Rejection frequencies of fractional integration tests (i.i.d. errors drawn from a Student-t distribution with v degrees of freedom)

θ	t_{LS}	t_{LS}^{HC}	$t_{\mathcal{M}}^{H}$	$t_{\mathcal{M}}^{B}$	DV	t_{QR}	θ	t_{LS}	t_{LS}^{HC}	$t_{\mathcal{M}}^{H}$	$t_{\mathcal{M}}^{B}$	DV	t_{QR}
$v = 1000, T = 250$							$v = 1000, T = 500$						
-0.3	100.00	100.00	99.98	99.98	71.78	98.72	-0.3	100.00	100.00	100.00	100.00	97.96	100.00
-0.2	97.58	97.64	96.84	96.80	41.14	84.62	-0.2	99.98	99.98	99.96	99.94	78.96	98.50
-0.1	53.88	54.06	50.62	50.48	11.80	37.76	-0.1	81.12	81.58	79.04	79.14	28.68	60.24
0.0	5.20	5.70	5.24	5.22	4.56	5.98	0	5.38	5.64	5.46	5.40	4.80	5.90
0.1	36.18	37.04	34.50	34.52	35.70	27.34	0.1	70.76	71.06	68.82	68.96	56.68	53.48
0.2	91.60	91.92	90.44	90.40	85.72	80.62	0.2	99.92	99.94	99.88	99.84	98.32	98.52
0.3	99.88	99.92	99.80	99.78	98.96	98.38	0.3	100.00	100.00	100.00	100.00	100.00	100.00
$v = 3, T = 250$							$v = 3, T = 500$						
-0.3	99.96	99.60	100.00	100.00	68.20	100.00	-0.3	100.00	99.94	100.00	100.00	96.48	100.00
-0.2	97.54	96.72	99.66	99.54	43.72	98.86	-0.2	99.96	99.68	100.00	100.00	81.72	100.00
-0.1	54.00	61.22	73.62	72.96	15.36	64.10	-0.1	82.24	83.64	96.04	95.74	37.20	90.98
0.0	5.16	5.82	5.26	5.42	4.40	4.86	0	4.92	5.68	5.30	5.36	4.84	4.54
0.1	37.12	42.20	59.92	62.86	51.38	54.68	0.1	71.58	73.04	91.50	92.00	76.80	88.24
0.2	91.72	90.82	98.36	98.62	95.12	97.70	0.2	99.88	99.58	100.00	100.00	99.82	100.00
0.3	99.76	99.50	99.98	99.98	99.82	99.98	0.3	100.00	99.98	100.00	100.00	100.00	100.00
$v = 2, T = 250$							$v = 2, T = 500$						
-0.3	99.8	98.84	99.98	99.98	50.66	99.98	-0.3	99.98	99.60	100.00	99.98	80.46	100.00
-0.2	97.06	95.62	99.94	99.96	33.70	99.82	-0.2	99.82	99.10	100.00	100.00	62.72	100.00
-0.1	53.22	69.42	91.16	91.22	15.18	85.34	-0.1	85.72	89.32	99.68	99.64	32.92	99.06
0.0	4.26	5.48	5.38	5.44	4.26	3.92	0	3.76	4.64	5.36	5.44	4.70	3.66
0.1	35.92	50.16	83.32	86.30	68.26	81.68	0.1	74.24	78.20	98.90	99.16	91.46	98.56
0.2	93.40	91.48	99.76	99.84	98.50	99.70	0.2	99.80	99.22	100.00	100.00	99.98	100.00
0.3	99.70	99.06	100.00	100.00	99.96	100.00	0.3	100.00	99.94	100.00	100.00	100.00	100.00

Table 2: Rejection frequencies of fractional integration tests (AR(1) short-run dynamics with coefficient 0.5 and errors drawn from a Student-t ditribution with v degrees of freedom)

θ	t_{LS}	t_{LS}^{HC}	$t_{\mathcal{M}}^{H}$	$t_{\mathcal{M}}^{B}$	DV	t_{QR}	θ	t_{LS}	t_{LS}^{HC}	$t_{\mathcal{M}}^{H}$	$t_{\mathcal{M}}^{B}$	DV	t_{QR}
$v = 1000$, $T = 250$							$v = 1000$, $T = 500$						
-0.3	49.84	51.50	47.82	47.14	0.78	36.42	-0.3	82.10	82.44	79.20	79.44	2.46	61.46
-0.2	26.82	27.90	25.88	25.72	0.78	21.22	-0.2	48.00	48.50	44.92	44.86	1.40	32.76
-0.1	13.26	13.98	12.94	12.84	0.86	12.00	-0.1	17.92	18.04	16.82	16.78	1.14	14.48
0.0	7.12	7.68	6.94	6.88	1.52	7.44	0	5.82	6.28	6.00	5.84	1.42	6.58
0.1	4.84	5.50	5.16	5.18	2.06	6.66	0.1	5.64	5.82	5.66	5.62	2.80	6.86
0.2	5.02	5.52	5.34	5.22	2.50	7.14	0.2	11.08	11.22	10.30	10.08	4.38	10.32
0.3	5.06	5.42	5.00	5.04	2.46	7.14	0.3	15.04	15.22	12.40	12.44	4.56	13.28
$v = 3$, $T = 250$							$v = 3$, $T = 500$						
-0.3	51.14	57.24	71.78	70.82	1.80	62.44	-0.3	82.78	84.08	96.04	95.46	2.62	90.54
-0.2	26.50	31.64	40.20	39.76	1.68	33.52	-0.2	48.00	52.56	70.04	68.28	2.74	59.24
-0.1	12.04	14.62	15.42	14.68	2.26	13.24	-0.1	17.38	20.14	24.62	24.02	2.86	20.42
0.0	5.86	6.86	6.12	5.44	3.48	5.66	0	5.96	6.42	5.88	5.54	4.36	5.18
0.1	3.96	4.68	4.78	5.64	5.52	5.64	0.1	5.90	6.88	9.72	10.90	8.26	9.68
0.2	4.22	5.28	7.82	9.10	9.30	9.64	0.2	11.16	12.88	23.04	25.76	14.52	24.08
0.3	4.42	5.36	9.00	10.44	11.42	12.18	0.3	14.56	16.60	30.42	33.76	18.84	34.08
$v = 2$, $T = 250$							$v = 2$, $T = 500$						
-0.3	50.12	67.34	90.60	88.74	9.88	84.68	-0.3	85.14	88.06	99.76	99.56	12.16	99.28
-0.2	24.28	40.86	62.50	61.60	7.48	55.34	-0.2	45.66	62.04	91.82	90.98	9.60	86.58
-0.1	10.28	18.14	23.46	22.60	5.72	18.36	-0.1	15.26	25.82	44.26	44.20	6.58	37.02
0.0	5.36	6.02	4.94	4.78	5.78	4.36	0	4.42	5.30	4.70	4.50	7.12	3.72
0.1	4.10	4.58	9.90	11.80	13.06	7.58	0.1	4.54	7.52	24.66	27.68	17.86	21.22
0.2	4.30	6.08	19.86	24.90	22.72	19.64	0.2	9.58	16.44	51.48	57.04	36.94	50.50
0.3	4.26	5.94	24.06	30.40	28.82	26.9	0.3	14.24	20.66	63.28	69.02	50.20	65.06

Table 3: Rejection frequencies of fractional integration tests (time-varying GARCH-type volatility and iid innovations drawn fom a Student-t distribution with v degrees of freedom)

	GARCH-A						GARCH-B				
θ	t_{LS}	t_{LS}^{HC}	$t_{\mathcal{M}}^{H}$	$t_{\mathcal{M}}^{B}$	t_{QR}		t_{LS}	t_{LS}^{HC}	$t_{\mathcal{M}}^{H}$	$t_{\mathcal{M}}^{B}$	t_{QR}
			v = 1000						v = 1000		
-0.3	100.00	100.00	100.00	100.00	100.00	-0.3	100.00	100.00	100.00	100.00	100.00
-0.2	99.96	99.96	99.94	99.94	98.22	-0.2	99.92	99.90	99.92	99.92	98.20
-0.1	73.16	72.70	75.24	75.08	57.26	-0.1	73.10	71.68	74.96	74.62	57.90
0	5.38	5.18	6.06	6.02	5.90	0	5.18	4.80	6.26	6.16	6.00
0.1	65.76	65.82	50.34	50.36	38.78	0.1	65.46	65.06	50.58	50.52	39.00
0.2	98.96	98.98	97.02	96.86	90.02	0.2	98.92	98.96	96.92	96.76	89.90
0.3	99.98	99.98	99.88	99.86	99.38	0.3	100.00	100.00	99.86	99.86	99.34
			v = 3						v = 3		
-0.3	100.00	99.90	100.00	100.00	100.00	-0.3	100	99.90	100.00	100.00	100.00
-0.2	99.88	99.60	100.00	100.00	100.00	-0.2	99.88	99.50	100.00	100.00	100.00
-0.1	71.28	73.48	93.52	93.12	88.24	-0.1	71.02	72.24	93.66	92.78	88.20
0	5.10	5.62	5.48	5.26	5.44	0	5.54	5.50	6.00	5.66	5.60
0.1	67.98	69.70	80.36	83.06	76.38	0.1	68.2	69.02	80.48	83.02	76.24
0.2	99.08	98.78	99.88	99.92	99.76	0.2	99.1	98.70	99.90	99.92	99.80
0.3	99.98	99.98	100.00	100.00	100.00	0.3	99.98	99.96	100.00	100.00	100.00
			v = 2						v = 2		
-0.3	99.98	99.70	100.00	100.00	100.00	-0.3	99.98	99.68	100.00	100.00	100.00
-0.2	99.78	99.04	100.00	100.00	100.00	-0.2	99.76	98.88	100.00	100.00	100.00
-0.1	75.44	81.02	99.66	99.52	99.32	-0.1	75.30	80.48	99.64	99.54	99.32
0	4.06	4.32	5.28	5.26	3.42	0	4.44	4.40	5.62	5.38	3.92
0.1	68.30	73.64	97.04	98.08	96.74	0.1	68.46	73.38	96.90	97.98	96.74
0.2	98.98	97.92	100.00	100.00	100.00	0.2	98.96	97.80	100.00	100.00	100.00
0.3	99.96	99.82	100.00	100.00	100.00	0.3	99.96	99.82	100.00	100.00	100.00

Table 4: Descriptive statistics of log absolute returns and log price range volatility estimates. By columns, mean, standard deviation, skewness, kurtosis, Ljung-Box test statistic for absence of autocorrelation up the the first 100 lags, Jarque-Bera test for normality, and total number of observations. Critical values Ljung-Box test: 118.50 (90%), 124.34 (95%), 135.81 (99%). Critical values JB test: 4.60 (90%), 5.99 (95%), 9.21 (99%).

	Panel A: Log absolute returns							Panel B: Log range estimates						
	Mean	Std.Dev.	Skewness	Kurtosis	LB Test	JB Test	Num.Obs.	Mean	Std.Dev.	Skewness	Kurtosis	LB Test	JB Test	Num.Obs.
SP500	-5.377	1.281	-1.094	5.271	4501.12	1771.54	4274	-4.477	0.609	0.263	3.094	63862.93	50.88	4277
NASDAQ C.	-5.070	1.240	-0.935	4.666	8144.98	1117.83	4276	-4.296	0.625	0.280	2.881	97317.13	58.40	4277
CAC40	-5.082	1.217	-1.175	5.515	2928.39	2145.28	4345	-4.266	0.569	0.144	2.925	65821.77	16.07	4347
DAX30	-5.064	1.215	-1.066	5.219	4412.51	1703.85	4318	-4.192	0.600	0.140	2.955	79346.00	14.56	4319
FTSE100	-5.301	1.194	-0.990	4.649	3549.44	1186.23	4288	-4.383	0.568	0.312	3.043	68350.23	69.85	4296
IBEX35	-5.032	1.183	-1.132	5.468	3650.40	2014.92	4310	-4.181	0.559	-0.002	2.873	78782.32	2.91	4315
NIKKEI225	-5.016	1.211	-1.182	5.441	1178.62	2005.39	4169	-4.392	0.554	0.397	5.180	29623.50	936.22	4174
BOVESPA	-4.782	1.158	-1.216	5.470	648.69	2107.87	4209	-3.885	0.473	0.302	3.662	18134.08	140.96	4215
HANG SENG	-5.115	1.241	-1.168	6.114	4189.69	2647.42	4192	-4.408	0.530	0.397	3.343	53040.87	130.51	4196
MERVAL	-4.728	1.243	-1.183	5.534	1050.80	2083.93	4162	-3.923	0.569	0.337	3.197	22960.19	85.85	4174
MXX	-5.211	1.204	-1.061	4.975	2619.15	1494.20	4269	-4.349	0.560	0.279	3.128	32349.31	58.46	4275

Table 5: 95% confidence interval estimates for the fractional parameter on log absolute returns according to different testing methodologies. By columns, LS-based estimates with robust standard errors to hetoroskedasticity, t_{LS}^{HC}; QR-based estimates at the median with robust standard errors, t_{QR}; Kolmogorov-Smirnov joint QR test-based estimates at quantiles in [0.1,0.9], KS; IRLS-based estimates with Huber, IRLS_H, and Biweight, IRLS_H, weighting functions; one-step Newton-Raphson test-based estimates with Huber, NR_H, and Biweight, NR_B, weigthing functions. The column labelled Intersection shows the values for which none of these tests can reject the null hypothesis of fractional integration, i.e., the intersection of the confidence intervals. ELW reports 95% confidence intervals based on exact local Whittle estimation. Finally, the last two columns show the average relative change in the amplitude of the estimated confidence intervals when LS is replaced by IRLS and NR estimation, respectively, with average values computed over the estimates from Huber and Bisquare weighthing functions.

	t_{LS}^{HC}	t_{QR}	KS	IRLS_H	IRLS_B	NR_H	NR_B	Intersection	ELW	ARC IRLS	ARC NR
SP500	[0.37,0.53]	[0.35,0.52]	[0.34,0.54]	[0.38,0.49]	[0.37,0.49]	[0.37,0.50]	[0.36,0.50]	[0.38,0.49]	[0.39.0.45]	-28.13%	-15.63%
NASDAQ C.	[0.37,0.54]	[0.37,0.48]	[0.35,0.48]	[0.37,0.45]	[0.38,0.47]	[0.36,0.46]	[0.36,0.46]	[0.38,0.45]	[0.39,0.46]	-50.00%	-38.24%
CAC40	[0.30,0.48]	[0.25,0.42]	[0.31,0.44]	[0.31,0.41]	[0.31,0.41]	[0.30.0.43]	[0.30,0.42]	[0.31,0.41]	[0.33,0.39]	-44.44%	-33.33%
DAX30	[0.36,0.52]	[0.33,0.46]	[0.32,0.48]	[0.36,0.46]	[0.35,0.45]	[0.35,0.47]	[0.34,0.46]	[0.36,0.45]	[0.37,0.44]	-37.50%	-25.00%
FTSE100	[0.32,0.48]	[0.35,0.47]	[0.33,0.49]	[0.36,0.46]	[0.36,0.47]	[0.34,0.48]	[0.35,0.48]	[0.36,0.46]	[0.33,0.39]	-34.38%	-15.63%
IBEX35	[0.37,0.53]	[0.34,0.48]	[0.35,0.51]	[0.37,0.47]	[0.37,0.47]	[0.36,0.48]	[0.36,0.48]	[0.37,0.47]	[0.37,0.44]	-37.50%	-25.00%
NIKKEI225	[0.27,0.44]	[0.23,0.39]	[0.26,0.43]	[0.28,0.40]	[0.28,0.40]	[0.27,0.41]	[0.27,0.41]	[0.28,0.39]	[0.30,0.37]	-29.41%	-17.65%
BOVESPA	[0.20,0.39]	[0.23,0.39]	[0.24,0.40]	[0.26,0.36]	[0.26,0.36]	[0.24,0.37]	[0.25,0.37]	[0.26,0.40]	[0.24,0.30]	-47.37%	-34.21%
HANG SENG	[0.36,0.50]	[0.37,0.51]	[0.35,0.48]	[0.37,0.46]	[0.37,0.46]	[0.36,0.47]	[0.36,0.47]	[0.37,0.46]	[0.36,0.43]	-35.71%	-21.43%
MERVAL	[0.19,0.35]	[0.18,0.36]	[0.23,0.36]	[0.23,0.33]	[0.23,0.34]	[0.21,0.35]	[0.22,0.35]	[0.23,0.33]	[0.23,0.29]	-25.00%	-3.57%
MXX	[0.32,0.50]	[0.27,0.41]	[0.28,0.43]	[0.32,0.41]	[0.32,0.41]	[0.31,0.43]	[0.31,0.42]	[0.32,0.41]	[0.32,0.38]	-44.44%	-36.11%

Table 6: 95% confidence interval estimates for the fractional parameter on log range volatility estimates according to different testing methodologies. By columns, LS-based estimates with robust standard errors to hetoroskedasticity, t_{LS}^{HC}; QR-based estimates at the median with robust standard errors, t_{QR}; Kolmogorov-Smirnov joint QR test-based estimates at quantiles in [0.1,0.9], KS; IRLS-based estimates with Huber, IRLS_H, and Biweight, IRLS_H, weighting functions; one-step Newton-Raphson test-based estimates with Huber, NR_H, and Biweight, NR_B, weigthing functions. The column labelled Intersection shows the values for which none of these tests can reject the null hypothesis of fractional integration, i.e., the intersection of the confidence intervals. DV and ELW reports 95% confidence intervals based on the sign test and exact local Whittle estimation, respectively. Finally, the last two columns show the average relative change in the amplitude of the estimated confidence intervals when LS is replaced by IRLS and NR estimation, respectively, with average values computed over the estimates from Huber and Bisquare weighthing functions.

	t_{LS}^{HC}	t_{QR}	KS	IRLS_H	IRLS_B	NR_H	NR_B	Intersection	DV	ELW	ARC IRLS	ARC NR
SP500	[0.50,0.69]	[0.36,0.62]	[0.41,0.68]	[0.49,0.66]	[0.48,0.66]	[0.47,0.68]	[0.47.0.67]	[0.49,0.62]	[-0.10,0.65]	[0.55,0.62]	-7.89%	7.89%
NASDAQ C.	[0.54,0.68]	[0.49,0.68]	[0.45,0.67]	[0.49,0.62]	[0.49,0.63]	[0.48,0.64]	[0.48,0.64]	[0.54,0.62]	[-0.05,0.78]	[0.60,0.67]	-3.57%	14.29%
CAC40	[0.46,0.64]	[0.44,0.64]	[0.47,0.68]	[0.48,0.61]	[0.48,0.62]	[0.46,0.63]	[0.47,0.63]	[0.48,0.61]	[-0.28,0.85]	[0.56,0.63]	-25.00%	-8.33%
DAX30	[0.54,0.73]	[0.51,0.78]	[0.55,0.77]	[0.54,0.70]	[0.54,0.70]	[0.52,0.72]	[0.52,0.72]	[0.55,0.70]	[0.26,0.66]	[0.60,0.67]	-15.79%	5.26%
FTSE100	[0.48,0.70]	[0.43,0.76]	[0.47,0.73]	[0.50,0.67]	[0.50,0.67]	[0.48,0.69]	[0.48,0.69]	[0.50,0.67]	[0.02,0.74]	[0.58,0.64]	-22.73%	-4.55%
IBEX35	[0.49,0.64]	[0.40,0.67]	[0.46,0.68]	[0.52,0.63]	[0.52,0.64]	[0.50,0.65]	[0.50,0.65]	[0.52,0.63]	[-0.14,0.86]	[0.53,0.60]	-16.67%	0.00%
NIKKEI225	[0.43,0.64]	[0.43,0.62]	[0.39,0.65]	[0.47,0.60]	[0.45,0.62]	[0.44,0.63]	[0.45,0.62]	[0.47,0.60]	[-0.13,0.69]	[0.51,0.58]	-28.57%	-14.29%
BOVESPA	[0.35,0.64]	[0.32,0.60]	[0.30,0.64]	[0.38,0.58]	[0.38,0.58]	[0.36,0.61]	[0.36,0.60]	[0.38,0.58]	[-0.24,0.81]	[0.48,0.54]	-31.03%	-15.52%
HANG SENG	[0.54,0.67]	[0.52,0.67]	[0.49,0.62]	[0.51,0.62]	[0.51,0.62]	[0.50,0.64]	[0.50,0.64]	[0.57,0.62]	[0.57,1.05]	[0.56,0.62]	-15.38%	7.69%
MERVAL	[0.39,0.56]	[0.45,0.65]	[0.40,0.59]	[0.42,0.55]	[0.42,0.56]	[0.40,0.57]	[0.40,0.57]	[0.45,0.55]	[-0.29,0.60]	[0.43,0.49]	-20.59%	0.00%
MXX	[0.48,0.65]	[0.39,0.61]	[0.41,0.65]	[0.45,0.60]	[0.45,0.60]	[0.43,0.62]	[0.43,0.61]	[0.48,0.60]	[-0.30,0.87]	[0.48,0.54]	-11.76%	8.82%

Figures

Figure 1: Time-series dynamics, sample histogram (confronted with the theoretical normal distribution), and sample autocorrelation function (with asymptotic 95% confidance bands) of the log absolute-valued returns of the SP500 index.

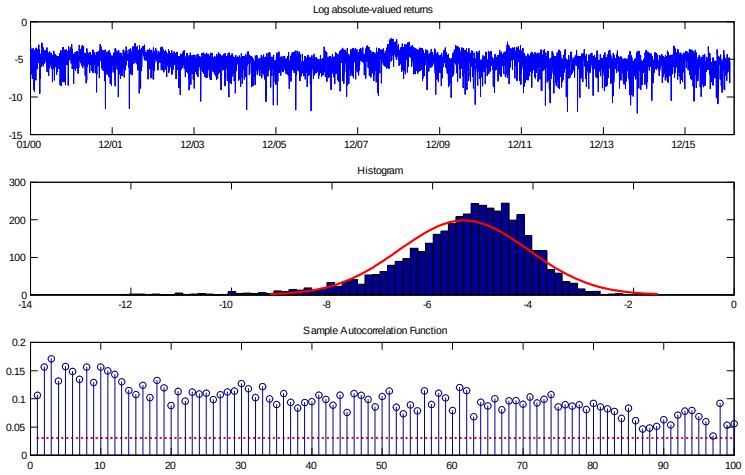

Figure 2: Time-series dynamics, sample histogram (confronted with the theoretical normal distribution), and sample autocorrelation function (with asymptotic 95% confidance bands) of the log high-low range estimator of the SP500 index.