Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 250–282
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

The quantilogram: With an application to evaluating directional predictability

O. Linton[a,*,1], Yoon-Jae Whang[b,2]

[a]Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK
[b]School of Economics, Seoul National University, Seoul 151-742, Korea

Available online 16 February 2007

Abstract

We propose a new diagnostic tool for time series called the quantilogram. The tool can be used formally and we provide the inference tools to do this under general conditions, and it can also be used as a simple graphical device. We apply our method to measure directional predictability and to test the hypothesis that a given time series has no directional predictability. The test is based on comparing the correlogram of quantile hits to a pointwise confidence interval or on comparing the cumulated squared autocorrelations with the corresponding critical value. We provide the distribution theory needed to conduct inference, propose some model free upper bound critical values, and apply our methods to S&P500 stock index return data. The empirical results suggest some directional predictability in returns. The evidence is strongest in mid range quantiles like 5–10% and for daily data. The evidence for predictability at the median is of comparable strength to the evidence around the mean, and is strongest at the daily frequency.

JEL classification: C12; C13; C14; C22

Keywords: Correlogram; Dependence; Efficient markets; Empirical process; Portmanteau; Quantiles

*Corresponding author. Tel.: +44 171 955 7864; fax: +44 171 831 1840.
E-mail addresses: lintono@lse.ac.uk (O. Linton), whang@snu.ac.kr (Y.-J. Whang).
[1]Research was supported by the Economic and Social Science Research Council of the United Kingdom.
[2]Research was supported by the Korea Research Foundation Grant (KRF-2004-041-B00066).

doi:10.1016/j.jeconom.2007.01.004

1. Introduction

We propose a new diagnostic tool for time series called the quantilogram. The tool can be used formally and we provide the inference tools to do this under general conditions, and it can also be used as a simple graphical device. A major reason for introducing this was to provide a simple method of measuring directional predictability and testing for the hypothesis that a given time series has no directional predictability. There is a large literature in empirical finance attempting to find predictability in the direction of stock prices, based on the statistics of signs and ranks. Cowles and Jones (1937) proposed a statistic for testing market efficiency based on the frequencies of up movements relative to down movements; specifically, they computed the Cowles–Jones ratio

$$CJ = \frac{N_{\mathrm{s}}}{N_{\mathrm{r}}},$$

where N_{s} is the number of sequences in the sample and N_{r} is the number of reversals in the sample. Both these quantities are simple functions of the sign $Z_t = 1(y_t > 0) - 1(y_t < 0) = \mathrm{sign}(y_t)$ of returns y_t. Under the assumption of i.i.d. $\mathrm{N}(0, \sigma^2)$ returns, which is what was taken to be the efficient markets in those days, $CJ = 1$. They computed CJ for a number of stock price indexes, finding some evidence of predictability relative to what would be expected. Campbell et al. (1997, Section 2.2.2) point out that when there is a non-zero drift in stock returns y_t, the sign of returns has a non-zero mean and this could account for some of the earlier violations. They compute the asymptotic distribution of the C–J ratio under the hypothesis of i.i.d. normal returns with drift $y_t \sim \mathrm{N}(\mu, \sigma^2)$. They show how the limiting distribution is affected by μ, and then how it can be corrected for μ. The limiting distribution and so the correction is heavily dependent on the normality assumption.

Christoffersen and Diebold (2002) examined these findings again. They questioned the basic model adopted and instead allowed for time-varying volatility. For the most part they assumed that returns satisfied

$$y_t = \mu + \sigma_t \varepsilon_t,$$

where σ_t is a conditional heteroskedastic (CH for short) process, measurable with respect to past information, and $\varepsilon_t \sim \mathrm{N}(0, 1)$. This process is consistent with the efficient markets hypothesis but is stronger than necessary. They argue that: (1) volatility dependence can induce sign dependence if expected returns are non-zero; (2) mean independence (market efficiency) is consistent with sign dependence and time-varying volatility. Thus a naive test of sign dependence is unlikely to reveal anything about market efficiency, unless we truly believe in a very simple null hypothesis. They further argued based on some continuous time models that (3) sign dependence is unlikely to be found in daily data, but more likely to be found in intermediate frequency data.

We question the choice of directional indicator. The motivation for analyzing pure signs (i.e., taking zero for centring), apart from the fact that many old studies including Cowles and Jones (1937) did, is presumably that zero is a neutral value delineating good from bad. We would argue that $\mathrm{E}(y_t)$, for example, is a better benchmark for this purpose, especially when $\mathrm{E}(y_t)$ is non-trivial, which is the case for say monthly data. Suppose that we examine the sign of centred returns

$$Z_t^* = 1(y_t > \mathrm{E}(y_t)) - 1(y_t < \mathrm{E}(y_t))$$

instead of the uncentred sign Z_t. In the strong CH model adopted by Christofferson and Diebold (and perforce in the special case of Cowles and Jones) there is no predictability in the centred sign Z_t^*.

Hong and Chung (2003) also modified the basic measure of direction: they chose an indicator based on the direction of stock returns relative to a threshold c that they consider non-random, i.e.,

$$Z_t(c) = 1(y_t > c), \quad -\infty < c < \infty.$$

They propose a new method for testing predictability based on this indicator. Theirs is an omnibus test based on a generalized spectrum, i.e., the test statistic is a functional of

$$\Delta_j(u, v) = \mathrm{cov}(\mathrm{e}^{\mathrm{i}uZ_t(c)}, \mathrm{e}^{\mathrm{i}uy_{t-|j|}})$$

over u, v, j. They show that it is consistent against a wide class of alternatives. They find evidence of predictability for a number of daily U.S. stock indexes for $c = 0, 0.5, 1, 1.5$ in 'units of standard deviations'. Note that by scaling by estimated standard deviation, the limiting distribution of their test statistics is affected, but they do not take account of this.

This paper is written in a specific to general style. We first propose a simple diagnostic statistic for measuring the extent of directional predictability based on a sample correlation. In contrast to the fixed threshold of Hong and Chung (2003) we take our threshold to be an estimated unconditional quantile. Our null hypothesis is that the chosen conditional quantile is not time varying. In the case of the median we are looking at the autocorrelation of returns signed relative to their unconditional median rather than the raw signs used in the C–J test and Christoffersen and Diebold (2002). We look at individual correlations but also aggregate into Box–Pierce-type statistics that take account of a number of lags. In practice we must replace the population quantile by an estimate. This in general affects the limiting distribution, and our theory captures the leading effect of this estimation.[3] The advantage of our approach is: (a) conceptual—using the quantile in connection with counts seems preferable from a statistical perspective to using a fixed threshold whose meaning is uncertain and depends on the time frame, etc.; (b) simplicity in computation and interpretation; (c) simple asymptotic theory. In some special cases the effect due to estimation of the quantiles disappears. In other cases, we give 'model free' upper bound critical values. We apply our test statistic to a sample of daily, weekly, and monthly returns on the S&P500. We find some evidence of predictability in the high-frequency data when a number of lags are taken into account. The evidence at the median threshold point is similar to the evidence of predictability in the mean for this data. The evidence at lower quantiles is much stronger. There is much less evidence of predictability for any quantile in monthly data. This seems to be (a little bit) contrary to the prediction of Christoffersen and Diebold (2002) that sign dependence is not likely to be found in high-frequency data but more likely to be found in data with frequency of two or three months. Having studied the issue of directional predictability we next generalize the setting. We allow the data to be residuals from a general time series model. In this case, the quantilogram provides a method for detecting model misspecification that is useful in cases where the data are heavy tailed. We provide the general tools for doing inference in this setting.

[3]Our approach is related to that taken in Engle and Manganelli (2004, Section 4) except that they take a more regression-based framework. We discuss their approach in comparison with ours later.

In Section 2 we discuss the model and null hypothesis, while in Section 3 we define the quantilogram for a series and describe its estimation. In Section 4 we describe the asymptotic properties of the estimated quantilogram, propose some test statistics, and discuss their implementation. Our procedures can be used more generally as a diagnostic for a specific time series model, and in Section 5 we generalize the setting to allow the data to be residuals from some separately estimated parametric model. In Section 6 we give some numerical results, and Section 7 concludes. The proofs of all theorems are in the Appendix.

2. Model and null hypothesis

Suppose that random variables $y_1, y_2, \ldots$ are from a stationary process whose marginal distribution has quantiles μ_α for $0<\alpha<1$. For now we suppose that y_t is observed although later we allow it to be a residual from some estimated model.

Our null hypothesis is that some conditional quantiles are time invariant, which can be written more formally as: for some $\alpha \in A \subseteq [0,1]$

$$\mathrm{E}[\psi_\alpha(y_t - \mu_\alpha)|\mathscr{F}_{t-1}] = 0 \quad \text{a.s.}, \tag{1}$$

where $\psi_\alpha(x) = \alpha - 1(x<0)$ denotes the check function, while $\mathscr{F}_{t-1} = \sigma(y_{t-1}, y_{t-2}, \ldots)$. One could call y_t a *quantilegale*, or in the special case where $\alpha = \frac{1}{2}$, a *mediangale*. Under this hypothesis, if you are below the unconditional α-quantile today, the chance is no more than α that you will be below it tomorrow. In the absence of this property there is obviously some predictability in the process. We can distinguish between the cases where the hypothesis is about a particular quantile, about a set of quantiles, or when it is about all quantiles. The latter hypothesis is obviously much harder to satisfy, and is equivalent to y_t being i.i.d. We are initially just going to consider the single α case, although later we generalize to the interval case. In the empirical work we look at a number of quantiles simultaneously. The hypothesis (1) is also strong with regard to the size of the information set $\mathscr{F}_{t-1}$; our approach will be to test finite-dimensional implications.

Compare (1) with the usual weak form efficient markets hypothesis that for some μ,

$$\mathrm{E}[y_t - \mu|\mathscr{F}_{t-1}] = 0. \tag{2}$$

It could be that the median is time invariant but that the mean is time varying or vice versa. In the first case, the market is not efficient according to the usual definition, but there is no predictability in the median. In the second case, the market is efficient but there is predictability in the median. In general, there is no necessary relation between (2) and (1). However, under (conditional) symmetry of y_t there is a one-to-one relationship between (2) and (1) with $\alpha = \frac{1}{2}$; in practice, symmetry can be approximately true for many financial series (reference) so that evidence about the median reflects also on the mean.

In view of the efficient markets story (2), why should we be interested in quantiles and not means? First, note that there is a sort of fair bet aspect to (1) that is relevant to investors—if you are positioned on α-type investments, you want to know whether the best current guess of outcome is always still in the α end of the distribution or more or less likely given current information. Second, there is a powerful statistical reason for preferring quantiles over moments, robustness—there is some evidence that high-frequency data especially may not have moments beyond the second and maybe not even that many. In such cases, the meaning of (2) is uncertain. Third, quantiles also have a role to play in decision-making. Bassett et al. (2003) discuss portfolio allocation under Choquet decision theory in which expected utility is distorted to accentuate the likelihood of the least favourable outcomes, so for example the individual might want to maximize $\alpha^{-1}\int_0^{\alpha} u(\mu_\gamma)\,d\gamma$, where $u(\cdot)$ is the utility function and α is a particular probability, instead of expected utility which can be written $\int_0^1 u(\mu_\gamma)\,d\gamma$. In this case the quantiles $\{\mu_\gamma : 0 \leqslant \gamma \leqslant \alpha\}$ are of interest to investors rather than mean/variance. Fourth, quantiles are at the centre of the concept of Value at Risk. Engle and Manganelli (2004) propose a class of models that makes the conditional quantiles time varying through past observations of y and past values of the conditional quantiles themselves. They also proposed some specification tests of their models based on essentially (1).

The null hypothesis (1) is quite broad and includes many dependent processes. For example, suppose that

$$y_t = \mu_\alpha + \varepsilon_t \sigma_t, \tag{3}$$

where ε_t has conditional α-quantile equal to zero, while σ_t^2 is some volatility process: stationary and measurable with respect to $\mathscr{F}_{t-1}$. This corresponds to the 'semi-strong' GARCH models discussed in Drost and Nijman (1993). The process (3) is quite general, since we do not restrict σ_t^2. Consider the 'strong' case where ε_t are i.i.d. with α-quantile zero for some single α. When ε_t has a symmetric distribution and $\alpha = \frac{1}{2}$ this would include the standard strong GARCH process, and the process y_t is consistent with the usual efficient markets hypothesis. The process (3) satisfies (1) even when there is considerable dependence in the process through σ_t^2. It is a more general straw man than the traditional i.i.d. assumption or even independent increments assumption, cf. CLM, Section 2.1.[4]

In the 'strong' case, if $\mu_\alpha = 0$, then, no matter what value $\mathrm{E}(y_t)$ takes, the sign sequence is independent over time, which is strictly speaking contrary to the finding (1) of Christoffersen and Diebold (2002), i.e., there is not necessarily volatility induced dependence.[5] However, if (3) holds for some quantile α, then the conditional quantile at another α' is generally time varying, so that there is volatility induced dynamics in some other quantile hit. Hong and Chung (2003) replace μ_α by some fixed threshold c.[6] Note that even if (3) is satisfied, then the sequence $1(y_t < c)$ will be predictable for any $c \neq \mu_\alpha$.

These facts tend to support the fundamental point of the volatility induced dependence in semi-strong models. However, in the semi-strong version of (3) at some α there is no implication of about the behaviour of other conditional quantiles, because the distribution of ε_t can be time varying and so one could obtain a cancellation to make the other conditional quantiles of y_t time invariant. This may appear to be a pathological case, but the following example illustrates that it need not be.

Consider the class of quantile autoregression processes of Koenker and Xiao (2003) which can (in the first order case) be represented as

$$y_t = a(U_t) + b(U_t) y_{t-1}, \tag{4}$$

[4]Note that σ_t^2 may not be a conditional variance in this case. See Koenker and Zhao (1996) for discussion of estimation of quantiles in the presence of ARCH effects.

[5]For the most part they work with conditional normality, which makes the mean equal to the median. We note that it may be dangerous to work with normality as an assumption in this case where one statistical reason for looking at signs is related to their robustness with respect to moments.

[6]Actually, they scale the fixed c by an estimated standard deviation. However, they do not take account of this estimation in their distribution theory.

where $b(.)$ is a monotonic function and U_t is standard uniform. When $a(u) = \Phi^{-1}(u)$ and $b(u) = b$ with $|b| < 1$ this would be a standard Gaussian autoregression. Note that the conditional quantile function of $y_t|y_{t-1}$ is $q(\alpha) = a(\alpha) + b(\alpha)y$, while the conditional mean function is $\mathrm{E}[a(U)] + \mathrm{E}[b(U)]y$, and conditional variance $\mathrm{var}[a(U)] + \mathrm{var}[b(U)]y^2 + 2\,\mathrm{cov}(a(U), b(U))y$. By varying the functions a, b one can get quite rich behaviour for y_t, as has been discussed in Koenker and Xiao (2003). Provided $\mathrm{var}[b(U)] > 0$, there is time-varying volatility; if $\mathrm{E}[b(U)] \neq 0$ there is time-varying conditional mean. Clearly, it is possible that: (1) $\mathrm{E}[b(U)] = 0$ but $b(\frac{1}{2}) \neq 0$; (2) $\mathrm{E}[b(U)] \neq 0$ but $b(\frac{1}{2}) = 0$. In case (1) the mean is unpredictable but the median is predictable, while in case (2) the median is unpredictable but the mean is predictable. These cases only arise when b is a non-trivial function and in such cases typically all other quantiles will be predictable.[7] However, the following example suggests caution on this point. Suppose that $b(u) = 0$ for $u \in [\alpha, \beta]$ for some $0 < \alpha < \beta < 1$ but $b(u) \neq 0$ for all quantiles in the range $(0, 1)/[\alpha, \beta]$, and suppose that $\mathrm{E}[b(U)] = 0$. The implied process is mean unpredictable and is quantile unpredictable for all quantiles in the range $[\alpha, \beta]$, but is quantile predictable for all quantiles in the range $(0, 1)/[\alpha, \beta]$. It is also variance predictable, but it is clearly not meaningful to say that the quantile predictability is caused by the variance predictability. In conclusion, there is no necessary implication of predictability or lack thereof in one quantile to predictability or lack thereof in another quantile within this class of models. It is therefore of interest to evaluate predictability for a range of quantiles.

It is possible to obtain a formal test of some version of the hypothesis (1) along the lines of Hong and Chung (2003), but we eschew this approach for a less formal and graphical approach. Our method is to compute the correlogram of the quantile hits, which we call the quantilogram, and to display this along with pointwise confidence bands. This approach is in wide use already, when applied to series y_t, and is available in any reputable software unlike most of the formal non-parametric hypothesis tests.

3. Quantilogram

3.1. Definition and properties

Suppose that random variables $y_1, y_2, \ldots$ are from a stationary process whose marginal distribution has quantiles μ_α for $0 < \alpha < 1$. Define the quantilogram

$$\rho_{\alpha k} = \frac{\mathrm{E}[\psi_\alpha(y_t - \mu_\alpha)\psi_\alpha(y_{t+k} - \mu_\alpha)]}{\mathrm{E}[\psi_\alpha^2(y_t - \mu_\alpha)]}, \quad k = 1, 2, \ldots$$

of the stationary series y_t for any α. Under the null hypothesis (1) the population quantity

$$\mathrm{E}[\psi_\alpha(y_t - \mu_\alpha)\psi_\alpha(y_{t+k} - \mu_\alpha)] = \mathrm{E}[\psi_\alpha(y_t - \mu_\alpha)\mathrm{E}[\psi_\alpha(y_{t+k} - \mu_\alpha)|\mathscr{F}_{t+k-1}] = 0$$

for all k. Therefore, $\rho_{\alpha k}$ is zero for all k. Under the alternative hypothesis $\rho_{\alpha k}$ can take a variety of shapes across α and k; however, under mixing, $\rho_{\alpha k} \to 0$ as $k \to \infty$ for all α.

[7] Likewise one can have variance predictability but not median (sign) predictability, and one can also have median predictability but not variance predictability.

Compared with the correlogram of y_t itself, $\rho_{\alpha k}$ measures a different type of association. In its favour, it is robust to the non-existence of moments: no moments are required of y_t for the existence of $\rho_{\alpha k}$ or for the distribution theory of estimators of it. This means it can be used in circumstances where the usual correlogram is suspect. Davies and Mikosch (1998) and Mikosch and Stărică (2000) have recently studied the behaviour of sample correlograms for time series processes with heavy tails and conditional heteroskedasticity. They show that when there are insufficient marginal moments the sample autocorrelation of a stationary mixing process can have a random limit, so is inconsistent in general. Runde (1997) examines the behaviour of the Box–Pierce statistic in the heavy tailed (stable) case and finds convergence to some stable limit (albeit at a slower rate than usual) under the null hypothesis of i.i.d. assuming at least one moment and no conditional heteroskedasticity. He provides alternative critical values for this test. Note that by contrast the quantilogram does not even require a single moment and has standard convergence rates and asymptotic distributions.

The connection with models is not so tight as in the correlogram/linear process case, see Hill (2005) for some discussion on this. To illustrate this point we compute $\rho_{\alpha k}$ in an example. Suppose that

$$y_t = \psi y_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim \mathrm{N}(0,1)$. Then the ordinary correlation function is ψ^k, whereas $\rho_{\alpha k}$ is quite a complicated function of ψ and k. However, the quantilogram does appear to behave sensibly in this case: $\rho_{\alpha k}$ can be shown to be increasing in ψ for all k and to asymptote to zero and one as $\psi \to 0$ and $\psi \to 1$, respectively. Fig. 1 shows the shape of this function for

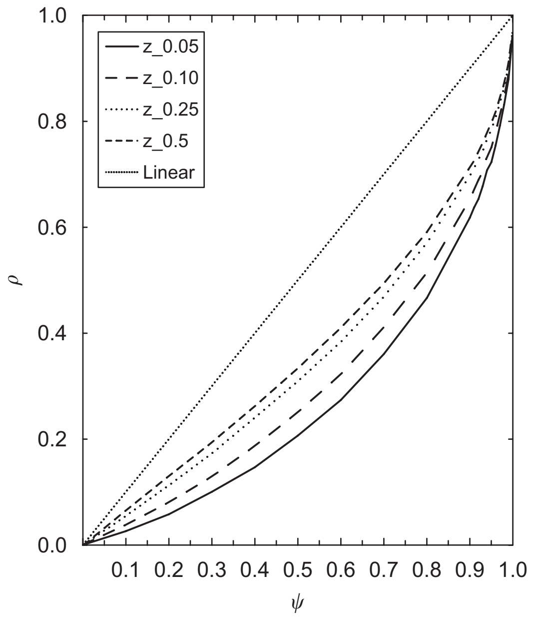

Fig. 1. Value of $\rho_{\alpha 1}$ for Gaussian AR(1) process as a function of AR parameter ψ for $\alpha = 0.5, 0.25, 0.10, 0.01$ in comparison with the value ψ.

$k = 1$ and $\alpha \in \{0.05, 0.1, 0.25, 0.5\}$ computed by simulation in comparison with the value for the ordinary correlation coefficient, ψ. The relationship is monotonic—the higher is ψ, the higher is $\rho_{\alpha 1}$—but there is some convexity that makes the curves lie below the 45° line. This convexity increases with the extremity of the quantile.

If instead ε_t were standard Cauchy, the usual correlogram is inconsistent, whereas the quantilogram gives sensible results. In this case, the function $\rho_{\alpha k}(\psi)$ is much closer to linear and there is much less difference across quantiles, as shown in the working paper version of this paper. For both these processes the graph of $\rho_{\alpha k}(\psi)$ as a function of k for given ψ shows similar sort of decay pattern to the usual correlogram and so conveys the right type of information about the dependence of the process.

3.2. Estimation

We compute the empirical counterpart of $\rho_{\alpha k}$. We first estimate μ_α by the sample quantile $\widehat{\mu}_\alpha$. We interpret this for mathematical reasons as a solution of the minimization problem:

$$\widehat{\mu}_\alpha = \arg\min_{\mu \in \mathbb{R}} \sum_{t=1}^{T} \rho_\alpha(y_t - \mu),$$

where $\rho_\alpha(x) = x[\alpha - 1(x < 0)]$. Then let

$$\widehat{\rho}_{\alpha k} = \frac{\sum_{t=1}^{T-k} \psi_\alpha(y_t - \widehat{\mu}_\alpha)\psi_\alpha(y_{t+k} - \widehat{\mu}_\alpha)}{\sqrt{\sum_{t=1}^{T-k} \psi_\alpha^2(y_t - \widehat{\mu}_\alpha)}\sqrt{\sum_{t=1}^{T-k} \psi_\alpha^2(y_{t+k} - \widehat{\mu}_\alpha)}}, \quad k = 1, 2, \ldots, T-1 \tag{5}$$

for any $\alpha \in (0, 1)$. Note that $-1 \leqslant \widehat{\rho}_{\alpha k} \leqslant 1$ for any α, k because this is just a sample correlation based on data $\psi_\alpha(y_t - \widehat{\mu}_\alpha)$.

4. Asymptotic properties

We now discuss the asymptotic properties of $\widehat{\rho}_{\alpha k}$. The usual Taylor series expansions cannot be applied because ψ_α is not a smooth function. A number of papers have established the properties of quantile estimators in a variety of settings, including Bassett and Koenker (1978), Bloomfield and Steiger (1983), and Pollard (1991), although we were not able to find the definitive reference suitable precisely for our purposes, and have consequently derived the theory from first principles. Our main result is provided under an assumption that is consistent with the set-up of Christoffersen and Diebold (2002), but is not the most general possible. Under our conditions we are able to derive a relatively simple limiting distribution and to propose a model free approach to inference. In Section 5 we extend our results to allow for much more general sampling schemes. The more complicated limiting distribution we obtain there requires a more complicated strategy for conducting inference, and we propose a method for doing this.

We assume:

Assumption 1. (a) $\{y_t : t = 1, \ldots, T\}$ is a strictly stationary and ergodic sequence (b) y_t has bounded unconditional density $f_y(\cdot)$ with respect to Lebesgue measure and has α-quantile μ_α and $f_y(\mu_\alpha) > 0$.

Under this assumption, we have the following Bahadur representation that is needed to discuss the asymptotic behaviour of the quantilogram:

Lemma 1. *Suppose Assumption* 1 *holds. Then, we have*

$$\sqrt{T}(\widehat{\mu}_\alpha - \mu_\alpha) = \frac{1}{f_y(\mu_\alpha)} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_\alpha(y_t - \mu_\alpha) + \mathrm{o_p}(1).$$

Next, we discuss the asymptotic property of $\widehat{\rho}_{\alpha k}$ under the null hypothesis. Here, we focus on the scale process (3) and strengthen Assumption 1.

Assumption 2. (a) Suppose that y_t satisfies (3) and satisfies Assumption 1(a). (b) $\{\varepsilon_t : t = 1, \ldots, T\}$ are i.i.d. with bounded density $f_\varepsilon(\cdot)$ with respect to Lebesgue measure and have α-quantile zero and $f_\varepsilon(0) > 0$. (c) σ_t^2 is strictly stationary and measurable with respect to $\mathscr{F}_{t-1}$ and $0 < \mathrm{E}[\sigma_t^{-1}] < \infty$.

Theorem 2. *Suppose Assumption* 2 *holds. Then, for* $p = 1, 2, \ldots$, *we have*

$$\sqrt{T}\widehat{\rho}_\alpha^{(p)} = \sqrt{T} \begin{bmatrix} \widehat{\rho}_{\alpha 1} \\ \vdots \\ \widehat{\rho}_{\alpha p} \end{bmatrix} \xrightarrow{\mathrm{d}} \mathrm{N}(0, V_\alpha^{(p)}) \quad \textit{where } V_\alpha^{(p)} = (V_{\alpha jk}),$$

$$V_{\alpha jk} = \begin{cases} 1 + \left(\dfrac{\mathrm{E}\left[\psi_\alpha(y_t - \mu_\alpha)\frac{1}{\sigma_{t+j}}\right]}{\mathrm{E}[\psi_\alpha^2(y_t - \mu_\alpha)]\mathrm{E}\left[\frac{1}{\sigma_t}\right]} \right)^2 \alpha(1-\alpha), & j = k, \\[2ex] \dfrac{\mathrm{E}\left[\psi_\alpha(y_t - \mu_\alpha)\frac{1}{\sigma_{t+k}}\right]\mathrm{E}\left[\psi_\alpha(y_t - \mu_\alpha)\frac{1}{\sigma_{t+j}}\right]}{\mathrm{E}[\psi_\alpha^2(y_t - \mu_\alpha)]\mathrm{E}^2\left[\frac{1}{\sigma_t}\right]}, & j \neq k. \end{cases}$$

The asymptotic variance does not explicitly depend on $f_y(\mu_\alpha)$—there has been a cancellation of this quantity from the two places where it shows up. Nevertheless, the asymptotic variance is quite complicated since it generally depends on the process σ_t^2. In some special cases, the correction factor in $V_{\alpha jj}$ due to the estimation of μ_α (this is the complicated term in parentheses) is zero. For example, when $\alpha = \frac{1}{2}$: if ε_t is symmetric about zero and σ_{t+k}^2 is an even function of ε_t (as in GARCH processes), then

$$\mathrm{E}\left[\psi_\alpha(\varepsilon_t)\frac{1}{\sigma_{t+k}}\right] = 0, \tag{6}$$

and $V_{\alpha jk} = 1$ if $j = k$ and 0 else. In this case, the estimation of a model for σ_t^2 can be avoided.

This may be too restrictive a special case. First, this generally only applies to the case $\alpha = \frac{1}{2}$. Second, symmetric error/even σ_{t+k}^2 is often thought inappropriate for stock return data. In the more general case, given an estimated parametric model for σ_t^2, one can estimate $V_{\alpha kk}$ by an obvious plug in approach based on an estimate $\widehat{\sigma}_t^2$ of σ_t^2 computed from the model estimates. This will be consistent when the volatility model is correct, but not necessarily otherwise.

We next explore an approach that avoids the necessity of specifying a volatility model. Using the facts that $|\psi_\alpha(\varepsilon_t)| \leqslant \max\{\alpha, 1-\alpha\}$ and σ_t^{-1} is stationary, we have

$$\underline{V}_\alpha \equiv 1 \leqslant V_{\alpha kk} \leqslant 1 + \frac{[\max\{\alpha, 1-\alpha\}]^2}{\alpha(1-\alpha)} \equiv 1 + \overline{v}_\alpha \equiv \overline{V}_\alpha.$$

The upper bound is independent of k, like the usual Bartlett intervals for ordinary correlations. The upper bound increases and without limit as $\alpha \to 0, 1$, and so provides less information in such cases.

Under the null hypothesis (1) and (3), the quantilogram at any point lies within $\pm z_{\gamma/2}\sqrt{\overline{V}_\alpha/T}$ with probability asymptotically greater than $1-\gamma$. If the additional conditions (6) are satisfied, the interval can be shrank to $\pm z_{\gamma/2}\sqrt{1/T}$. We call the smaller band liberal and the larger one conservative. In effect this test is designed for the hypothesis that $\rho_{\alpha k} = 0$ versus $\rho_{\alpha k} \neq 0$ for some k, but has power against other alternatives. The nice feature is that the upper bound holds independently of the model for σ_t^2 and of the error distribution.

We next turn to omnibus tests, which require us to bound the covariances. Note that

$$|V_{\alpha jk}| = \left| \frac{1}{\alpha(1-\alpha)} \frac{\mathrm{E}\left[\psi_\alpha(\varepsilon_t)\frac{1}{\sigma_{t+k}}\right]\mathrm{E}\left[\psi_\alpha(\varepsilon_t)\frac{1}{\sigma_{t+j}}\right]}{\mathrm{E}^2\left[\frac{1}{\sigma_t}\right]} \right| \leqslant \overline{v}_\alpha,$$

although $V_{\alpha jk}$ itself can be positive or negative. The matrix $V_\alpha^{(p)}$ is dominated by a matrix $I + \overline{v}_\alpha ii^\top$ whose largest eigenvalue is $1 + p\overline{v}_\alpha$. Consider the omnibus test statistic

$$Q_p = T\widehat{\rho}_\alpha^{(p)\top}\widehat{\rho}_\alpha^{(p)} = T\sum_{k=1}^{p} \widehat{\rho}_{\alpha k}^2 \tag{7}$$

for any p. Then, $Q_p \leqslant [1 + p\overline{v}_\alpha] \times T\widehat{\rho}_\alpha^{(p)\top}[V_\alpha^{(p)}]^{-1}\widehat{\rho}_\alpha^{(p)}$. Let $\chi_\gamma^2(p)$ be the level γ critical value of a chi-squared(p) distribution. Under the null hypothesis, the rule:

reject at level γ if $Q_p > [1 + p\overline{v}_\alpha]\chi_\gamma^2(p)$

has asymptotic size less than or equal to γ. A liberal test can be constructed using the lower bound of one instead of $1 + p\overline{v}_\alpha$. The 'conservative' test is likely to be very conservative for even moderate p: the bound is clearly too large, since it violates summability of the covariance matrix as $p \to \infty$. In small samples or large p, practice, better performance might be expected if one uses the modified Box–Ljung-type statistic $Q_p = T(T+2)\sum_{j=1}^{p} \widehat{\rho}_{\alpha j}^2/(T-j)$.

In effect this test is designed for the hypothesis that $\rho_{\alpha 1} = \cdots = \rho_{\alpha p} = 0$ versus $\rho_{\alpha k} \neq 0$ for some $k \leqslant p$, but has power against other alternatives.

A related test can be based on the partial quantilogram, which can be defined in the same way as in Brockwell and Davis (1991, p. 102). Specifically, define $\widehat{\phi}_{\alpha k}$ for each k as $\widehat{\phi}_{\alpha k} = \widehat{\phi}_{\alpha kk}$, where

$$\begin{bmatrix} \widehat{\phi}_{\alpha k1} \\ \widehat{\phi}_{\alpha k2} \\ \vdots \\ \widehat{\phi}_{\alpha kk} \end{bmatrix} = \begin{bmatrix} 1 & \widehat{\rho}_{\alpha 1} & \cdots & \widehat{\rho}_{\alpha,k-1} \\ \widehat{\rho}_{\alpha 1} & 1 & \widehat{\rho}_{\alpha 1} & \cdots \\ \vdots & & & \\ \widehat{\rho}_{\alpha,k-1} & \widehat{\rho}_{\alpha,k-2} & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\rho}_{\alpha 1} \\ \widehat{\rho}_{\alpha 2} \\ \vdots \\ \widehat{\rho}_{\alpha,k} \end{bmatrix}.$$

Theorem 3. *Suppose Assumption* 2 *holds. Then, for* $p = 1, 2, \ldots,$ *we have*

$$\sqrt{T}\widehat{\phi}_{\alpha}^{(p)} = \sqrt{T}\begin{bmatrix} \widehat{\phi}_{\alpha 1} \\ \vdots \\ \widehat{\phi}_{\alpha p} \end{bmatrix} \xrightarrow{\mathrm{d}} \mathrm{N}(0, V_{\alpha}^{(p)}).$$

Define the portmanteau statistic

$$Q_p^* = T\widehat{\phi}_{\alpha}^{(p)\top}\widehat{\phi}_{\alpha}^{(p)} = T\sum_{k=1}^{p}\widehat{\phi}_{\alpha k}^{2}. \tag{8}$$

The same considerations apply as in the case of Q_p. That is, we can apply exactly the same bounding arguments either to the individual partial autocorrelations or the portmanteau statistic. Dufour and Roy (1986) suggest alternative portmanteau tests.

Dufour et al. (1998) establish various properties of the signogram (which corresponds to the case $\alpha = \frac{1}{2}$) under independent sampling: when the median is known, they provide a test whose null distribution is known exactly; when the median is estimated, they show that this test is asymptotically distribution free under independent observations. In the dependent stochastic scaling process (3), this property no longer holds.

4.1. Comparison with the DQQ test

Engle and Manganelli (2004) propose a dynamic conditional quantile (DQQ) test of the hypothesis (1). Their test involves running the regression

$$\psi_{\alpha}(y_t - \widehat{\mu}_{\alpha}) = \widehat{a}_0 + \widehat{a}_1\psi_{\alpha}(y_{t-1} - \widehat{\mu}_{\alpha}) + \cdots + \widehat{a}_k\psi_{\alpha}(y_{t-k} - \widehat{\mu}_{\alpha}) + error \tag{9}$$

(and perhaps including other variables on the right-hand side) and then testing whether the coefficients in this regression are jointly zero using the usual Wald quadratic form. They give the asymptotic distribution of the coefficients and a consistent estimator of the weighting matrix.

This approach is similar to ours; indeed, the population regression of $\psi_{\alpha}(y_t - \mu_{\alpha})$ on $\psi_{\alpha}(y_{t-1} - \mu_{\alpha})$ is exactly equal to $\rho_{\alpha 1}$. The regression approach achieves similar objectives to the quantilogram, and has advantages and disadvantages over the methods we propose. Specifically, the disadvantages of the regression approach are: (1) If k is large, the covariate matrix in the regression (9) can be rank deficient making the test infeasible, whereas the quantilogram can be computed for any $k < T$. In their application, Engle and Manganelli (2004) choose $k = 5$, whereas we take $k = 100$. Given the recent interest in long memory time series, it may be prudent to allow for long lags in theory and practice. (2) The second advantage of the quantilogram approach over the Wald test is the graphical element and the directional information conveyed. (3) The third advantage of our approach is that we are able to provide model-free bounds to the asymptotic covariance matrix in an important special case, whereas they rely on a smoothing-based estimator of the conditional error density in all cases. Of course this latter advantage only applies in the special case where there are no other estimated parameters; also it has less value in the extreme quantile case where the bounds become large. For this reason we develop below an approach based on estimating the asymptotic variance that is applicable under even more general conditions than we used below.

4.2. Extension to multivariate case

The quantilogram can easily be extended to the vector case, where it is of interest to detect directional predictability from one series to another. In the bivariate case, define the cross-quantilogram for two series $\{y_t, x_t\}$:

$$\rho_{\alpha\beta k} = \frac{\mathrm{E}[\psi_\alpha(y_t - \mu_\alpha)\psi_\beta(x_{t-k} - \mu_\beta)]}{(\mathrm{E}[\psi_\alpha^2(y_t - \mu_\alpha)]\mathrm{E}[\psi_\beta^2(x_t - \mu_\beta)])^{1/2}}$$

for $\alpha, \beta \in (0,1)$ and $k \in \{\dots, -1, 0, 1, \dots\}$. The leading case is when $\alpha = \beta$. In the vector case, $\rho_{\alpha\beta k}$ is not necessarily equal to $\rho_{\alpha\beta(-k)}$. One can measure the degree of temporal asymmetry in the relationship by $\rho_{\alpha\alpha k} - \rho_{\alpha\alpha(-k)}$.

The distribution theory presented above can be extended to the vector case.

5. Extension to estimated residuals and hit processes

We now generalize the setting to allow the data to be residuals from a previously estimated parametric model. This broadens the range of applicability of our procedures but at the cost that the bounding approach no longer works and we have to employ explicitly estimated asymptotic variances to compute consistent critical values.

Suppose that u_t are generalized residuals that depend on an unknown parameter $\beta_0 \in \mathbb{B} \subset \mathbb{R}^q$, i.e.,

$$u_t := u_t(\beta_0) = q(\varepsilon_t, \Omega_t, \beta_0),$$

where $\Omega_t \subset \{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, X_t, X_{t-1}, \dots\}$ denotes the information set at time t and $(\varepsilon_t, X_t^\top)^\top$ are random vectors that satisfy the assumptions below. We assume that the function $q(\varepsilon, \Omega_t, \beta)$ is continuous and strictly increasing in $\varepsilon \in \mathbb{R}$, so that the inverse function

$$q^{-1}(\cdot, \Omega_t, \beta) := g_t(\beta, \cdot)$$

is well defined. Note that $\varepsilon_t = g_t(\beta_0, u_t)$. We also have: $\forall \beta \in \mathbb{B}$, $\forall \mu \in \mathbb{R}$ and $\forall t \geqslant 1$

$$q(\varepsilon, \Omega_t, \beta) < \mu \;\Leftrightarrow\; \varepsilon < g_t(\beta, \mu).$$

For simplicity, we sometimes denote $g_t(\beta, \mu) = g_t(\theta)$, where $\theta = (\beta^\top, \mu)^\top \in \Theta \equiv \mathbb{B} \times \mathbb{R}$.

Example 1 (*Nonlinear time series regression*). Let

$$y_t = m(\Omega_t, \beta_0) + \varepsilon_t,$$

where $\Omega_t = (\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, X_t, X_{t-1}, \dots)$. If we take $u_t = \varepsilon_t = y_t - m(\Omega_t, \beta_0)$, then

$$q(\varepsilon, \Omega_t, \beta) = \varepsilon + m(\Omega_t, \beta_0) - m(\Omega_t, \beta),$$

$$g_t(\beta, \mu) = \mu + m(\Omega_t, \beta) - m(\Omega_t, \beta_0).$$

Example 2 (GARCH(1, 1)). Let

$$y_t = X_t^\top \delta_0 + \varepsilon_t \sigma_t,$$

$$\begin{aligned}\sigma_t^2 &= \gamma_{10} + \gamma_{20}\sigma_{t-1}^2 + \gamma_{30}(y_{t-1} - X_{t-1}^\top \delta_0)^2 \\ &\equiv \sigma_t^2(\beta_0),\end{aligned}$$

where $\beta_0=(\gamma_{10},\gamma_{20},\gamma_{30},\delta_0^\top)^\top, \mathrm{E}[\varepsilon_t|\mathscr{F}_{t-1}]=0$ and $\mathrm{E}[\varepsilon_t^2|\mathscr{F}_{t-1}]=1$. The parameters are assumed to satisfy $\gamma_{10}>0,\gamma_{20}>0,\gamma_{30}>0$, and $\gamma_{20}+\gamma_{30}<1$. Let $\Omega_t=(\varepsilon_{t-1},\varepsilon_{t-2},\ldots,X_t,X_{t-1},\ldots)$. If we take u_t to be the standardized residual, we have $u_t=\varepsilon_t=(y_t-X_t^\top\delta_0)/\sigma_t$. In this case,

$$q(\varepsilon,\Omega_t,\beta)=\frac{\varepsilon\sigma_t+X_t^\top(\delta_0-\delta)}{\sigma_t(\beta)},$$

$$g_t(\beta,\mu)=\frac{\sigma_t(\beta)\mu+X_t^\top(\delta-\delta_0)}{\sigma_t}.$$

The null hypothesis of interest is: for some $\alpha\in(0,1)$,

$$\mathrm{E}[\psi_\alpha(u_t(\beta_0)-\mu_\alpha)|\mathscr{F}_{t-1}]=0 \quad \text{a.s.},$$

where $\psi_\alpha(x)=\alpha-1(x<0)$ as before. This can be used as a way of evaluating the parametric model.

Let $\widehat{\beta}$ be an estimate of β_0 satisfying Assumption 3(h). We then estimate μ_α by

$$\widehat{\mu}_\alpha=\arg\min_{\mu\in\mathbb{R}}\sum_{t=1}^T\rho_\alpha(u_t(\widehat{\beta})-\mu), \tag{10}$$

where $\rho_\alpha(x)=x[\alpha-1(x<0)]$ and the sample quantilogram is now defined as in (5) with $u_t(\widehat{\beta})$ instead of y_t.

We need the following assumptions:

Assumption 3. (a) $\{(\varepsilon_t,X_t):t=1,\ldots,T\}$ is a strictly stationary and ergodic sequence.

(b) The distribution $F_u(u)$ of u_t has a Lebesgue density $f(u)$ such that $\inf_{c\leqslant\alpha\leqslant 1-c} f_u(F_u^{-1}(\alpha))>0$ for all $c\in(0,\frac{1}{2})$. The conditional distribution $F(\varepsilon|\mathscr{F}_{t-1})\equiv F_t(\varepsilon)$ of ε_t given $\mathscr{F}_{t-1}$ has α-quantile μ_α and has Lebesgue density $f(\varepsilon|\mathscr{F}_{t-1})\equiv f_t(\varepsilon)$, which, with probability one, satisfies $\sup_{\{0<F_t(\varepsilon)<1\}}f_t(\varepsilon)\leqslant C_{1t}$ and a Lipschitz condition $|f_t(x_1)-f_t(x_2)|\leqslant C_{2t}|x_1-x_2|$ for all $t\geqslant 1$, where $EC_{1t}^2<\infty$ and $EC_{2t}<\infty$.

(c) There exists a function $h_t=(h_{1t}^\top,h_{2t})^\top$ with $h_t:\Theta\longrightarrow\mathbb{R}^q\times\mathbb{R}$, measurable with respect to $\mathscr{F}_{t-1}$, such that for all $k<\infty$, for all $c\in(0,\frac{1}{2})$,

$$\max_{1\leqslant t\leqslant T}\sup_{c\leqslant\alpha\leqslant 1-c}\sup_{\|\Delta\|\leqslant k}\sqrt{T}|g_t(\theta_\alpha+T^{-1/2}\Delta)-g_t(\theta_\alpha)-T^{-1/2}h_t(\theta_\alpha)^\top\Delta|=o_p(1),$$

where $\theta_\alpha=(\beta_0^\top,\mu_\alpha)^\top$ and $\Delta\in\mathbb{R}^{q+1}$.

(d) For all $\Delta<\infty$ and $\theta\in\Theta, \mathrm{E}|g_t(\theta+T^{-1/2}\Delta)-g_t(\theta)|^2=o(1)$.

(e) $\max_{1\leqslant t\leqslant T}T^{-1/2}\sup_{c\leqslant\alpha\leqslant 1-c}\|h_t(\beta_0,\mu_\alpha)\|^2=o_p(1)$ and $\max_{1\leqslant t\leqslant T}\sup_{c\leqslant\alpha\leqslant 1-c}\|h_t(\beta_0,\mu_\alpha)\|=O_p(1)$ for all $c\in(0,\frac{1}{2})$.

(f) $\inf_{c\leqslant\alpha\leqslant 1-c}|h_{2t}(\beta_0,\mu_\alpha)|>0$ a.s. for all $t\geqslant 1$ for all $c\in(0,\frac{1}{2})$.

(g) $\|h_t(\beta_1,\mu_1)-h_t(\beta_2,\mu_2)\|\leqslant D_t\{\|\beta_1-\beta_2\|+|\mu_1-\mu_2|\}$ for all $\beta_1,\beta_2\in\mathbb{B}$ and $\alpha_1,\alpha_2\in\mathbb{R}$, where $\max_{1\leqslant t\leqslant T}D_t=O_p(1)$.

(h) $\sqrt{T}(\widehat{\beta}-\beta_0)=T^{-1/2}\sum_{t=1}^T\xi_t(\beta_0)+o_p(1)\xrightarrow{d}N(0,V_\beta)$, where the function $\xi_t(\beta):\Theta\rightarrow\mathbb{R}^q$ is $\mathscr{F}_t$-measurable and satisfies $\mathrm{E}[\xi_t(\beta_0)|\mathscr{F}_{t-1}]=0$ a.s. for all $t\geqslant 1$ and $V_\beta=\lim_{T\to\infty}\mathrm{var}(T^{-1/2}\sum_{t=1}^T\xi_t(\beta_0))$ exists.

We now briefly discuss Assumption 3. Assumption 3(c) imposes a smoothness condition on the function g_t. In Example 1, Assumption 3(c) is satisfied with $h_{1t}(\theta_\alpha)=\partial m(\Omega_t,\beta_0)/\partial\beta$

and $h_{2t}(\theta_\alpha) = 1$, provided $m(\Omega_t, \cdot)$ is twice continuously differentiable with $\max_{1 \leqslant t \leqslant T} T^{-1/2} \sup_{\beta \in B_0} \partial^2 m(\Omega_t, \beta)/\partial\beta\partial\beta^\top = o_p(1)$, where B_0 is a neighbourhood of β_0. If the underlying random variables are strictly stationary as is assumed in Assumption 3(a), the latter holds if $\mathrm{E} \sup_{\beta \in B_0} \|\partial^2 m(\Omega_t, \beta)/\partial\beta\partial\beta^\top\|^{2+\delta} < \infty$ for some $\delta > 0$ by Chebyshev's inequality. On the other hand, in the case of Example 2, Assumption 3(c) holds with

$$h_{1t}(\theta_\alpha) = \frac{\partial g_t(\theta_\alpha)}{\partial \beta} = \left(\frac{\mu_\alpha}{2\sigma_t^2}, \frac{\sigma_{t-1}^2 \mu_\alpha}{2\sigma_t^2}, \frac{\varepsilon_{t-1}^2 \sigma_{t-1}^2 \mu_\alpha}{2\sigma_t^2}, \frac{\sigma_t X_t - \gamma_{30}\varepsilon_{t-1}\sigma_{t-1}\mu_\alpha X_{t-1}}{\sigma_t^2} \right)^\top$$

and $h_{2t}(\theta_\alpha) = 1$, provided $\max_{1 \leqslant t \leqslant T} T^{-1/2} \sup_{c \leqslant \alpha \leqslant 1-c} \sup_{\beta \in B_0, \mu \in B_\alpha} \partial^2 g_t(\beta, \mu_\alpha)/\partial\theta\partial\theta^\top = o_p(1)$, where B_α is a neighbourhood of μ_α. With strict stationarity, the latter in turn holds if $\mathrm{E} \sup_{c \leqslant \alpha \leqslant 1-c} \sup_{\beta \in B_0, \mu \in B_\alpha} \|\partial^2 g_t(\beta, \mu_\alpha)/\partial\theta\partial\theta^\top\|^{2+\delta} < \infty$ for some $\delta > 0$.

Assumption 3(d) further impose a smoothness condition on g_t and assumes it is L_2-continuous. Assumptions 3(e)–(g) requires that $h_t(\cdot, \cdot)$ satisfies suitable moment conditions and smoothness condition. Note that Assumption 3(f) trivially holds in Examples 1 and 2, because $h_{2t}(\beta_0, \mu_\alpha) = 1$for all α. Finally, Assumption 3(h) assumes that $\widehat{\beta}$ satisfies a linear expansion and is satisfied by many $T^{1/2}$-consistent estimators of β_0 including the maximum likelihood estimator.

Under these assumptions, we first show that the quantile estimator $\widehat{\mu}_\alpha$ based on the estimator residuals has a Bahadur representation. For the statement of this theorem we need the following notation. Suppose that $A_n(\alpha)$ and $B_n(\alpha)$ be two stochastic processes with realization in $D(0,1)^q$, i.e., the space of cadlag functions on $(0,1)^q$. We shall write

$$A_n(\alpha) = B_n(\alpha) + o_p^*(1)$$

if

$$\sup_{c \leqslant \alpha \leqslant 1-c} \|A_n(\alpha) - B_n(\alpha)\| = o_p(1) \quad \text{for all } c \in (0, 1/2).$$

Theorem 4. *Suppose Assumption* 3 *holds. Then, we have* (i)

$$\begin{aligned} W_T(\alpha) &\equiv \sqrt{T}(\widehat{\mu}_\alpha - \mu_\alpha) \\ &= H_2(\alpha)^{-1} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \psi_\alpha(u_t(\beta_0) - \mu_\alpha) - H_1(\alpha)^\top \frac{1}{\sqrt{T}} \sum_{t=1}^T \xi_t(\beta_0) \right] + o_p^*(1), \end{aligned}$$

where $H(\alpha) = (H_1(\alpha)^\top, H_2(\alpha))^\top$, $H_1(\alpha) = \mathrm{E}[f_t(F_t^{-1}(\alpha))h_{1t}(\theta_\alpha)]$, *and* $H_2(\alpha) = \mathrm{E}[f_t(F_t^{-1}(\alpha)) h_{2t}(\theta_\alpha)]$. *Moreover, we have* (ii)

$$W_T(\alpha) \;\Rightarrow\; H_2(\alpha)^{-1}[B(\alpha) - H_1(\alpha)^\top \cdot v],$$

where $B(\alpha)$ *is the Brownian bridge on* $(0,1)$ *and* v *is the normal random variable* $\mathrm{N}(0, V_\beta)$.

We note from Theorem 5 that the asymptotic distribution of the sample quantile $\widehat{\mu}_\alpha$ now depends on that of $\widehat{\beta}$.

Of interest are hypotheses of the form previously considered, namely

$$\mathrm{H}_0 : \rho_{\alpha k} = 0, \quad k = 1, \ldots, p$$

for some α. As before, compute the quantilogram and the Box–Pierce statistics. For $k = 1, \ldots, p$, let

$$J_k(\alpha) = (J_{1k}(\alpha)^\top, J_{2k}(\alpha))^\top,$$

$$J_{1k}(\alpha) = \mathrm{E}[\psi_\alpha(u_t - \mu_\alpha) f_{t+k}(F_{t+k}^{-1}(\alpha)) h_{1,t+k}(\theta_\alpha)],$$

$$J_{2k}(\alpha) = \mathrm{E}[\psi_\alpha(u_t - \mu_\alpha) f_{t+k}(F_{t+k}^{-1}(\alpha)) h_{2,t+k}(\theta_\alpha)], \tag{11}$$

and

$$D_k(\alpha) = H_1(\alpha) J_{2k}(\alpha) - H_2(\alpha) J_{1k}(\alpha).$$

Then, the sample quantilogram has the following distribution under the null hypothesis:

Theorem 5. *Suppose Assumption* 3 *holds. Then, under the null hypothesis, we have: for* $p = 1, 2, \ldots,$

$$\sqrt{T}\widehat{\rho}_\alpha^{(p)} = \sqrt{T}\begin{bmatrix}\widehat{\rho}_{\alpha 1}\\ \vdots \\ \widehat{\rho}_{\alpha p}\end{bmatrix} \xrightarrow{\mathrm{d}} \mathrm{N}(0, \overline{V}_\alpha^{(p)}) \quad \text{where } \overline{V}_\alpha^{(p)} = (\overline{V}_{\alpha jk}),$$

$$\overline{V}_{\alpha jk} = \begin{cases} 1 + \dfrac{J_{2k}^2(\alpha)}{\alpha(1-\alpha)H_2^2(\alpha)} + \dfrac{D_k(\alpha)^\top V_\beta D_k}{[\alpha(1-\alpha)H_2(\alpha)]^2}, & j = k, \\ \dfrac{J_{2j}(\alpha) J_{2k}(\alpha)}{\alpha(1-\alpha)H_2^2(\alpha)} + \dfrac{D_j(\alpha)^\top V_\beta D_k}{[\alpha(1-\alpha)H_2(\alpha)]^2}, & j \neq k, \end{cases}$$

and V_β *is defined in Assumption* 3(h).

Compare the asymptotic variance $\overline{V}_\alpha^{(p)}$ with $V_\alpha^{(p)}$ in Theorem 2. The essential difference is the presence of the last term in $\overline{V}_{\alpha jk}$, which is due to estimation of the unknown parameter β_0. Furthermore, $\overline{V}_\alpha^{(p)}$ now depend on the (conditional) density $f_t(\cdot)$ of ε_t. Note that if ε_t are independent of $\mathscr{F}_{t-1}$, as is usually assumed in the "strong GARCH" case, then terms $J_{2k}^2(\alpha)/(\alpha(1-\alpha)H_2^2(\alpha))$ and $J_{2j}(\alpha)J_{2k}(\alpha)/(\alpha(1-\alpha)H_2^2(\alpha))$ no longer depend on f_t due to a cancellation as in Theorem 2. However, even in this case, the last terms in $\overline{V}_{\alpha jk}$ still depend on the density function. Therefore, an omnibus test needs an estimator of the density function in general.

We next discuss how to conduct inference in this case. Note that, although $\overline{V}_\alpha^{(p)}$ depends on the conditional density function $f_t(\cdot)$, which is a very complicated object, it only depends on it through an expectation over the joint distribution of the high-dimensional conditioning set. It follows that $\overline{V}_\alpha^{(p)}$ can be estimated much more accurately than $f_t(\cdot)$. For example, we can estimate $\overline{V}_\alpha^{(p)}$ by

$$\widehat{\overline{V}}_\alpha^{(p)} = (\widehat{\overline{V}}_{\alpha jk}),$$

where

$$\widehat{\overline{V}}_{\alpha jk} = \begin{cases} 1 + \dfrac{\widehat{J}_{2k}^2(\alpha)}{\alpha(1-\alpha)\widehat{H}_2^2(\alpha)} + \dfrac{\widehat{D}_k(\alpha)^\top \widehat{V}_\beta \widehat{D}_k}{[\alpha(1-\alpha)\widehat{H}_2(\alpha)]^2}, & j = k, \\ \dfrac{\widehat{J}_{2j}(\alpha)\widehat{J}_{2k}(\alpha)}{\alpha(1-\alpha)\widehat{H}_2^2(\alpha)} + \dfrac{\widehat{D}_j(\alpha)^\top \widehat{V}_\beta \widehat{D}_k}{[\alpha(1-\alpha)\widehat{H}_2(\alpha)]^2}, & j \neq k, \end{cases}$$

$$\widehat{H}_l(\alpha) = \frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{u}_t - \widehat{\mu}_\alpha) h_{lt}(\widehat{\beta}, \widehat{\mu}_\alpha) \quad \text{for } l = 1, 2,$$

$$\widehat{J}_{lk}(\alpha) = \frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{u}_{t+k} - \widehat{\mu}_\alpha)\psi_\alpha(\widehat{u}_t - \widehat{\mu}_\alpha) h_{l,t+k}(\widehat{\beta}, \widehat{\mu}_\alpha) \quad \text{for } l = 1, 2,$$

$$\widehat{D}_k(\alpha) = \widehat{H}_1(\alpha)\widehat{J}_{2k}(\alpha) - \widehat{H}_2(\alpha)\widehat{J}_{1k}(\alpha),$$

where $\widehat{V}_\beta$ is a consistent estimator of V_β, $\widehat{u}_t = u_t(\widehat{\beta})$ is an estimated residual, $K_b(\cdot) = K(\cdot/b)/b$, $K(\cdot) = 1(|\cdot| \leqslant 1)/2$ is a (uniform) kernel function and b is a bandwidth parameter such that $b = o(1)$ and $b^{-1} = o(T^{1/2})$ as $T \to \infty$. Using a standard argument (see, e.g., Powell, 1991, Proof of Theorem 3, pp. 380–381), it is not difficult to show that $\widehat{\overline{V}}_\alpha^{(p)}$ is consistent for $\overline{V}_\alpha^{(p)}$.

Define the omnibus test statistic to be

$$Q_p = T\widehat{\rho}_\alpha^{(p)\top}[\widehat{\overline{V}}_\alpha^{(p)}]^{-1}\widehat{\rho}_\alpha^{(p)} \tag{12}$$

for any p. Then, Theorems 5 and 6 imply that, under the null hypothesis, the rule:

reject at level γ if $Q_p > \chi^2_\gamma(p)$

will have asymptotic size equal to γ.

We are also interested in hypotheses of the form

$$\mathrm{H}_0 : \rho_{\alpha k} = 0 \quad \text{for all } \alpha \in \mathscr{A}, \quad k = 1, \ldots, p,$$

where $\mathscr{A}$ is some subset of $[0, 1]$. If the set $\mathscr{A} = [0, 1]$ then this is a very strong hypotheses. Of more interest is the case where $\mathscr{A}$ is a strict subset of the unit interval, i.e., $\mathscr{A} = [\alpha, \beta]$ with $0 < \alpha < \beta < 1$. For example, $\mathscr{A} = [\frac{1}{2}, 1]$ or $\mathscr{A} = [\frac{1}{4}, \frac{3}{4}]$: the former case corresponds to 'no upside dependence', while the latter hypothesis corresponds to 'no dependence in mid-range'. To test this hypothesis we examine functionals of the form $\sup_{\alpha\in\mathscr{A}} \sum_{k=1}^{p} \rho^2_{\alpha k}$ or $\int_{\mathscr{A}} \sum_{k=1}^{p} \rho^2_{\alpha k}\, \mathrm{d}\lambda(\alpha)$ for some measure λ. Consider the test statistics

$$Q_{1p} = T \sup_{\alpha\in\mathscr{A}} \sum_{k=1}^{p} \widehat{\rho}^2_{\alpha k},$$

$$Q_{2p} = T \int_{\mathscr{A}} \sum_{k=1}^{p} \widehat{\rho}^2_{\alpha k}\, \mathrm{d}\lambda(\alpha).$$

Let $(\widetilde{B}_1(\cdot)\ \cdots\ \widetilde{B}_p(\cdot)\ B(\cdot)\ v^\top)^\top$ be a mean zero Gaussian process on $(0, 1)$ with covariance function given by

$$C(\alpha_1, \alpha_2) = \begin{pmatrix} \delta_{12}^2 I_p & \mathbf{0}_{p\times 1} & \mathbf{0}_{p\times q} \\ \mathbf{0}_{1\times p} & \delta_{12} & \mathbf{0}_{1\times q} \\ \mathbf{0}_{q\times p} & \mathbf{0}_{q\times 1} & V_\beta \end{pmatrix},$$

where $\delta_{12} = \min\{\alpha_1, \alpha_2\} - \alpha_1\alpha_2$, I_p is the identity matrix of dimension p and $\mathbf{0}_{p\times q}$ is the zero matrix of dimension $p \times q$. Then, the asymptotic distributions of the test statistics Q_{1p} and Q_{2p} are given by functionals of the Gaussian process.

Theorem 6. *Suppose Assumption* 3 *holds. Then, under the null hypothesis, we have: for* $p = 1, 2, \ldots,$

$$Q_{1p} \Rightarrow T \sup_{\alpha \in \mathscr{A}} \sum_{k=1}^{p} D_{\alpha k}^2,$$

$$Q_{2p} \Rightarrow T \int_{\mathscr{A}} \sum_{k=1}^{p} D_{\alpha k}^2 \, \mathrm{d}\lambda(\alpha),$$

where

$$D_{\alpha k} = \frac{1}{\alpha(1-\alpha)} \widetilde{B}_k(\alpha) - \frac{J_{2k}(\alpha)}{H_2(\alpha)} B(\alpha) + \frac{D_k(\alpha)}{H_2(\alpha)} v.$$

Since the asymptotic null distribution in Theorem 6 depends on the true data generating processes, the critical values cannot be tabulated once and for all. However, the latter can be simulated by a resampling schemes such as a bootstrap or subsampling procedure, see Lahiri (2003) or Politis et al. (1999) for a recent survey.

6. Numerical results

6.1. Application

We investigate samples of daily, weekly, and monthly returns on the S&P500 from 1955 to 2002, a total of 11,893, 2464, and 570 observations, respectively. The daily data are quite

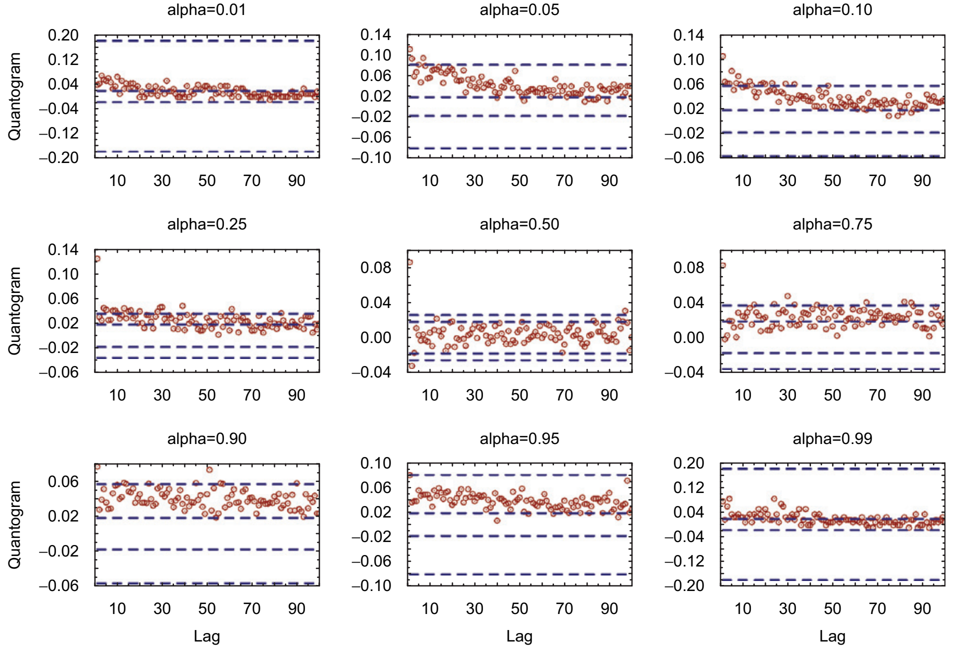

Fig. 2. S&P500 daily data: shown are the values of $\widehat{\rho}_{\alpha k}$ along with the liberal and conservative 95% confidence intervals.

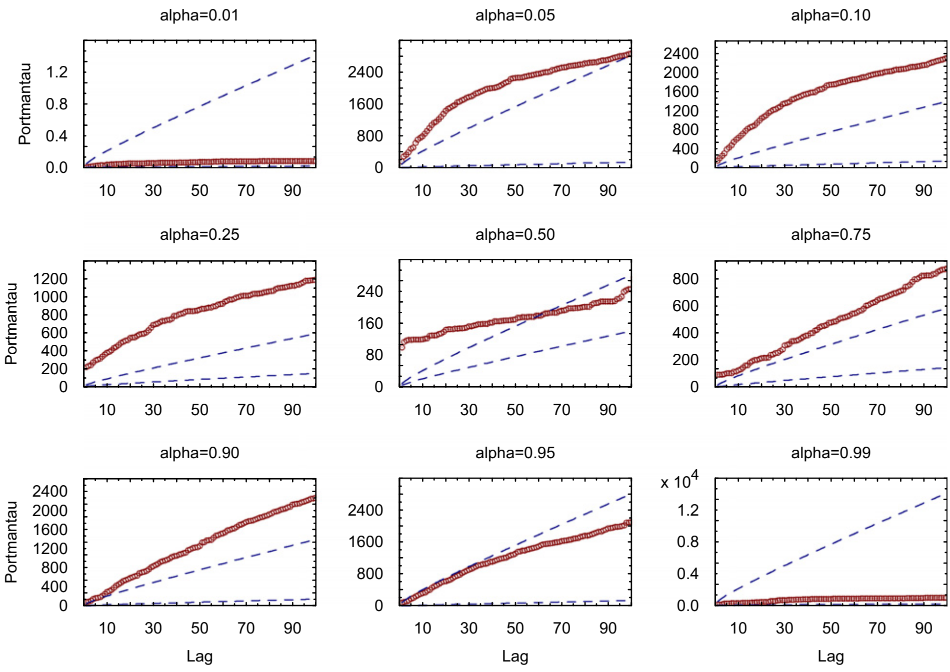

Fig. 3. S&P500 daily data: Box–Ljung test statistic Q_p for each lag p and quantile α along with 95% liberal and conservative critical values.

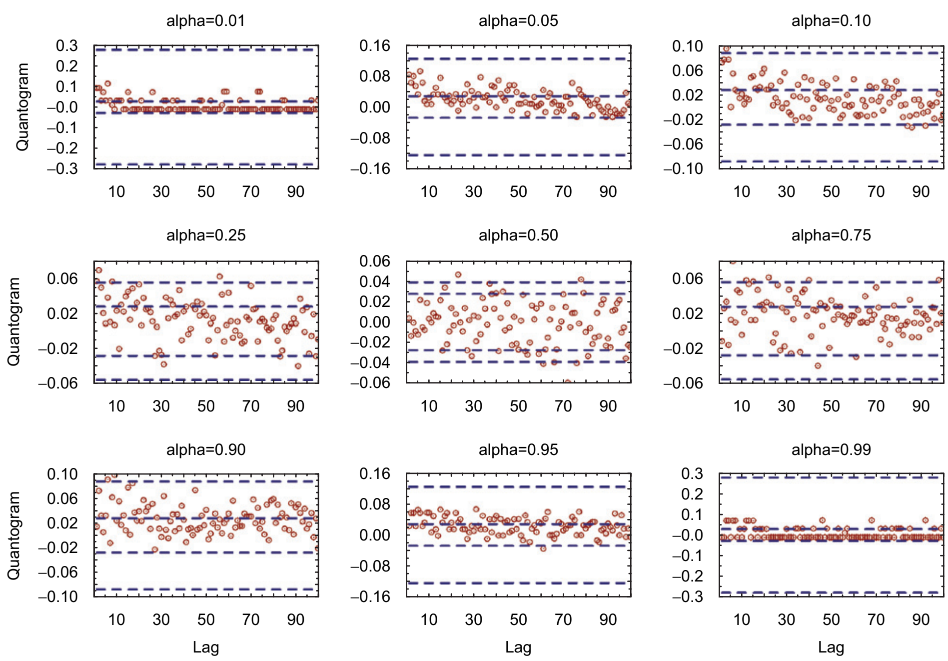

Fig. 4. S&P500 weekly data: shown are the values of $\widehat{\rho}_{\alpha k}$ along with the liberal and conservative 95% confidence intervals.

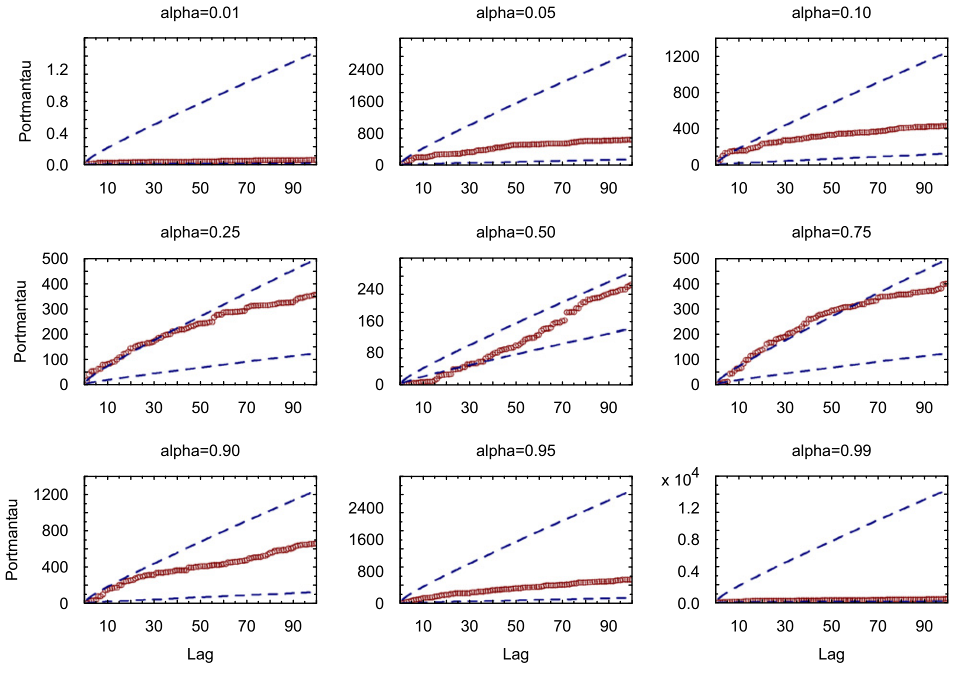

Fig. 5. S&P500 weekly data: Box–Ljung test statistic Q_p for each lag p and quantile α along with 95% liberal and conservative critical values.

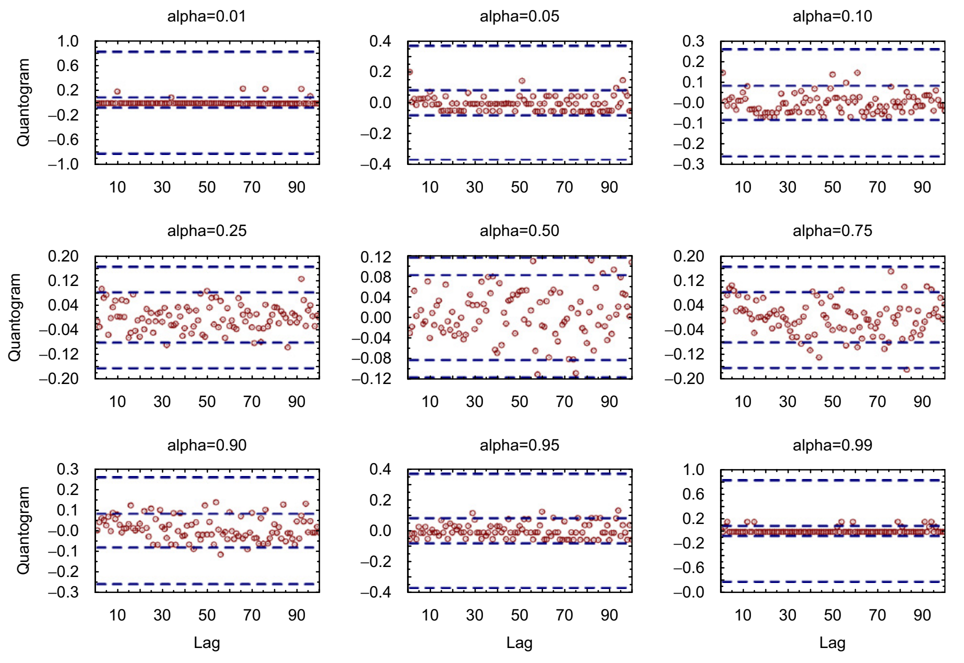

Fig. 6. S&P500 monthly data: shown are the values of $\widehat{\rho}_{\alpha k}$ along with the liberal and conservative 95% confidence intervals.

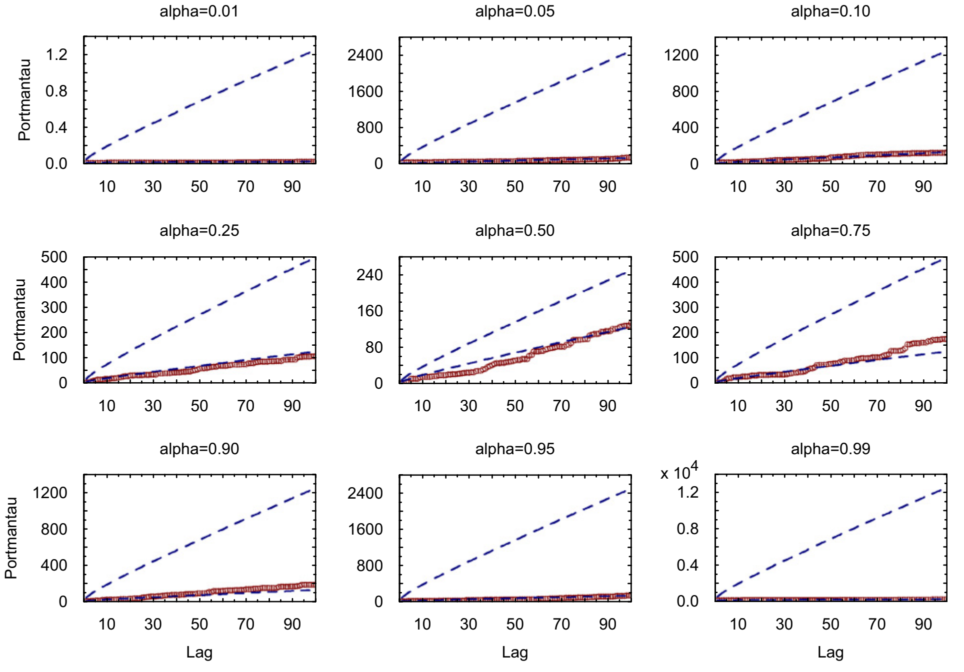

Fig. 7. S&P500 monthly data: Box–Ljung test statistic Q_p for each lag p and quantile α along with 95% liberal and conservative critical values.

heavy tailed and there is evidence that the fourth moment does not exist for the daily data, which makes analysis of the correlation function of the squared series, for example, suspect. These data were used in Linton and Mammen (2005) where a full set of sample moments and parametric estimates can be found.

In Figs. 2, 4, 6 we give the quantilogram for quantiles in the range $\alpha = 0.01$ to 0.99 and out to $k = 100$ lags. We also show the 95% confidence intervals (centred at zero) based on the lower and upper bound. There seems to be some evidence of predictability, but it depends on the data frequency and in some cases on which confidence interval you use. The evidence of predictability is strongest in the daily data. For the daily data the first median autocorrelation is positive and significant, while the second is negative and significant.[8] Thereafter there is not much individual evidence of predictability in the median. In the lower quantiles, $\alpha = 0.05$ and 0.10 especially, there are many individually significant correlations, all of which are positive. Thus when we have large negative losses in one period there is an increased likelihood of having large negative losses in the next period. In the corresponding higher quantiles of $\alpha = 0.95$ and 0.90 there are also some significant individual correlations, mostly positive, but evidence is less clear than for $\alpha = 0.05$ and 0.10. The picture is similar for the weekly data except that the evidence is less strong, in that although there are many exceedances of the liberal bands there are many fewer exceedences of the conservative bands. For monthly data there are very few observations outside the liberal confidence bands, and none outside the conservative ones.

[8] For comparison, the first autocorrelation of the data itself is 0.09 and the second is -0.04; the correlogram is very similar to the medianogram.

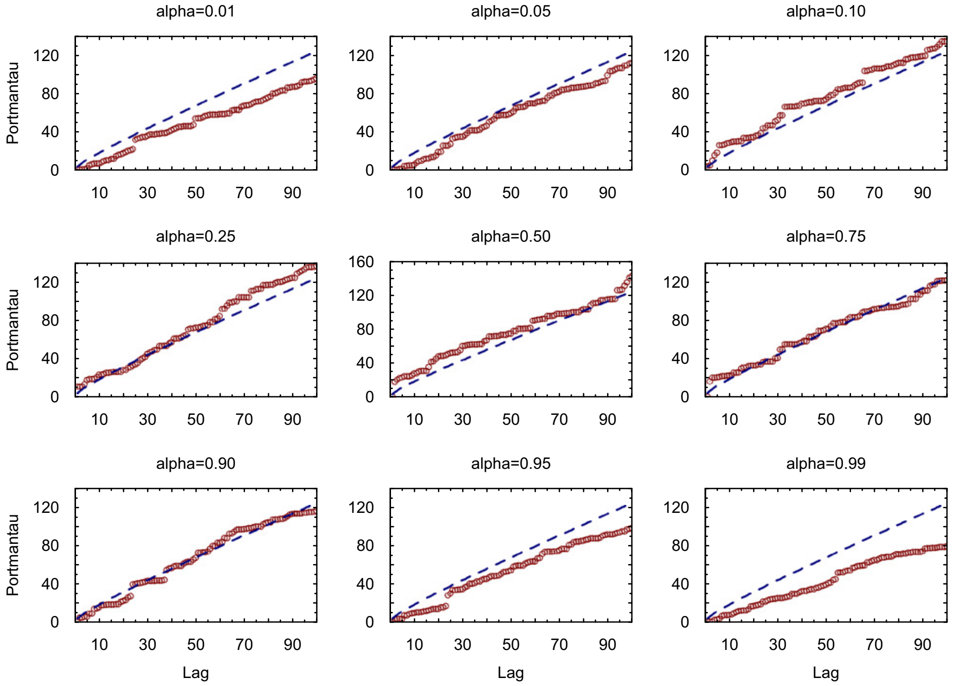

Fig. 8. Standardized residuals from AR(2)/AGARCH(1,1) model fit on S&P500 daily data: Box–Ljung test statistic Q_p for each lag p and quantile α along with 95% liberal critical values.

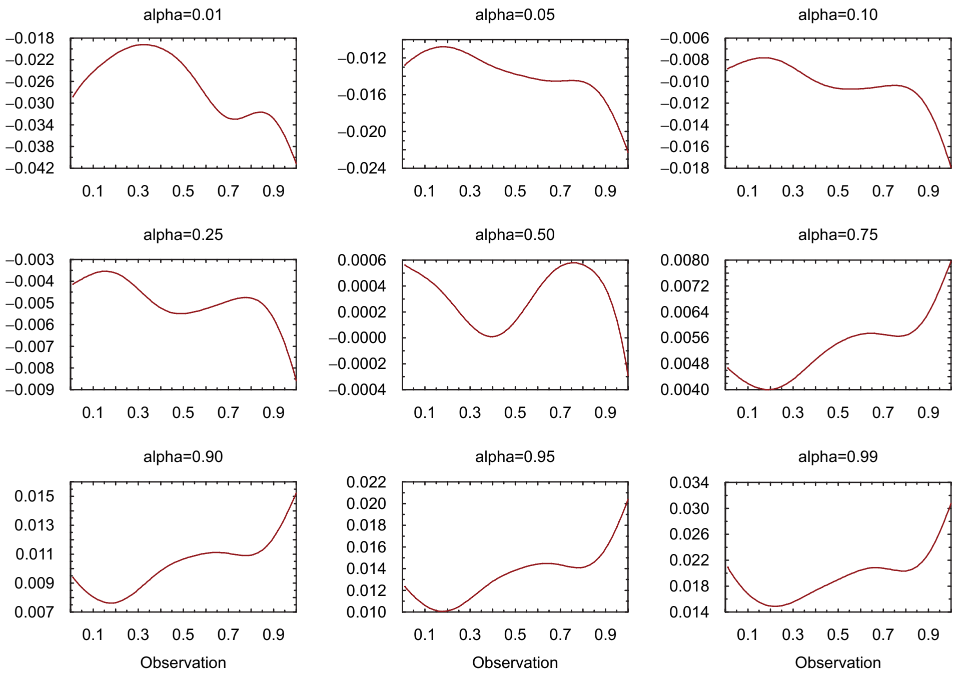

Fig. 9. Smoothed time-varying quantiles of daily returns.

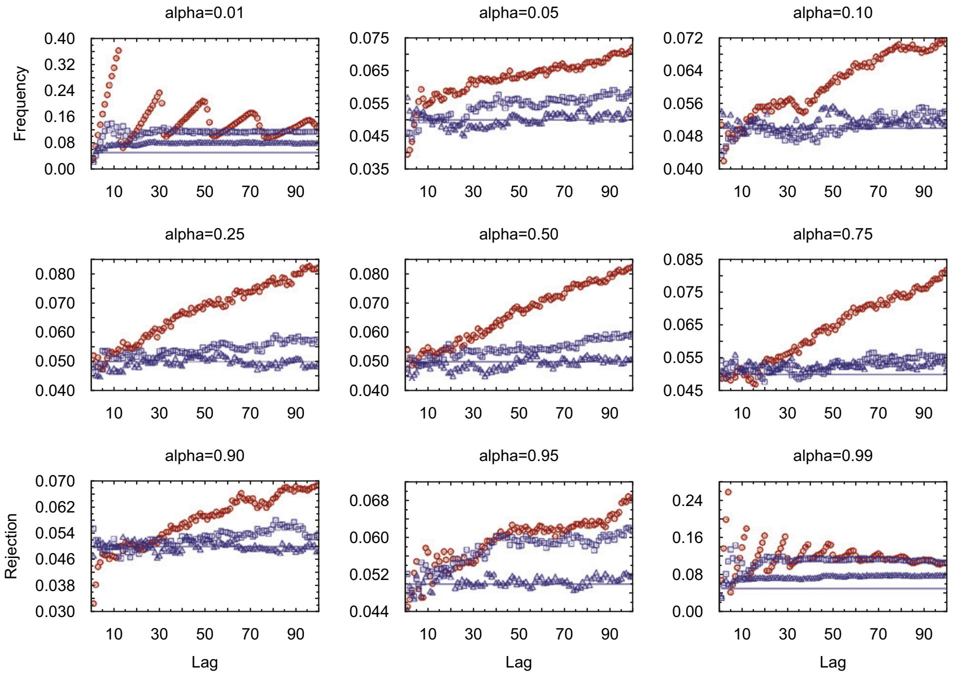

Fig. 10. Shows empirical rejection frequency of the test based on Box–Ljung statistics and liberal critical values at the 5% level against lag p. Design (1) with $n = 570$ (circles), $n = 2464$ (square), and $n = 11,893$ (triangle); $ns = 10,000$.

The portmanteau tests (we use the Box–Ljung versions throughout) in Figs. 3, 5, and 7 give stronger evidence of predictability: very pronounced in the daily data for all except the most extreme quantiles (where there is insufficient data). This is consistent with the finding in Hong and Chung (2003) that the magnitude of predictability is small for any given lag but large when combined across many lags. It is interesting that the $\alpha = 0.05$ case has much more pronounced dependence than the $\alpha = 0.95$ case.

Note that the upper bound confidence interval/critical value becomes very large in the extreme quantile case and is perhaps too pessimistic (Figs. 2–7).

The conclusion is that there is evidence of directional predictability that is not consistent with the pure strong quantile volatility model, that is, for no quantile does (3) appear consistent with the data. The predictability could be coming through mean effects or time varying higher moments as discussed in Christoffersen and Diebold (2002).

We next consider two possible explanations for the predictability. First, that the data are generated by the following AR(2)/AGARCH(1, 1) model

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t \sigma_t,$$

$$\sigma_t^2 = \gamma_0 + \gamma_1 \sigma_{t-1}^2 + \gamma_2 u_{t-1}^2 + \gamma_3 u_{t-1}^2 1(u_{t-1} < 0),$$

where $u_t = y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 y_{t-2}$. This model would generate the type of quantile dependence we have seen in the raw data. We check whether after whitening the data through this model, the dependence remains. We estimated this model on the daily data using the Gaussian QMLE. We then examine the standardized residuals from this estimated model. In Fig. 8 we show the portmanteau test statistic along with the lower bound critical values. Clearly, there is much less evidence of sign predictability left in the residuals, but there is still some.

The second model we consider is just that the conditional quantiles are time varying in a smooth fashion, i.e., $y_t = \mu_\alpha(t/T) + u_t$, where $\mathrm{E}[\psi_\alpha(u_t)|\mathscr{F}_{t-1}] = 0$ a.s., and $\mu_\alpha(.)$ is a smooth function. We compute the time smoothed quantiles using a nearest neighbour/kernel smoothing procedure and report the results below (for the daily data) in Fig. 9.

This suggests that the quantiles may have been changing over time, with lower quantiles decreasing and upper quantiles increasing, this implies a conditional distribution spreading out (the local interquartile range increases quite substantially over the period) over time. This change would perhaps provide an explanation of the positive autocorrelations in the lower and upper quantiles.

6.2. Simulation study

To ascertain the finite sample performance of our test statistics we conducted a small Monte Carlo experiment. We consider three designs: (1) $y_t \sim \mathrm{N}(0,1)$; (2) $y_t = \varepsilon_t \sigma_t$ with $\sigma_t^2 = \gamma_0 + \gamma_1 \sigma_{t-1}^2 + \gamma_2 y_{t-1}^2$, where $\varepsilon_t \sim \mathrm{N}(0,1)$ and parameters estimated from the

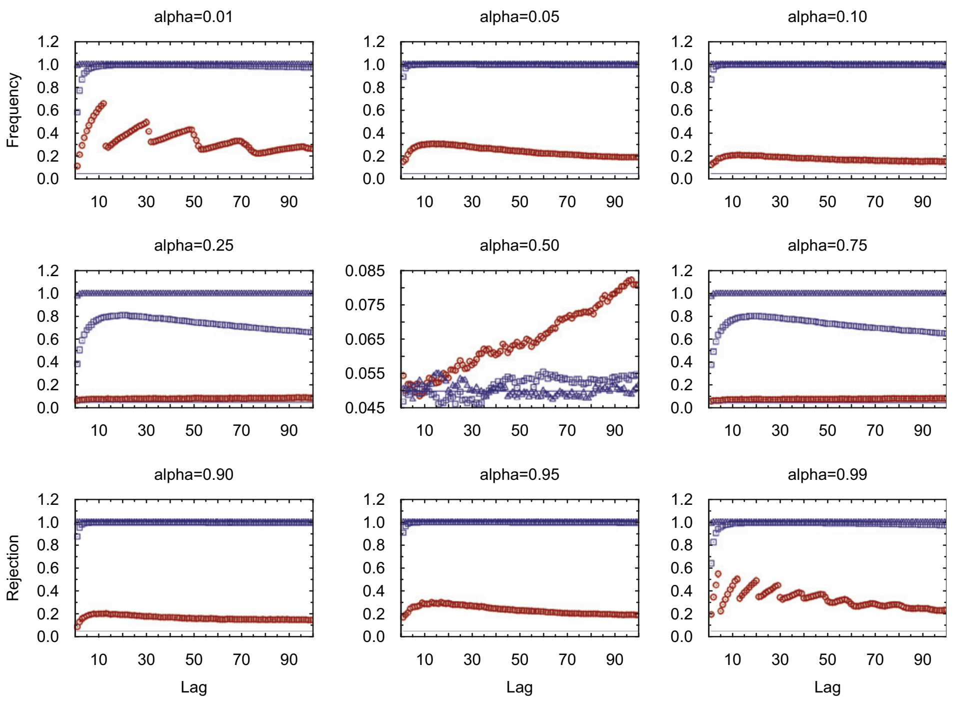

Fig. 11. Shows empirical rejection frequency of the test based on Box–Ljung statistics and liberal critical values at the 5% level against lag p. Design (2) with $n = 570$ (circles), $n = 2464$ (square), and $n = 11{,}893$ (triangle); $ns = 10{,}000$.

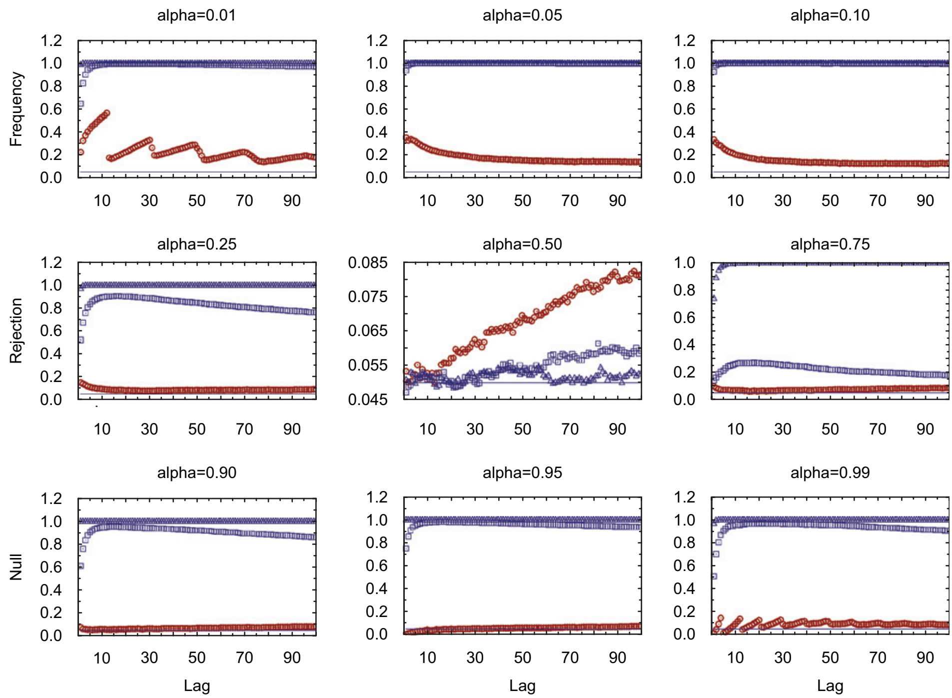

Fig. 12. Shows empirical rejection frequency of the test based on Box–Ljung statistics and liberal critical values at the 5% level against lag p. Design (3) with $n = 570$ (circles), $n = 2464$ (square), and $n = 11,893$ (triangle); $ns = 10,000$.

standardized daily return series; (3) y_t as before with $\sigma_t^2 = \gamma_0 + \gamma_1\sigma_{t-1}^2 + \gamma_2 y_{t-1}^2 + \gamma_3 y_{t-1}^2 1(y_{t-1} < 0)$ and parameters estimated from the standardized return series. We chose sample sizes $n = 11,893$, 2464, and 570 to match the sample sizes in the empirical work. Design (1) is completely under the null hypothesis for all quantiles, while (2) and (3) are under the null for the median but otherwise under the alternative. Design (2) satisfies (6), while Design (3) does not. The parameter values are given in Linton and Mammen (2005). For the daily data the parameter values are such that the second but not the fourth moment of y_t exists. This certainly would affect the performance of the correlogram of y_t^2 for which eight moments would be required for the standard theory but also the correlogram of y_t, see the earlier references for simulation studies.

We computed $ns = 10,000$ replications. Figs. 10–12 show the empirical rejection frequency of the Box–Ljung statistics for lags $p = 1, \ldots, 100$ and level 5% for the designs 1–3.

In Fig. 10, the rejection frequencies are close to 0.05 except for the 0.01 and 0.99 cases. There is a tendency to over-reject for the smaller sample sizes, and this seems to get worse with long lags. However, for the median with the largest sample size the rejection frequency is uniformly within simulation error of the nominal value for all lag lengths. In Figs. 11 and 12 we have similar performance at the median, as this is under the null hypotheses. For the other quantiles the rejection frequency approaches one in the largest sample size for all lags. This is as expected.

7. Conclusions

We have proposed using a standard time series methodology for measuring nonlinear dependence based on the correlogram of the quantile hits. We developed the distribution theory needed for the application of correlogram methods to the quantile case under general conditions. This methodology is used widely in econometrics for analyzing time series data and its computation is available in most standard packages. We think it is therefore likely to be a useful practical technique for analyzing directional predictability. See Garel and Hallin (1999) for related work based on ranks.

The empirical results show that it is important to take account of the effects of many small contributions from different lags as is done in the Box–Pierce-type statistics. We found very strong evidence of predictability in daily stock index returns at many different quantiles, and especially in the lower tails. This evidence remains, although it is much more muted, after fitting a time series model to the mean and variance of returns. The next question would be to model the underlying phenomena. There seem to be two approaches. In the first, like Engle and Manganelli (2004) and Koenker and Xiao (2003), you model directly the quantiles of the process y_t. In the second, like Rydberg and Shephard (2003), you model the binary hit sequences themselves. We think both approaches have merit.

Hill (2005) has recently proposed a co-relation measure for dependence in the extremes of heavy tailed dependent time series. Perhaps the theory for the quantilogram can also be extended to extreme quantiles $\alpha \to 0$ and $\alpha \to 1$ in a similar way but is beyond the scope of this paper.

Acknowledgements

We would like to thank J. Danielsson, J.M. Dufour, M. Hallin, Y. Hong, and R. Koenker, for interesting discussions. We thank Neil Shephard for providing a reference.

Appendix

Proof of Lemma 1. Lemma 1 is a special case of Theorem 4(a) in which there is no estimated parameter and hence can be proved using similar steps as in the proof of Theorem 4. □

Proof of Theorem 2. Define

$$\widetilde{\sigma}_k(\mu) = \frac{1}{T-k}\sum_{t=1}^{T-k} \psi_\alpha(y_t - \mu)\psi_\alpha(y_{t+k} - \mu),$$

$$\sigma_k(\mu) = \mathrm{E}[\psi_\alpha(y_t - \mu)\psi_\alpha(y_{t+k} - \mu)]$$

and let $\widetilde{\sigma}_k = \widetilde{\sigma}_k(\mu_\alpha)$ and $\sigma_k = \sigma_k(\mu_\alpha)$. By rearranging terms and a Taylor expansion, we have

$$\sqrt{T}(\widetilde{\sigma}_k(\widehat{\mu}_\alpha) - \sigma_k) = \sqrt{T}(\widetilde{\sigma}_k(\widehat{\mu}_\alpha) - \sigma_k(\widehat{\mu}_\alpha)) + \left(\frac{\partial}{\partial\mu}[\sigma_k(\mu)]\right)_{\mu=\mu^*} \sqrt{T}(\widehat{\mu}_\alpha - \mu_\alpha), \tag{13}$$

where μ^* lies between $\widehat{\mu}_\alpha$ and μ_α.

Using the stochastic equicontinuity arguments analogous to (but substantially simpler than) those used to verify (23) below, we have

$$\sqrt{T}(\widetilde{\sigma}_k(\widehat{\mu}_\alpha) - \sigma_k(\widehat{\mu}_\alpha)) = \sqrt{T}(\widetilde{\sigma}_k - \sigma_k) + o_p(1). \tag{14}$$

Consider the second term on the rhs of (13). We have

$$\begin{aligned}
\sigma_k(\mu) &= \mathrm{E}[\psi_\alpha(y_t - \mu)\psi_\alpha(y_{t+k} - \mu)] \\
&= \mathrm{E}\left[\alpha - 1\left(\varepsilon_t < \frac{\mu - \mu_\alpha}{\sigma_t}\right)\right]\left[\alpha - 1\left(\varepsilon_{t+k} < \frac{\mu - \mu_\alpha}{\sigma_{t+k}}\right)\right] \\
&= \mathrm{E}\left\{\left[\alpha - 1\left(\varepsilon_t < \frac{\mu - \mu_\alpha}{\sigma_t}\right)\right]\left[\alpha - F_\varepsilon\left(\frac{\mu - \mu_\alpha}{\sigma_{t+k}}\right)\right]\right\} \\
&= \mathrm{E}\left[\int_{-\infty}^{\varepsilon_t \leqslant (\mu-\mu_\alpha)/\sigma_t)} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{F_\varepsilon\left(\frac{\mu - \mu_\alpha}{\sigma_{t+k}}\right) - \alpha\right\} \prod_{j=0}^{k-1} f(\varepsilon_{t+j})\,\mathrm{d}\varepsilon_{t+j}\right] \\
&\quad - \alpha\mathrm{E}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{F_\varepsilon\left(\frac{\mu - \mu_\alpha}{\sigma_{t+k}}\right) - \alpha\right\} \prod_{j=0}^{k-1} f(\varepsilon_{t+j})\,\mathrm{d}\varepsilon_{t+j}\right]
\end{aligned}$$

by the law of iterated expectations. Therefore,

$$\begin{aligned}
\left(\frac{\partial}{\partial\mu}\sigma_k(\mu)\right)_{\mu=\mu_\alpha} &= f_\varepsilon(0)\mathrm{E}\left[(1(\varepsilon_t < 0) - \alpha)\frac{1}{\sigma_{t+k}}\right] \\
&\quad + f_\varepsilon(0)\mathrm{E}\left[\frac{1}{\sigma_t}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \{F_\varepsilon(0) - \alpha\} \prod_{j=1}^{k-1} f(\varepsilon_{t+j})\,\mathrm{d}\varepsilon_{t+j}\right] \\
&= -f_\varepsilon(0)\mathrm{E}\left[\psi_\alpha(\varepsilon_t)\frac{1}{\sigma_{t+k}}\right]
\end{aligned} \tag{15}$$

because $F_\varepsilon(0) = \alpha$. Furthermore, note that

$$f_u(0) = \left(\frac{\partial}{\partial x}\mathrm{E}\left[F_\varepsilon\left(\frac{x}{\sigma_t}\right)\right]\right)_{x=0} = f_\varepsilon(0)\mathrm{E}\left[\frac{1}{\sigma_t}\right]. \tag{16}$$

Therefore, by (13)–(16) and consistency of $\widehat{\mu}_\alpha$ for μ_α due to Lemma 1, we have

$$\sqrt{T}(\widetilde{\sigma}_k(\widehat{\mu}_\alpha) - \sigma_k) = \sqrt{T}(\widetilde{\sigma}_k - \sigma_k) - \frac{\mathrm{E}\left[\psi_\alpha(\varepsilon_t)\frac{1}{\sigma_{t+k}}\right]}{\mathrm{E}\left[\frac{1}{\sigma_t}\right]}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\alpha(\varepsilon_t) + o_p(1). \tag{17}$$

On the other hand, for each $k = 0, 1, \ldots$, and $\forall\varepsilon > 0$, we have

$$\begin{aligned}
&\Pr\left[\left|\frac{1}{T-k}\sum_{t=1}^{T-k}\psi_\alpha^2(y_{t+k} - \widehat{\mu}_\alpha) - \mathrm{E}\psi_\alpha^2(y_{t+k} - \mu_\alpha)\right| > \varepsilon\right] \\
&\leqslant \Pr\left[\sup_{\mu\in\Theta}\left|\frac{1}{T-k}\sum_{t=1}^{T-k}\{\psi_\alpha^2(y_{t+k} - \mu) - \mathrm{E}\psi_\alpha^2(y_{t+k} - \mu)\}\right| > \frac{\varepsilon}{2}\right] \\
&\quad + \Pr\left[|(\mathrm{E}\psi_\alpha^2(y_{t+k} - \mu))_{\mu=\widehat{\mu}_\alpha} - \mathrm{E}\psi_\alpha^2(y_{t+k} - \mu_\alpha)| > \frac{\varepsilon}{2}\right] + o(1) \\
&\quad \to 0,
\end{aligned} \tag{18}$$

where the inequality holds by triangle inequality and consistency of $\widehat{\mu}_\alpha$ for μ_α and the convergence to zero holds by the following arguments: the first term on the rhs of (18) converges to zero by the stochastic equicontinuity argument as in the proof of (23) and ergodic theorem. Next, the second term on the rhs of (18) is also o(1) by Assumption 2, Lemma 1 and a one term Taylor expansion, as desired.

Combining (17) and (18), we have

$$\sqrt{T}(\widehat{\rho}_{\alpha k}-\rho_{\alpha k}) = \frac{1}{\mathrm{E}[\psi_\alpha^2(y_t-\mu_\alpha)]}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\alpha(\varepsilon_t)\psi_\alpha(\varepsilon_{t+k}) - \frac{\mathrm{E}\left[\psi_\alpha(\varepsilon_t)\frac{1}{\sigma_{t+k}}\right]}{\mathrm{E}[\psi_\alpha^2(y_t-\mu_\alpha)]\mathrm{E}\left[\frac{1}{\sigma_t}\right]}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\alpha(\varepsilon_t)+\mathrm{o_p}(1).$$

Therefore, by a CLT for a martingale difference sequence, see, e.g., Hall and Heyde (1980, Corollary 3.1, p. 58), we have the desired asymptotic normality result of Theorem 2. The covariance between $\sqrt{T}(\widehat{\rho}_{\alpha k}-\rho_{\alpha k})$ and $\sqrt{T}(\widehat{\rho}_{\alpha j}-\rho_{\alpha j})$ for $j\neq k$ is determined by the second term, i.e.,

$$\mathrm{acov}(\sqrt{T}(\widehat{\rho}_{\alpha k}-\rho_{\alpha k}),\sqrt{T}(\widehat{\rho}_{\alpha j}-\rho_{\alpha j})) = \frac{\mathrm{E}\left[\psi_\alpha(\varepsilon_t)\frac{1}{\sigma_{t+k}}\right]\mathrm{E}\left[\psi_\alpha(\varepsilon_t)\frac{1}{\sigma_{t+j}}\right]}{\mathrm{E}[\psi_\alpha^2(y_t-\mu_\alpha)]\mathrm{E}^2\left[\frac{1}{\sigma_t}\right]}. \qquad \square$$

Proof of Theorem 4. We first verify part (i). Let $\theta_\alpha=(\beta_0^\top,\mu_\alpha)^\top$, $\Delta=(\Delta_1^\top,\Delta_2)^\top\in R^{q+1}$ and

$$V_T(\alpha,\Delta)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\alpha[\varepsilon_t-g_t(\theta_\alpha+T^{-1/2}\Delta)].$$

Setting $\widehat{\Delta}=\sqrt{T}(\widehat{\theta}_\alpha-\theta_\alpha)=(\sqrt{T}(\widehat{\beta}-\beta_0)^\top,\sqrt{T}(\widehat{\mu}_\alpha-\mu_\alpha))^\top=(\widehat{\Delta}_1^\top,\widehat{\Delta}_2)^\top$, we have

$$V_T(\alpha,\widehat{\Delta})=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\alpha[\varepsilon_t-g_t(\widehat{\theta}_\alpha)]=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\alpha[u_t(\widehat{\beta})-\widehat{\mu}_\alpha]=\mathrm{o}_\mathrm{p}^*(1), \tag{19}$$

using the continuity of the distribution of u_t, see e.g., Ruppert and Carroll (1980, Lemma A.2).

Let $A(c)=(c,1-c)$. We first establish that

$$\sup_{\alpha\in A(c)}\sup_{\|\Delta\|\leqslant M}|V_T(\alpha,\Delta)-V_T(\alpha,0)+H(\alpha)^\top\Delta|=\mathrm{o_p}(1) \tag{20}$$

for every $c\in(0,\frac{1}{2})$ and $M\in(0,\infty)$. Then the result of Theorem 4(i) holds by the following arguments. By Theorem 2.2 of Shorack (1979), $V_T(\cdot,0)$ converges weakly to Brownian bridge on $(0,1)$ and therefore $|V_T(\alpha,0)|$ is $\mathrm{O_p}(1)$ uniformly in $\alpha\in A(c)$. Given this result and (20), by using the method of Jurečková (1977, Lemma 5.2) (also see of Koenker and Zhao, 1996, Lemma A.4), we can show that for all $\varepsilon_0>0$, there exist η, M, and T_0 such that

$$\mathrm{P}\left(\sup_{\alpha\in A(c)}\inf_{|\Delta_2|>K}|V_T(\alpha,\widehat{\Delta}_1,\Delta_2)|<\eta\right)<\varepsilon_0 \quad \text{for } T\geqslant T_0. \tag{21}$$

Therefore, (19) and (21) imply that

$$\begin{aligned}
&\mathrm{P}\left(\sup_{\alpha\in A(c)} |\widehat{\Delta}_2|\geqslant K\right)\\
&\leqslant \mathrm{P}\left(\sup_{\alpha\in A(c)} |\widehat{\Delta}_2|\geqslant K, \sup_{\alpha\in A(c)} |V_T(\alpha,\widehat{\Delta})|<\eta\right) + \mathrm{P}\left(\sup_{\alpha\in A(c)} |V_T(\alpha,\widehat{\Delta})|\geqslant\eta\right)\\
&\leqslant \mathrm{P}\left(\sup_{\alpha\in A(c)} \inf_{|\Delta_2|>K} |V_T(\alpha,\widehat{\Delta}_1,\Delta_2)|<\eta\right) + \varepsilon_0\\
&\leqslant 2\varepsilon_0.
\end{aligned} \tag{22}$$

Part (i) of Theorem 5 now follows from (19), (20) and (22).

We now establish (20). Let

$$\begin{aligned}
Z_T(\alpha,\Delta) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} & 1(\varepsilon_t\leqslant g_t(\theta_\alpha+T^{-1/2}\Delta)) - F_t(g_t(\theta_\alpha+T^{-1/2}\Delta))\\
& - 1(\varepsilon_t\leqslant g_t(\theta_\alpha)) + \alpha.
\end{aligned}$$

For (20), it suffices to verify: for every $c\in(0,\frac{1}{2})$ and $M\in(0,\infty)$,

$$\sup_{\alpha\in A(c)} \sup_{\|\Delta\|\leqslant M} |Z_T(\alpha,\Delta)| = \mathrm{o_p}(1) \tag{23}$$

and

$$\sup_{\alpha\in A(c)} \sup_{\|\Delta\|\leqslant M} \left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\{F_t(g_t(\theta_\alpha+T^{-1/2}\Delta)) - F_t(g_t(\theta_\alpha))\} - H(\alpha)^\top\Delta\right| = \mathrm{o_p}(1). \tag{24}$$

We prove (23) and (24) by generalizing the results of Gutenbrunner and Jurečková (1992, Lemma 1), Bai (1994) and Koul (2002, Section 8). First, consider (23). We first establish a pointwise convergence result, i.e.,

$$Z_T(\alpha,\Delta) = \mathrm{o_p}(1) \tag{25}$$

for every given $\Delta\in\{\|\Delta\|\leqslant M\}$ and $\alpha\in A(c)$. Let

$$\eta_{Tt}(\alpha,\Delta) = 1(\varepsilon_t\leqslant g_t(\theta_\alpha+T^{-1/2}\Delta)) - 1(\varepsilon_t\leqslant g_t(\theta_\alpha)),$$

$$d_{Tt}(\alpha,\Delta) = g_t(\theta_\alpha+T^{-1/2}\Delta) - g_t(\theta_\alpha).$$

Then, for any $\varepsilon>0$,

$$\begin{aligned}
\mathrm{P}(|Z_T(\alpha,\Delta)|>\varepsilon) &\leqslant \frac{1}{T\varepsilon^2}\mathrm{E}\left\{\sum_{t=1}^{T}(\eta_{Tt}(\alpha,\Delta) - \mathrm{E}[\eta_{Tt}(\alpha,\Delta)|\mathscr{F}_{t-1}])\right\}^2\\
&\leqslant \frac{1}{T\varepsilon^2}\sum_{t=1}^{T} \mathrm{E}\{\mathrm{E}[\eta_{Tt}(\alpha,\Delta)^2|\mathscr{F}_{t-1}]\}\\
&\leqslant \frac{1}{\varepsilon^2}\,\mathrm{E}\{F_t(g_t(\theta_\alpha+T^{-1/2}\Delta)) - F_t(g_t(\theta_\alpha))\}\\
&\leqslant \frac{1}{\varepsilon^2}\,\mathrm{E}C_{1t}|d_{Tt}(\alpha,\Delta)| = \mathrm{o}(1),
\end{aligned} \tag{26}$$

where the first inequality holds by Chebyshev's inequality, the third inequality follows from $\mathrm{E}[1(X<a)-1(X<b)]^2 \leqslant |F(a)-F(b)|$ with F being the cdf of a random variable X, and the last convergence to zero holds by Cauchy–Schwarz inequality and Assumptions 3(b) and (d). This establishes (25).

We next show that (25) in fact holds uniformly over $\{\Delta : \|\Delta\| \leqslant M\}$ and $\alpha \in A(c)$ for every given $c \in (0, \frac{1}{2})$ and $M \in (0, \infty)$. Fix a $0 < \delta \leqslant M$. By Assumption 3(c), for any $\varepsilon_1 > 0$, there exists T_1 such that

$$\mathrm{P}\left(\max_{1 \leqslant t \leqslant T} \sup_{\alpha \in A(c)} \sup_{\|\Delta\| \leqslant \delta} |g_t(\theta_\alpha + T^{-1/2}\Delta) - g_t(\theta_\alpha) - T^{-1/2} h_t(\theta_\alpha)^\top \Delta| \leqslant T^{-1/2}\delta\varepsilon_1 \right) \geqslant 1 - \varepsilon_1 \quad \text{for all } T > T_1. \tag{27}$$

Now, fix $\varepsilon_1 > 0$ and let

$$A_T = \left\{ \sup_{\alpha \in A(c)} \sup_{\|\Delta_1 - \Delta_2\| \leqslant \delta} |d_{Tt}(\alpha, \Delta_1) - d_{Tt}(\alpha, \Delta_2)| \leqslant \delta B_{Tt}, 1 \leqslant t \leqslant T \right\}, \tag{28}$$

where

$$B_{Tt} = T^{-1/2} \left[\sup_{\alpha \in A(c)} \|h_t(\theta_\alpha)\| + 2\varepsilon_1 \right]. \tag{29}$$

Then, Assumption 3(c) and (27) imply that there exists T_1 such that

$$\mathrm{P}(A_T) > 1 - \varepsilon_1 \quad \text{for all } T > T_1. \tag{30}$$

Define for every $\lambda \in R$,

$$\widetilde{Z}_T(\alpha, \Delta, \lambda) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{ 1(\varepsilon_t \leqslant F_t^{-1}(\alpha) + d_{Tt}(\alpha, \Delta) + \lambda B_{Tt}) - F_t(F_t^{-1}(\alpha) + d_{Tt}(\alpha, \Delta) + \lambda B_{Tt}) - 1(\varepsilon_t \leqslant F_t^{-1}(\alpha)) + \alpha \}.$$

Note that $\widetilde{Z}_T(\alpha, \Delta, 0) = Z_T(\alpha, \Delta)$. Similarly to Koul (1991) and Bai (1994), we shall show that the desired result (23) is a consequence of the following result:

$$\sup_{\alpha \in A(c)} |\widetilde{Z}_T(\alpha, \Delta, \lambda)| = o_p(1) \quad \text{for every given } \Delta \text{ and } \lambda. \tag{31}$$

Let $\Gamma = \{\Delta : \|\Delta\| \leqslant M\}$. Due to its compactness, the set Γ can be partitioned into a finite number $\mathrm{N}(\delta)$ of subsets $\{\Gamma_1, \ldots, \Gamma_{\mathrm{N}(\delta)}\}$ such that the diameter of each subset is not greater than δ. Fix r and consider Γ_r. Choose $\Delta_\mathrm{r} \in \Gamma_\mathrm{r}$. Due to (30), for all $\Delta \in \Gamma_\mathrm{r}$, we have

$$\sup_{\alpha \in A(c)} |d_{Tt}(\alpha, \Delta) - d_{Tt}(\alpha, \Delta_\mathrm{r})| \leqslant \delta B_{Tt} \quad \forall 1 \leqslant t \leqslant T \tag{32}$$

with probability that goes to one (with probability tending to one). By monotonicity of the indicator function and the result (32),

$$Z_T(\alpha, \Delta) \leqslant \widetilde{Z}_T(\alpha, \Delta_\mathrm{r}, \delta) + T^{-1/2} \sum_{t=1}^{T} \{ F_t(F_t^{-1}(\alpha) + d_{Tt}(\alpha, \Delta_\mathrm{r}) + \delta B_{Tt}) - F_t(F_t^{-1}(\alpha) + d_{Tt}(\alpha, \Delta)) \}$$

and the lower bound is defined with δ replaced by $-\delta$, for all $\alpha \in A(c)$, where the inequality holds with probability tending to one. But

$$\begin{aligned}
&\sup_{\alpha\in A(c)} T^{-1/2}\left|\sum_{t=1}^{T} F_t(F_t^{-1}(\alpha)+d_{Tt}(\alpha,\Delta_{\mathrm{r}})+\delta B_{Tt})-F_t(F_t^{-1}(\alpha)+d_{Tt}(\alpha,\Delta))\right| \\
&\quad\leqslant T^{-1/2}\sum_{t=1}^{T} C_{1t}\left\{\sup_{\alpha\in A(c)}|d_{Tt}(\alpha,\Delta)-d_{Tt}(\alpha,\Delta_{\mathrm{r}})|+\delta B_{Tt}\right\} \\
&\quad\leqslant 2\delta T^{-1}\sum_{t=1}^{T} C_{1t}\left[\sup_{\alpha\in A(c)}\|h_t(\theta_\alpha)\|+2\varepsilon_1\right]=\delta \mathrm{O_p}(1),
\end{aligned}$$

where the $\mathrm{O_p}(1)$ is uniform for all $\Delta \in \Gamma$, the first inequality holds by Assumption 3(b) and the second inequality holds with probability tending to one, and the last equality holds by Assumptions 3(b) and (e)) and ergodic theorem. Therefore,

$$\begin{aligned}
\sup_{\alpha\in A(c)}\sup_{\|\Delta\|\leqslant M}|Z_T(\alpha,\Delta)| &\leqslant \max_{r\leqslant \mathrm{N}(\delta)}\sup_{\alpha\in A(c)}|\widetilde{Z}_T(\alpha,\Delta_{\mathrm{r}},\delta)| \\
&\quad+\max_{r\leqslant \mathrm{N}(\delta)}\sup_{\alpha\in A(c)}|\widetilde{Z}_T(\alpha,\Delta_{\mathrm{r}},-\delta)|+\delta\mathrm{O_p}(1).
\end{aligned} \tag{33}$$

Now, by choosing δ small enough, the $\delta\mathrm{O_p}(1)$ can be made small with arbitrarily large probability for all T sufficiently large. Therefore, the proof of (23) is complete once we verify (31).

To establish (31), we also use a chaining argument similar to the above. Now, fix Δ and λ. Divide the interval $A(c)$ into R sub-intervals by points $-c=\alpha_0<\alpha_1<\cdots<\alpha_R=c$. Let $\delta^*=2c/R$ be the length of each sub-interval. By Assumption 3(b), there exists a constant $C_0 \in (0,\infty)$ such that $\forall \alpha_1,\alpha_2 \in A(c)$,

$$|\mu_{\alpha_1}-\mu_{\alpha_2}|\leqslant \frac{1}{\inf_{\alpha\in A(c)} f_u(F_u^{-1}(\alpha))}|\alpha_1-\alpha_2|\leqslant C_0|\alpha_1-\alpha_2|, \tag{34}$$

where f_u and F_u denote the unconditional pdf and cdf of u_t, respectively. Now, fix $\varepsilon_2>0$ and let

$$A_T^*=\left\{\sup_{\alpha_1,\alpha_2\in A(c):|\alpha_1-\alpha_2|<\delta^*}|d_{Tt}(\alpha_1,\Delta)-d_{Tt}(\alpha_2,\Delta)|\leqslant B_{Tt}^*,\ 1\leqslant t\leqslant T\right\}, \tag{35}$$

where

$$B_{Tt}^*=T^{-1/2}[\delta^* C_0 D_t+2\varepsilon_2]. \tag{36}$$

Then, by Assumptions 3(c) and 3(g) and (34), there exists T_2 such that

$$\mathrm{P}(A_T^*)>1-\varepsilon_2 \quad \text{for all } T>T_2. \tag{37}$$

Define

$$\begin{aligned}
\widetilde{Z}_T^*(\alpha,\Delta,\eta,\lambda)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\{&1(\varepsilon_t\leqslant F_t^{-1}(\alpha)+d_{Tt}(\alpha,\Delta)+\eta B_{Tt}^*+\lambda B_{Tt}) \\
&-F_t(F_t^{-1}(\alpha)+d_{Tt}(\alpha,\Delta)+\eta B_{Tt}^*+\lambda B_{Tt})-1(\varepsilon_t\leqslant F_t^{-1}(\alpha))-\alpha\}.
\end{aligned}$$

With $\alpha_r < \alpha < \alpha_{r+1}$, by monotonicity of $1(\varepsilon_t \leqslant \cdot)$ and $F_t(\cdot)$ and (37), we have: with probability tending to one,

$$\begin{aligned}
&\sup_{\alpha \in A(c)} |\widetilde{Z}_T(\alpha, \Delta, \lambda)| \\
&\quad \leqslant \max_{r} |\widetilde{Z}_T^*(\alpha, \Delta, 1, \lambda)| + \max_{r} |\widetilde{Z}_T^*(\alpha, \Delta, -1, \lambda)| \\
&\qquad + \max_{r} \frac{1}{\sqrt{T}} \left| \sum_{t=1}^{T} \{ F_t(F_t^{-1}(\alpha_{r+1}) + d_{Tt}(\alpha_{r+1}, \Delta) + B_{Tt}^* + \lambda B_{Tt}) \right. \\
&\qquad \left. - F_t(F_t^{-1}(\alpha_{r+1}) + d_{Tt}(\alpha_{r+1}, \Delta) - B_{Tt}^* + \lambda B_{Tt}) \} \right| \\
&\qquad + \sup_{a,b \in A(c): |a-b| \leqslant \delta} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{ 1(\varepsilon_t \leqslant F_t^{-1}(a)) - a - 1(\varepsilon_t \leqslant F_t^{-1}(b)) + b \} \right|. \qquad (38)
\end{aligned}$$

The first two terms on the rhs of (38) are $o_p(1)$ since they are the maximum of finite number of $o_p(1)$ terms using an argument similar to (26) and Assumptions 3(b), (d), (e) and (g). The second term of (38) is also $o_p(1)$ by a mean-value expansion, Assumption 3(g) and the result (37). Finally, the last term of (38) is $o_p(1)$ due to the stochastic equicontinuity of empirical processes based on i.i.d. random variables (see, e.g., Pollard, 1984), since $F_t(\varepsilon_t)$ are i.i.d. U(0, 1), see Diebold et al. (1998). This now establishes (31) and hence (23).

Now it remains to show (24). The latter holds since

$$\begin{aligned}
&\sup_{\alpha \in A(c)} \sup_{\|\Delta\| \leqslant M} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{ F_t(g_t(\theta_\alpha + T^{-1/2}\Delta)) - F_t(g_t(\theta_\alpha)) \} - H(\alpha)^\top \Delta \right| \\
&\quad \leqslant \sup_{\alpha \in A(c)} \sup_{\|\Delta\| \leqslant M} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} d_{Tt}(\Delta, \alpha) \int_0^1 \{ f_t(F_t^{-1}(\alpha) + s d_{Tt}(\Delta, \alpha)) - f_t(F_t^{-1}(\alpha)) \} \, \mathrm{d}s \right| \\
&\qquad + \sup_{\alpha \in A(c)} \sup_{\|\Delta\| \leqslant M} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(F_t^{-1}(\alpha)) \{ d_{Tt}(\Delta, \alpha) - T^{-1/2} h_t(\theta_\alpha)^\top \Delta \} \right| \\
&\qquad + M \cdot \sup_{\alpha \in A(c)} \left\| \frac{1}{T} \sum_{t=1}^{T} f_t(F_t^{-1}(\alpha)) h_t(\theta_\alpha) - H(\alpha) \right\| \\
&\quad = o_p(1), \qquad (39)
\end{aligned}$$

where the first term on the rhs of (39) is $o_p(1)$ by Assumptions 3(b) and (e) and the result (30), the second term is $o_p(1)$ using Assumptions 3(b) and (c), and the last term goes to zero in probability by a uniform weak law of large numbers (see Andrews, 1992, Theorem 3) applied to Lipschitz continuous functions (in α) due to Assumptions 3(b) and (g). This establishes part (i) of Theorem 4.

Finally, part (ii) is a direct consequence of continuous mapping theorem using Assumption 3(h), Lemma 2.2 of Shorack (1979), and the fact that $F_t(\varepsilon_t)$ are i.i.d. U(0, 1). □

Proof of Theorem 5. The proof is similar to the proof of Theorem 2. Define

$$\widetilde{\sigma}_k(\theta) = \frac{1}{T-k}\sum_{t=1}^{T-k} \psi_\alpha(u_t(\beta) - \mu)\psi_\alpha(u_{t+k}(\beta) - \mu),$$

$$\sigma_k(\theta) = \mathrm{E}[\psi_\alpha(u_t(\beta) - \mu)\psi_\alpha(u_{t+k}(\beta) - \mu)]$$

and let $\widetilde{\sigma}_k = \widetilde{\sigma}_k(\theta_\alpha)$ and $\sigma_k = \sigma_k(\theta_\alpha)$. Using a one-term Taylor expansion as in (13), the result

$$\frac{\partial}{\partial\theta}\widetilde{\sigma}_k(\theta_\alpha) = -\mathrm{E}\left[\psi_\alpha(u_t - \mu_\alpha) f_{t+k}(F_{t+k}^{-1}(\alpha))\frac{\partial g_{t+k}(\theta_\alpha)}{\partial\theta}\right] = -J_k(\alpha),$$

and an argument similar to (23), we may show that: uniformly in $\alpha \in A(c)$,

$$\begin{aligned}\sqrt{T}(\widehat{\rho}_{\alpha k} - \rho_{\alpha k}) = {} & \frac{1}{\mathrm{E}[\psi_\alpha^2(u_t - \mu_\alpha)]}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \psi_\alpha(u_t - \mu_\alpha)\psi_\alpha(u_{t+k} - \mu_\alpha) \\ & - \frac{J_{2k}(\alpha)}{\mathrm{E}[\psi_\alpha^2(u_t - \mu_\alpha)]H_2(\alpha)}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \psi_\alpha(u_t - \mu_\alpha) \\ & + \frac{J_{2k}(\alpha)H_1(\alpha)^\top - J_{1k}(\alpha)^\top H_2(\alpha)}{\mathrm{E}[\psi_\alpha^2(u_t - \mu_\alpha)]H_2(\alpha)}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \xi_t(\beta_0) \\ & + o_p(1). \end{aligned} \tag{40}$$

Therefore, by a CLT for a martingale difference sequence, see, e.g., Hall and Heyde (1980, Corollary 3.1, p. 8), we have the desired asymptotic normality result of Theorem 5. □

Proof of Theorem 6. The results of Theorem 6 directly follow from the uniform convergence result (40) and continuous mapping theorem. □

References

Andrews, D.W.K., 1992. Generic uniform convergence. Econometric Theory 8, 241–257.

Bai, J., 1994. Weak convergence of the sequential empirical processes of residuals in ARMA models. Annals of Statistics 22, 2051–2061.

Bassett, G.W., Koenker, R., 1978. Asymptotic theory of least absolute error regression. Journal of the American Statistical Association 73, 618–622.

Bassett, G.W., Koenker, R., Korda, G., 2003. Pessimistic portfolio allocation and Choquet expected utility. http://www.econ.uiuc.edu/~roger/research/risk/risk.html

Bloomfield, P., Steiger, W.L., 1983. Least Absolute Deviations: Theory, Applications, and Algorithms. Birkhauser, Boston.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, second ed. Springer, Berlin.

Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton.

Christoffersen, P.F., Diebold, F.X., 2002. Financial asset returns, market timing, and volatility dynamics. Manuscript.

Cowles, A., Jones, H., 1937. Some a posteriori probabilities in stock market action. Econometrica 5, 280–294.

Davies, R.A., Mikosch, T., 1998. The sample autocorrelations of heavy tailed processes with applications to ARCH. Annals of Statistics 26, 2049–2080.

Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. Evaluation density forecasts with applications to financial risk management. International Economic Review 39, 863–883.

Drost, F.C., Nijman, T.E., 1993. Temporal aggregation of GARCH processes, Econometrica 61, 909–927.

Dufour, J.M., Roy, R., 1986. Generalized portmanteau statistics and tests of randomness. Communications in Statistics: Theory and Methods 15, 2953–2972.

Dufour, J.M., Hallin, M., Mizera, I., 1998. Generalized runs tests for heteroscedastic time series. Nonparametric Statistics 9, 39–86.

Engle, R.F., Manganelli, S., 2004. CAViaR: conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics 22, 367–381.

Garel, B., Hallin, M., 1999. Rank-based autoregressive order identification. Journal of the American Statistical Association 94, 1357–1371.

Gutenbrunner, C., Jurečková, J., 1992. Regression rank scores and regression quantiles. Annals of Statistics 20, 305–330.

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and Its Applications. Academic Press, New York.

Hill, J.B., 2005. Gaussian tests of extremal white noise for dependent, heterogeneous heavy tailed time series with an application. Manuscript, Florida International University.

Hong, Y., Chung, J., 2003. Are the directions of stock price changes predictable? Statistical Theory and Evidence. Manuscript.

Jurečková, J., 1977. Asymptotic relations of M-estimates and R-estimates in linear regression model. Annals of Statistics 5, 464–472.

Koenker, R., Xiao, Z., 2003. Quantile autoregression, unit roots and asymmetric interest rate dynamics. Manuscript.

Koenker, R., Zhao, Q., 1996. Conditional quantile estimation and inference for ARCH models. Econometric Theory 12, 793–814.

Koul, H.L., 1991. A weak convergence result useful in robust autoregression. Journal of Statistical Planning and Inference 29, 291–308.

Koul, H.L., 2002. Weighted Empirical Processes in Dynamic Nonlinear Models, second Ed. Springer, New York.

Lahiri, S.N., 2003. Resampling Methods for Dependent Data. Springer, New York.

Linton, O., Mammen, E., 2005. Estimating semiparametric ARCH(∞) models by kernel smoothing. Econometrica 73, 771–836.

Mikosch, T., Stărică, C., 2000. Limit theory for the sample autocorrelations and extremes of a GARCH(1, 1) process. Annals of Statistics 28, 1427–1451.

Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer, New York.

Pollard, D., 1984. Convergence of Stochastic Processes. Springer, New York.

Pollard, D., 1991. Asymptotics for least absolute deviation regression estimators. Econometric Theory 7, 186–199.

Powell, J., 1991. Estimation of monotonic regression under quantile restrictions. In: Barnett, W., et al. (Eds.), Nonparametric and Semiparametric Methods in Economics and Statistics. Cambridge University Press, Cambridge.

Runde, R., 1997. The asymptotic null distribution of the Box-Pierce Q-statistic for random variables with infinite variance: an application to German stock returns. Journal of Econometrics 78, 205–216.

Ruppert, D., Carroll, R.J., 1980. Trimmed least squares estimation in the linear model. Journal of the American Statistical Association 75, 828–838.

Rydberg, T.H., Shephard, N., 2003. Dynamics of trade-by-trade price movements: decomposition and models. Journal of Financial Econometrics 1, 2–25.

Shorack, G., 1979. Weighted empirical process of row independent r. v's with arbitrary df's. Statistica Neerlandica 35, 169–189.