Cleiton G. Taufemback, Victor Troster* and Muhammad Shahbaz

A Robust Test for Monotonicity in Asset Returns

https://doi.org/10.1515/jtse-2019-0068
Received December 2, 2019; accepted July 25, 2021

Abstract: In this paper, we propose a robust test of monotonicity in asset returns that is valid under a general setting. We develop a test that allows for dependent data and is robust to conditional heteroskedasticity or heavy-tailed distributions of return differentials. Many postulated theories in economics and finance assume monotonic relationships between expected asset returns and certain underlying characteristics of an asset. Existing tests in literature fail to control the probability of a type 1 error or have low power under heavy-tailed distributions of return differentials. Monte Carlo simulations illustrate that our test statistic has a correct empirical size under all data-generating processes together with a similar power to other tests. Conversely, alternative tests are nonconservative under conditional heteroskedasticity or heavy-tailed distributions of return differentials. We also present an empirical application on the monotonicity of returns on various portfolios sorts that highlights the usefulness of our approach.

Keywords: monotonicity tests, expected asset returns, heavy-tailed distributions, sign test, portfolio sorts

JEL Classification: C12, C58

1 Introduction

Many asset pricing models suggest a monotonic relationship between expected asset returns and certain underlying characteristics of an asset. For instance,

***Corresponding author**: **Victor Troster**, Department of Applied Economics, Universitat de les Illes Balears, Palma de Mallorca 07122, Spain, E-mail: victor.troster@uib.es. https://orcid.org/0000-0002-8798-5601
Cleiton G. Taufemback, Institute of Mathematics and Statistics, Universidade Federal do Rio Grande do Sul, Porto Alegre 90040-060, Brazil, E-mail: cleiton.taufemback@ufrgs.br
Muhammad Shahbaz, Center for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing 100081, China, E-mail: muhdshahbaz77@gmail.com

some models assume a monotonically increasing relationship between returns on a bond and time to maturity. A common procedure in the existing literature was to test for the significance of the difference in returns between the highest and lowest return characteristic. Nevertheless, this significance test is inaccurate since it ignores the relationship between expected returns and asset characteristics for intermediate return categories. Therefore, only a test that considers strictly monotonic relationships for all return characteristics can solve this problem.

Patton and Timmermann (2010) developed a monotonicity test that calculates individual t-statistics for expected return differentials over the entire range of return categories, where the alternative hypothesis states that all expected return differentials are positive. Nevertheless, the test proposed by Patton and Timmermann (2010) rules out a weakly increasing relationship under the null hypothesis; since it assumes that the relationship is weakly decreasing if it is not strictly increasing. As a result, their test does not successfully control the probability of a type 1 error. Romano and Wolf (2013) proposed monotonicity tests that also consider a monotonic relationship under the alternative hypothesis and that provide a correct size. However, their test statistic based on expected return differentials may be misleading when asset returns display conditional heteroskedasticity or are drawn from heavy-tailed distributions. Monte Carlo experiments show (on Section 3) that their test statistic has low power and an incorrect size under heavy-tailed distributions of the return differentials.

This paper proposes a robust test of monotonicity in asset returns that is valid under a general setting. We develop a test that allows for dependent data and is robust to conditional heteroskedasticity or heavy-tailed distributions of return differentials. Monte Carlo simulations illustrate that our test statistic is conservative for finite samples under all data-generating processes considered. The tests of Patton and Timmermann (2010) fail to control the type 1 error under all data-generating processes, while the test of Romano and Wolf (2013) is non-conservative under conditional heteroskedasticity or heavy-tailed distributions of the return differentials. Also, our test displays a similar power to other tests when the sample size is larger than 200 observations. Therefore, our method allows for testing for monotonicity in a variety of situations by extending the approaches of Patton and Timmermann (2010) and Romano and Wolf (2013).

It is important to consider heavy-tailed distributions in practice since many financial series display thick tails (see Cont 2001; Embrechts, Kluppelberg, and Mikosch 1997; Fama 1965; Gabaix 2009; Ibragimov and Walden 2007; Ibragimov, Ibragimov, and Kattuman 2013; Loretan and Phillips 1994; Rachev and Mittnik 2000). Mandelbrot (1963) initiated research on thick-tailed distributions, where a random variable R follows a distribution with tails displaying power-law decay such that $P(|R| > r) \sim r^{-\alpha}$, for $r > 0$, with a tail parameter $\alpha > 0$. In this instance,

$f(r) \sim g(r)$ denotes $f(r) = g(r)(1 + o(1))$ as $r \to \infty$. The tail parameter α defines the greatest order of finite moments of the random variable R. The mean of R is finite if and only if $\alpha > 1$, and the variance of R is finite if and only if $\alpha > 2$. Finiteness of first and second moments is crucial for obtaining an optimal diversification (Ibragimov and Walden 2007) and for implementing standard econometric methods. In addition, Ibragimov, Jaffee, and Walden (2009) report evidence that gains of diversification drop remarkably when distributions have thick tails. Scherer, Harhoff, and Kukies (2000) and Silverberg and Verspagen (2007) also found that financial returns on technological innovations have thick-tailed distributions with infinite means.

Rachev, Menn, and Fabozzi (2005) provides a review of research papers supporting heavy-tailed distributions for returns on equity and bonds. As considered in Ibragimov and Walden (2007), many empirical studies found that financial returns display heavy-tailed distributions (see Gabaix 2009; Gabaix et al. 2003, 2006; Ibragimov, Ibragimov, and Kattuman 2013; Jansen and de Vries 1991; Loretan and Phillips 1994; McCulloch 1996, 1997). In addition, such authors as Lux (1996), Guillaume et al. (1997), Gabaix et al. (2003), Gabaix (2009), and Ibragimov, Ibragimov, and Kattuman (2013) reported evidence that these estimates of the tail parameters are very much alike in different countries. As a result, financial returns may display undefined mean, infinite variance, or infinite fourth moments.

We develop a sign-based test that tests a null hypothesis of no systematic relationship against an alternative hypothesis of strict monotonicity. Sign-based tests are robust under observations with infinite variance (Capanu, Jones, and Randles 2006; Delgado and Velasco 2005; Gerard and Schucany 2007). Boldin, Simonova, and Tyurin (1997) proposed locally optimal nonparametric sign tests for linear models under independent errors. Coudin and Dufour (2009) developed sign-based inference for linear median regression models that allows for non normally distributed and serially correlated errors. Campbell and Dufour (1991, 1995, 1997) proposed nonparametric tests of conditional independence based on signed-rank statistics.

Our test has an alternative hypothesis in line with Patton and Timmermann (2010) and Romano and Wolf (2013). Under the alternative hypothesis, a strictly increasing monotonic relationship exists. Fama (1984) and Wolak (1987, 1989) developed the inverse procedure of testing the null hypothesis of a strictly monotonic relationship against the alternative hypothesis of no relationship. However, the null hypothesis of monotonicity may not be rejected because of the lack of power of the test. Therefore, this approach may provide poor statistical support for the associated hypothesis when the null is not rejected. Hence, we will overlook these tests in our paper.

The remainder of the paper is organized as follows. Section 2 explains our testing procedure. Section 3 presents Monte Carlo simulations results. Section 4 shows an application of our test to returns on portfolio sorted on firm characteristics. Finally, Section 5 concludes.

2 A Robust Test for Monotonicity in Asset Returns

Let $(r_{0,t}, r_{1,t}, \dots, r_{N,t})'$ be a vector of strictly stationary time series returns of dimension $N+1$ with T observations. As in Patton and Timmermann (2010) and Romano and Wolf (2013), we assume that the order of the return categories is predetermined and independent from the data. For each return category i, we define its expected return as μ_i, with an estimator $\hat{\mu}_i \equiv (1/T)\sum_{t=1}^{T} r_{i,t}$. Then, we denote the expected return vector as $\boldsymbol{\mu} \equiv (\mu_0,\ \mu_1, \dots,\ \mu_N)'$. We define the expected return differentials between the categories $i-1$ and i as $\Delta_i \equiv\ \mu_i -\ \mu_{i-1}$. Following the notation of Patton and Timmermann (2010) and Romano and Wolf (2013), let the vector of observed return differentials be $\mathbf{d}_t \equiv (d_{1,t}, d_{2,t}, \dots, d_{N,t})' = (r_{1,t} - r_{0,t}, r_{2,t} - r_{1,t}, \dots, r_{N,t} - r_{N-1,t})'$. Then, the associated vector of expected return differentials is $\boldsymbol{\Delta} \equiv (\mu_1 -\ \mu_0,\ \mu_2 -\ \mu_1, \dots,\ \mu_N -\ \mu_{N-1})' = (\Delta_1, \Delta_2, \dots, \Delta_N)'$. We can also write $\boldsymbol{\Delta} = E(\mathbf{d}_t)$, with an estimator $\hat{\Delta}_i \equiv (1/T)\sum_{t=1}^{T} d_{i,t}$. Therefore, we can test whether the expected returns $r_{i,t}$ are monotonically increasing over the $N+1$ categories by verifying whether $\Delta_i > 0$ holds for all $i = 1, \dots, N$.

Patton and Timmermann (2010) specify weakly decreasing return differentials under the null hypothesis against strictly increasing return differentials under the alternative hypothesis:

$$H_0 : \Delta_i \leq 0 \quad \text{for all } i \text{ versus } H_A : \min_{i=1,\dots,N} \Delta_i > 0, \tag{1}$$

since $\min_{i=1,\dots,N} \Delta_i > 0$ implies $\Delta_i > 0$ for all $i = 1, \dots, N$. Nevertheless, the alternative hypothesis of strict monotonicity in (1) is not the rejection of the null, when the parameter space for $\boldsymbol{\Delta}$ is $\mathbb{R}^N$, the N-dimensional Euclidian space. Romano and Wolf (2013) suggest partitioning $\mathbb{R}^N$ as $\mathbb{R}^N = R_1 \cup R_2 \cup R_3$, where

$$\begin{aligned} R_1 &\equiv \{r \in \mathbb{R}^n : r_i \leq 0 \quad \text{for all } i\}, \\ R_2 &\equiv \{r \in \mathbb{R}^n : r_i \leq 0 \quad \text{for some } i \text{ and } r_j > 0 \text{ for some other } j\}, \\ R_3 &\equiv \{r \in \mathbb{R}^n : r_i > 0 \quad \text{for all } i\}. \end{aligned} \tag{2}$$

Thus, we can specify the hypotheses in (1) as

$$H_0 : \Delta \in R_1 \text{ versus } H_A : \Delta \in R_3.$$

Therefore, Patton and Timmermann (2010) exclude *a priori* the possibility that $\boldsymbol{\Delta} \in R_2$, and the null hypothesis is rejected under the assumption that $\boldsymbol{\Delta} \in R_1 \cup R_3$. Ignoring *a priori* the possibility that $\boldsymbol{\Delta} \in R_2$ may lead to a false decision supportive of a strictly monotonic increasing relationship under the alternative hypothesis.

Romano and Wolf (2013) proposed a monotonicity test of expected asset returns allowing for $\Delta \in \mathbb{R}^N$ *a priori*, where the hypotheses are written as follows:

$$H_0 : \Delta \in R_1 \cup R_2 \text{ versus } H_A : \Delta \in R_3. \tag{3}$$

The hypotheses in (3) can be specified as

$$H_0 : \min_{i=1,\ldots,N} \Delta_i \leq 0 \text{ versus } H_A : \min_{i=1,\ldots,N} \Delta_i > 0. \tag{4}$$

The testing problem of (4) is composite since the null parameter space contains several parameters Δ_i. Thus, it is difficult to define the values of Δ_i for the sampling distribution under the null hypothesis. Romano and Wolf (2013) proposed three different methods to calculate their test critical values, but they recommended for practical purposes the so-called conservative test. Although the conservative test of Romano and Wolf (2013) has a correct size, it may have low power and incorrect finite-sample size under conditional heteroskedasticity or heavy-tailed distributions of the return differentials.

We draw statistical inferences based on sign tests. Let the vector of signs of return differentials be $\boldsymbol{\omega}_t \equiv (\omega_{1,t}, \omega_{2,t}, \ldots, \omega_{N,t})' = (\text{sign}(d_{1,t}), \text{sign}(d_{2,t}), \ldots, \text{sign}(d_{N,t}))'$, where sign$(r)$ is equal to $1/2$ if $r > 0$ and $-1/2$ otherwise. We assume that $\omega_{i,t}$ has no mass at 0, for all $i = 1, \ldots, N$. Thus, the associated vector of expected signs of return differentials is $\boldsymbol{\Omega} \equiv (\Omega_1, \Omega_2, \ldots, \Omega_N)' = (E(\omega_{1,t}), E(\omega_{2,t}), \ldots, E(\omega_{N,t}))'$, with an estimator $\hat{\Omega}_i \equiv (1/T)\sum_{t=1}^{T} \omega_{i,t} = \bar{\omega}_{i,t}$. We can also write $\boldsymbol{\Omega} \equiv E(\boldsymbol{\omega}_t)$. Let the minimum of expected signs of return differentials be $\Omega^* \equiv \min_{i=1,\ldots,N} \Omega_i$. Besides, let $m = \arg\min_{i=1,\ldots,N} \Omega_i$, and we assume $\Omega^* < \Omega^{*,-m}$ a.s. for $\Omega^{*,-m} = \min_{i=1,\ldots,N, i \neq m} \Omega_i$. Then, we propose to test the null hypothesis of no monotonicity of the return differentials as follows:

$$H_0 : \Omega^* \leq 0 \text{ versus } H_A : \Omega^* > 0, \tag{5}$$

where the hypotheses in (5) are equivalent to the testing problem in (3). The hypothesis testing in (5) provides a precise definition of the null space since the sign function restricts the range of values of Δ_i in the set $\{-1/2, 1/2\}$, for each $i = 1, \ldots, N$. Following Hansen (2005) and Romano and Wolf (2005), we develop a studentized version of Ω^* for testing H_0 in (5), which leads to an increased power of the test statistic. Let $\hat{\Omega}^*$ be the sample analog of Ω^*, where $\hat{\Omega}^* = \min_{i=1,\ldots,N} \hat{\Omega}_i$. Then, we define our test statistic as follows:

$$\hat{M} = \frac{\hat{\Omega}^*}{\hat{\sigma}_*}, \tag{6}$$

where $\hat{\sigma}_*$ is a consistent estimator of the standard error of $\hat{\Omega}^*$, $\hat{\sigma}_* \equiv \sqrt{\mathrm{Var}(\hat{\Omega}^*)}$.

Patton and Timmermann (2010) proposed a block-bootstrap version of their test statistic when the returns follow a time series process. Romano and Wolf (2013) recommended applying a suitable bootstrap when current values are correlated with a set of past values. However, they do not explicitly provide a bootstrap algorithm for calculating critical values under dependent data.

We estimate σ_*^2 by applying a consistent subsampling estimator proposed by Politis and Romano (1993) and Politis, Romano, and Wolf (1999). If the vector of signs of the return differentials $\boldsymbol{\omega}_t$ is strictly stationary satisfying $\sum_{k=1}^{\infty} |\mathrm{Cov}(\omega_{i,1}, \omega_{i,1+k})| < \infty$, for all $i = 1, \ldots, N$, then

$$\hat{\sigma}_*^2 = \mathrm{Var}(\omega_{i,1}) + 2\sum_{k=1}^{T} \left(1 - \frac{k}{T}\right) \mathrm{Cov}(\omega_{i,1}, \omega_{i,1+k}). \tag{7}$$

We could estimate σ_*^2 by plugging the estimates of $\mathrm{Cov}(\omega_{i,1}, \omega_{i,1+k})$ in (7), for all $i = 1, \ldots, N$, but the covariance estimates are unreliable for lags close to T since they rely on a gradually lower sample size. We can just aspire to obtain good estimates of $\mathrm{Cov}(\omega_{i,1}, \omega_{i,1+k})$ for $k = 1, \ldots, b$, where $b \ll T$ (see Politis and Romano 1993).

Nevertheless, it is possible to estimate σ_*^2 by calculating the sample variability of $(1/\sqrt{b})\sum_{t=j}^{j+b-1} \omega_{i,t}$, for $j = 1, \ldots, T - b + 1$, which is a subsampling estimator of σ_*^2. Hence, we apply a consistent subsampling scheme developed by Politis and Romano (1993) and Politis, Romano, and Wolf (1999) to estimate σ_*^2. We describe a subsampling algorithm for estimating σ_*^2 in (6) as follows. For each return category, $i = 1, \ldots, N$, given a series of signs of return differentials $\{\omega_{i,1}, \ldots, \omega_{i,T}\}$, we define $B = T - b + 1$ subsamples with size b as $\{\omega_{i,j}, \ldots, \omega_{i,j+b-1}\}$. For each subsample $j \le B$, we calculate the subsampling variability $(1/\sqrt{b})\sum_{t=j}^{j+b-1} \omega_{i,t}$. Then, we estimate σ_*^2 by applying the following subsampling estimator:

$$\hat{\sigma}_{*,b}^2 = \frac{1}{T - b + 1} \sum_{j=1}^{T-b+1} \left[\frac{1}{\sqrt{b}} \left(\sum_{t=j}^{j+b-1} \omega_{*,t} - \sqrt{b}\bar{\omega}_{*,t} \right)^2 \right], \tag{8}$$

where b is a subsample size that satisfies the restriction $1/b + b/T \to 0$ as $T, b \to \infty$. Thus, we calculate our test statistic as follows:

$$\hat{M}_b = \frac{\hat{\Omega}^*}{\hat{\sigma}_{*,b}}. \tag{9}$$

The choice of the subsample size has a significant effect on the result for finite samples (Sakov and Bickel 2000). Politis and Romano (1993) suggest applying a subsample size that satisfies $\kappa T^{0.3}$ without providing a guideline for choosing the value of κ. Although the value of κ has no asymptotic estimation effect, it plays an important role in finite samples. Series with strong serial correlation require high values of κ to capture its serial correlation, while independent and identically distributed (IID) series need small values of κ. We address this problem by proposing a functional form for κ. Let $\kappa = \sum_{j=0}^{\lambda} |\mathrm{Corr}(\omega_{i,t}, \omega_{i,t-j})|$, with $1/\lambda + \lambda/T \to 0$, so that κ is proportional to the serial correlation of the series, where $\kappa = 1$ for an IID series and $\kappa > 1$ for a serially correlated series.

We need to ensure that certain regularity conditions hold to determine the asymptotic distribution of the $\hat{M}_b$ test in (9). In what follows, let "$\xrightarrow{p}$" and "$\xrightarrow{d}$" denote convergence in probability and in distribution, respectively. Theorem 1 below delivers the limit distribution of our $\hat{M}_b$ test statistic in (9).

Theorem 1. *Suppose that (i)* $\boldsymbol{d}_t = (d_{1,t}, d_{2,t}, \dots, d_{N,t})'$ *is a strictly stationary and* α*-mixing process with coefficients satisfying* $\sum_{k=1}^{\infty} k^{p-1}(\alpha_{\mathbf{d}_t}(k))^{\delta/(2p+\delta)} < \infty$*, and* $\boldsymbol{\omega}_t = (\omega_{1,t}, \omega_{2,t}, \dots, \omega_{N,t})'$ *is strictly stationary, (ii)* $\mathrm{E}|\boldsymbol{\omega}_1|^{2p+\delta} < C$*, where* $p \in \mathbb{Z}$ *and* $p > 2$*,* $0 < \delta \leq 2$*, and* $C > 0$ *are constants, (iii)* $\sqrt{T}(\hat{\Omega} - E(\hat{\Omega})) \xrightarrow{d} N(0, \mathbf{V})$*, where* $\boldsymbol{V}$ *is a positive definite* $N \times N$ *covariance matrix with typical element* $v_{i,j}$*, (iv)* $\Pr\left[\left(\hat{\Omega}^{*,-m} - \hat{\Omega}^{*}\right) > 0\right] = 1$ *as* $t \to \infty$*, (v) the subsample size b satisfies* $b \to \infty$ *and* $b = o(T)$*, and (vi) under* H_0^**, it holds* $H_0^*: (\Omega_1, \dots, 0, \dots, \Omega_N)$*, with* $\Omega_i > 0$ *for all* $i \neq m$*. Then,*

(i) $\hat{\sigma}_*^2 \xrightarrow{p} \sigma_*^2$,

(ii) $\sqrt{T}\hat{M}_b \xrightarrow{d} N(0, 1)$.

Assumptions (i)–(iii) of Theorem 1 provide standard conditions for return differentials under time series data. As the sign(·) function is a Borel-measurable function, Assumption (i) implies that $\boldsymbol{\omega}_t = (\omega_{1,t}, \omega_{2,t}, \dots, \omega_{N,t})'$ is also an α-mixing process. Assumptions (i)–(ii) are necessary to restrain the dependence of $\mathbf{d}_t = (d_{1,t}, d_{2,t}, \dots, d_{N,t})'$ and $\boldsymbol{\omega}_t = (\omega_{1,t}, \omega_{2,t}, \dots, \omega_{N,t})'$; they accommodate commonly used models for asset returns such as generalized autoregressive conditional heteroscedasticity (GARCH) and factor models under mild additional assumptions. Assumption (iv) ensures that the minimum function over $\boldsymbol{\Omega}$ returns an independent random variable of $\boldsymbol{\Omega}^{-m}$, where $\boldsymbol{\Omega}^{-m}$ represents the $\boldsymbol{\Omega}$ sample space excluding Ω^*. Hence, it guarantees that asymptotically, for some m, $\Omega^* \equiv \min_{i=1,\dots,N} \Omega_i = \Omega_m$ always, reducing our multivariate scenario to a univariate case. Assumption (v) imposes restrictions on the subsample sizes, as in Politis and Romano (1993) and Politis, Romano, and Wolf (1999). Finally, H_0^* of

Assumption (vi) refers to the worst-case scenario where all Ω_i are greater than zero except for Ω^*, implying a weakly monotonic relationship that falls into the R_2 region of (2).

The term "worst-case scenario" of Assumption (vi) was coined by Romano and Wolf (2013) since any other scenario that has more zeros is easier to identify as nonmonotonic than the worst-case scenario. Theorem 1 shows that our test statistic has correct size for the worst-case scenario. Therefore, the presence of more zeros only pushes our test statistic towards the nonrejection of the null hypothesis of no monotonicity. Unreported simulations corroborate this result, and they are available upon request to the corresponding author.

Let the power function of a test that rejects H_0 when a test statistic S_T belongs to the critical region K be $\pi_T(\theta) = P_\theta(S_T \in K)$. We define the asymptotic relative efficiency (ARE) of test 1 with respect to test 2, for a given level α and power γ, as

$$\text{ARE} \equiv \lim_{\nu\to\infty} \frac{T_{\nu,1}}{T_{\nu,2}},$$

where $T_{\nu,j}$ is the minimal number of observations such that $\pi_{T_{\nu,j}}(0) \leq \alpha$ and $\pi_{T_{\nu,j}}(\theta_\nu) \geq \gamma$, for $j = 1, 2$. When testing a location parameter, the sign test has higher ARE over the t-test if the underlying distribution is heavy-tailed, with an ARE of $4f^2(0) \int r^2 f(r)\mathrm{d}r$ (Van der Vaart 2000, Chapter 14). In this instance, $f(r) \sim g(r)$ denotes $f(r) = g(r)(1 + o(1))$ as $r \to \infty$. This result underscores the advantage of the sign-based test under thick-tailed distributions. For instance, the ARE of the sign test over the t-test under Laplace and Cauchy distributions is 2 and ∞, respectively; conversely, the ARE is $2/\pi$ and $1/3$ under Gaussian and uniform distributions, respectively.

3 Monte Carlo Study

In this section, we perform Monte Carlo simulations to evaluate the finite-sample performance of our proposed test statistic. We investigate two sets of scenarios. The first set, Exp. H_0, considers the case where H_0 in (5) is valid and the relationship between expected returns and portfolio sorts is nonmonotonic. It evaluates the finite-sample size of various tests in the worst-case scenario, where $\boldsymbol{\Delta}$ lies on the boundary part of R_2, but never in R_3 of (2). Under this scenario, the relationship is not strictly monotonic, but close to. The test of Patton and Timmermann (2010) is nonconservative under this scenario since it overlooks the R_2 region of (2). The second set of scenarios, Exp.H_A, evaluates the power performance of many tests under H_A in (5). We design these scenarios as follows:

$$\text{Exp.}\, H_0 : \Delta = (\Delta, \Delta, \ldots, \Delta, 0)',$$

$$\text{Exp.}\, H_A : \Delta = (\Delta, \Delta, \ldots, \Delta, \Delta)',$$

where $\mathbf{\Delta}$ is a column-vector with N rows (return categories) and $\Delta \geq 0$. We propose the following data-generating processes (DGPs) for the return differentials:

$$\text{DGP.1:}\quad \mathbf{d}_t \sim N(\Delta, \Sigma), \Sigma_{i,j} = 0.9^{|i-j|},$$

$$\text{DGP.2:}\quad d_{i,t} = \Delta_i + \varepsilon_{i,t}, \varepsilon_{i,t} = \sigma_{i,t} u_{i,t}, \sigma_{i,t}^2 = 0.1 + 0.9\varepsilon_{i,t-1}^2, u_{i,t} \overset{\text{i.i.d.}}{\sim} N(0,1),$$

$$\text{DGP.3:}\quad d_{i,t} = \Delta_i + 0.7 d_{i,t-1} + 0.3 u_{i,t}, u_{i,t} \overset{\text{i.i.d.}}{\sim} N(0,1),$$

$$\text{DGP.4:}\quad d_{i,t} = \Delta_i + \varepsilon_{i,t}, \varepsilon_{i,t} = \sigma_{i,t} u_{i,t}, u_{i,t} \overset{\text{i.i.d.}}{\sim} N(0,1), \eta_{i,t} = \varepsilon_{i,t}\sigma_{i,t}^{-1},$$
$$\log(\sigma_{i,t}^2) = 0.001 + \left[0.25(\eta_{i,t-1}) + 0.10(|\eta_{i,t-1}| - E|\eta_{i,t-1}|)\right]$$
$$+ 0.99 \log(\sigma_{i,t-1}^2),$$

$$\text{DGP.5:}\quad d_{i,t} = \Delta_i + 0.5\varepsilon_{i,t}, \varepsilon_{i,t} \overset{\text{i.i.d.}}{\sim} \text{Cauchy}(\delta, c), \{\delta; c\} = \{0; 1\},$$

$$\text{DGP.6:}\quad d_{i,t} = \Delta_i + \varepsilon_{i,t}, \varepsilon_{i,t} \overset{\text{i.i.d.}}{\sim} S(\alpha, \beta, c, \delta), \alpha \in [0.2; 1.8],$$
$$\{\beta; c; \delta\} = \{0; 0.01, 0\}.$$

where $\mathbf{\Delta} = (\Delta_1, \ldots, \Delta_N)'$, $\{\delta, c\}$ are the location and the scale parameter of a Cauchy distribution, and $\mathbf{\Sigma}$ is a Toeplitz covariance matrix with typical element $\Sigma_{i,j} = 0.9^{|i-j|}$. Let $S(\alpha, \beta, c, \delta)$ be the stable distribution with the location parameter δ, the scale parameter c, the skewness parameter β, and the stability index (characteristic exponent) α. We denote the standard stable distribution (with $c = 1$ and $\delta = 0$) by $S(\alpha, \beta)$. The characteristic exponent α determines the decay rate of the tails of stable distributions. When $\alpha = 2$, a normal distribution arises with variance 2. The variance is infinite for $\alpha < 2$. The mean is defined if $\alpha > 1$, but it is undefined when $\alpha \leq 1$.

Romano and Wolf (2013) examined DGP.1, where the return differentials close to each other are strongly dependent, whereas the return differentials far from each other are weakly dependent. DGP.2 considers return differentials that follow an autoregressive conditional heteroskedasticity (ARCH) process of order one. DGP.3 evaluates the performance of the tests under short-term dependence for return differentials.

Under DGP.4, the return differentials follow an exponential GARCH, E-GARCH(1,1), process of Nelson (1991). In the E-GARCH model, a negative return change increases volatility more than an equivalent positive return increase. This model is useful for estimating asymmetries on the volatility of financial assets

(Brandt and Jones 2006). DGP.5 considers errors following a heavy-tailed distribution with an undefined mean. The Cauchy distribution is a special case of the stable distribution. The Cauchy distribution is a symmetric standard stable distribution ($\beta = 0$) with $\alpha = 1$. Hence, DGP.5 assesses the finite-sample performance of various tests under a return differentials distribution with undefined mean and heavy tails. Finally, DGP.6 determines the finite-sample behavior of various tests under heavy-tailed distributions for different degrees of the characteristic exponent $\alpha = \{0.2, 0.4, \dots, 1.8\}$ that determines the thickness of the distribution tails.

We use an empirical rejection frequency of 5% in all experiments and 10,000 Monte Carlo simulations to calculate the empirical rejection probabilities. We calculate our test statistic with a subsample size of $b = \kappa T^{0.3}$, where $\kappa = \sum_{j=0}^{\lambda} |\text{Corr}(\omega_{i,t}, \omega_{i,t-j})|$ and $\lambda = T^{0.15}$. We compare our findings with the MR and MR_{all} tests of Patton and Timmermann (2010) calculated on the minimal set of possible inequalities (MR) and on all possible inequalities implied by monotonicity (MR_{all}), and the conservative test of Romano and Wolf (2013) (RW).

We employ a block-bootstrap approach to calculate the MR and MR_{all} test statistics and the critical values of the RW test, with $B = 999$ bootstrap replications and a block length of $b = \lfloor T^{1/3} \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function, following the suggestion of Künsch (1989) for the choice of the block length.

We present two sets of simulation experiments. The first set verifies the empirical rejection rates for different values of $\Delta = \{0.00, 0.05, \dots, 0.50\}$, with a sample of size $T = 120$. We consider $N = 10$ return categories. Figure 1 displays the rejection frequencies of the various tests under H_0 in (5). The empirical size of the $\hat{M}_b$ test is almost always correct under H_0 in (5). On the other hand, both MR and MR_{all} tests present size distortions for all DGPs. Thus, for clarity, we only report these test results for DGP.1. Although the RW test is conservative for DGP.1 and DGP.2, it tends to under-reject under DGPs 4–5; besides, it presents size distortions under short-term dependence (DGP.3).

Figure 2 displays the empirical rejection probabilities of the various tests under the alternative hypothesis. Due to size distortions under H_0, we only report the empirical power of the MR and MR_{all} tests for DGP.1. In general, these tests present the highest rejection rate among all tests. The RW test has the highest rejection rate for DGP.1 and DGP3. Nevertheless, this test is non-conservative under DGP.3. The $\hat{M}_b$ test outperforms the RW test under DGP.2, DGP.4, and DGP.5. The presence of conditional heteroskedasticity, under DGP.2 and DGP.4, affects the finite-sample behavior of the RW test. Conversely, our test is more powerful under conditional heteroskedasticity and thick-tailed distributions with an undefined mean (DGP.5) due to the robustness of the sign test.

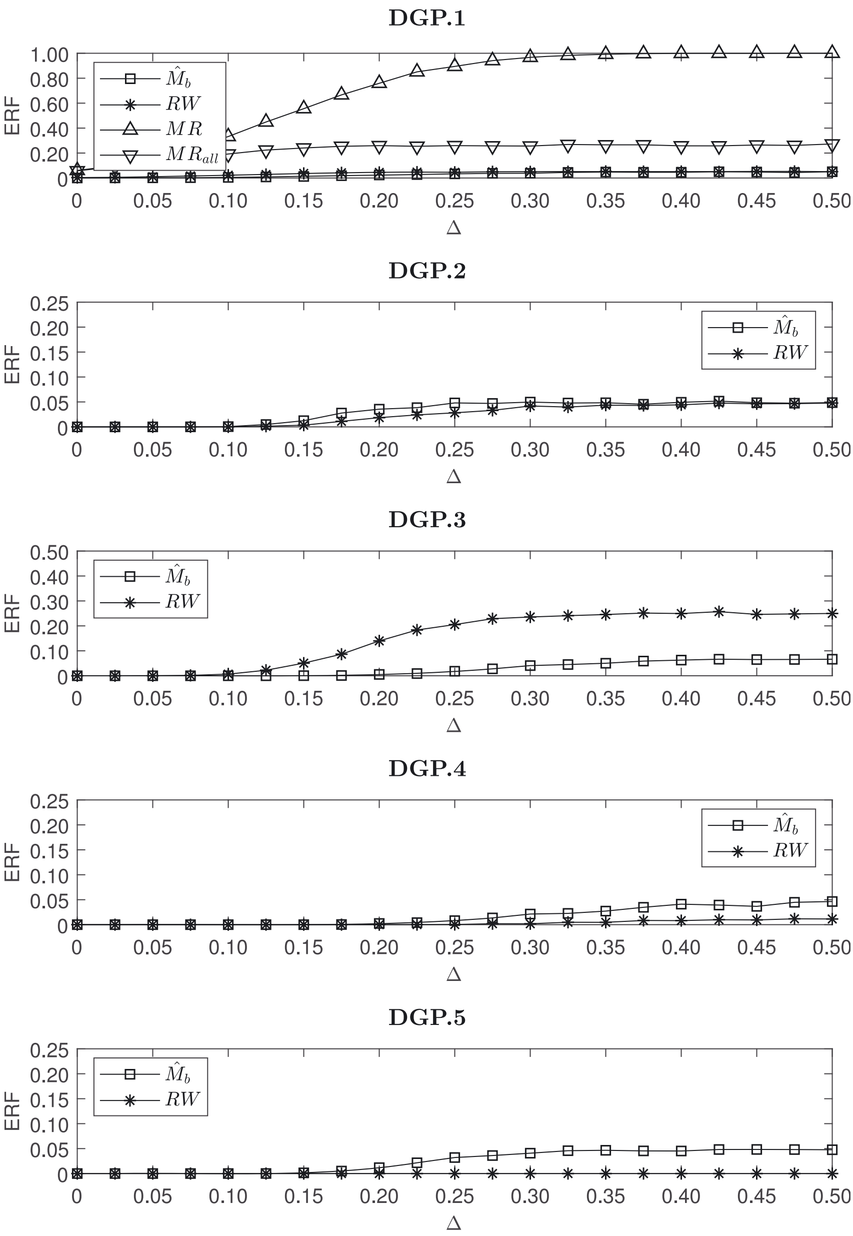

Figure 1: Empirical rejection frequencies under H_0 (Exp. H_0) for many tests with $N = 10$ return categories and different values of Δ. We perform 10,000 Monte Carlo repetitions.

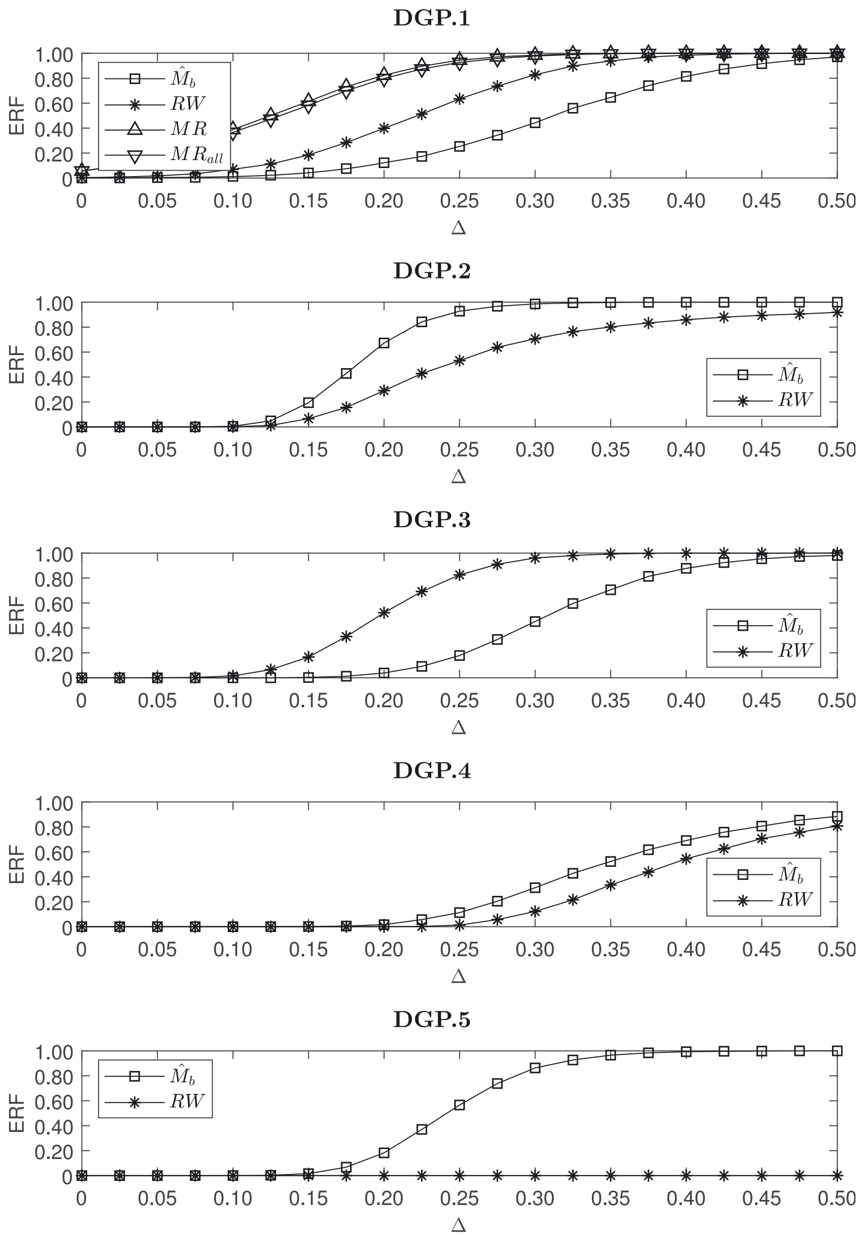

Figure 2: Empirical rejection frequencies under H_A (Exp. H_A) for many tests with $N = 10$ return categories and different values of Δ. We perform 10,000 Monte Carlo repetitions.

For comparison, we also consider the case of $N = 5$ return characteristics as the test power is inversely related to the amount of return differentials N. The larger the amount of categories, the more difficult it is to verify positive expected return differentials over the entire range of categories. Figure 3 shows the frequency rejection rates for $N = 5$ return categories. Consistent with the results presented in Romano and Wolf (2013), the power of all tests increase when N is decreased. In addition, the discrepancies in power are smaller when $N = 5$. The RW test still outperforms the $\hat{M}_b$ test for DGP.1 and DGP.3; on the other hand, our $\hat{M}_b$ test presents the highest power under DGP.2, DGP.4, and DGP.5.

Figure 4 reports the simulated rejection probabilities of the $\hat{M}_b$ and RW tests under DGP.6, with $T = 120$ and $\Delta = 0.50$. We assess the finite-sample behavior of these tests under heavy-tailed distributions for several rates of decay of the distribution tails. The $\hat{M}_b$ test delivers a correct finite-sample size and maximum power for all heavy-tailed characteristic exponents $\alpha \in [0.2; 1.8]$. Conversely, the RW test is non-conservative for $\alpha < 1.2$. The RW test statistic has a power close to zero for $\alpha \leq 0.6$. Thus, the $\hat{M}_b$ test presents reliable finite-sample performance under heavy-tailed distributions.

Next, we report another set of simulations, where we present the empirical rejection frequencies for five sample sizes $T = \{100, 200, 300, 400, 500\}$ with $N = 10$ return categories and $\Delta = 0.50$. We also calculate our test statistic of (9) for three subsample sizes of $b_k = \kappa_k T^{0.3}$, where $\kappa_k = \sum_{j=0}^{\lambda_k} |\text{Corr}(\omega_{i,t}, \omega_{i,t-j})|$ and $\lambda_k \in \{T^{0.05}, T^{0.10}, T^{0.15}\}$ for $k = 1, 2, 3$. We denote our test statistic of (9) by $\hat{M}_{\lambda_k}$ depending on the λ_k applied. To save space, we omit the results obtained for DGP.6.

Tables 1 and 2 present the rejection frequencies of the various tests under H_0 and H_A in (5), respectively. Consistent with the results presented in Figure 1, the $\hat{M}_b$ test displays a finite-sample size close to the nominal level for DGPs 1–5; on the other hand, the MR and MR_{all} tests present size distortions for all DGPs, while the RW test is conservative only for DGPs 1–2. These findings are robust to different numbers of observations and subsample sizes of the $\hat{M}_b$ test. Further, Table 2 suggests that there are no significant differences in power for sample sizes larger than $T = 200$.

Our findings from the Monte Carlo experiments indicate that the $\hat{M}_b$ test should be selected for practical use since it is robust to heavy-tailed distributions and to conditional heteroskedasticity. The RW test provides a zero rejection rate under some heavy-tailed distributions with undefined means, and it is anticonservative in the presence of conditional heteroskedasticity and autocorrelated data. The MR and MR_{all} tests fail to control the probability of a type 1 error for all DGPs. Conversely, our proposed test statistic presents a correct finite-sample size under all DGPs together with similar power to alternative approaches

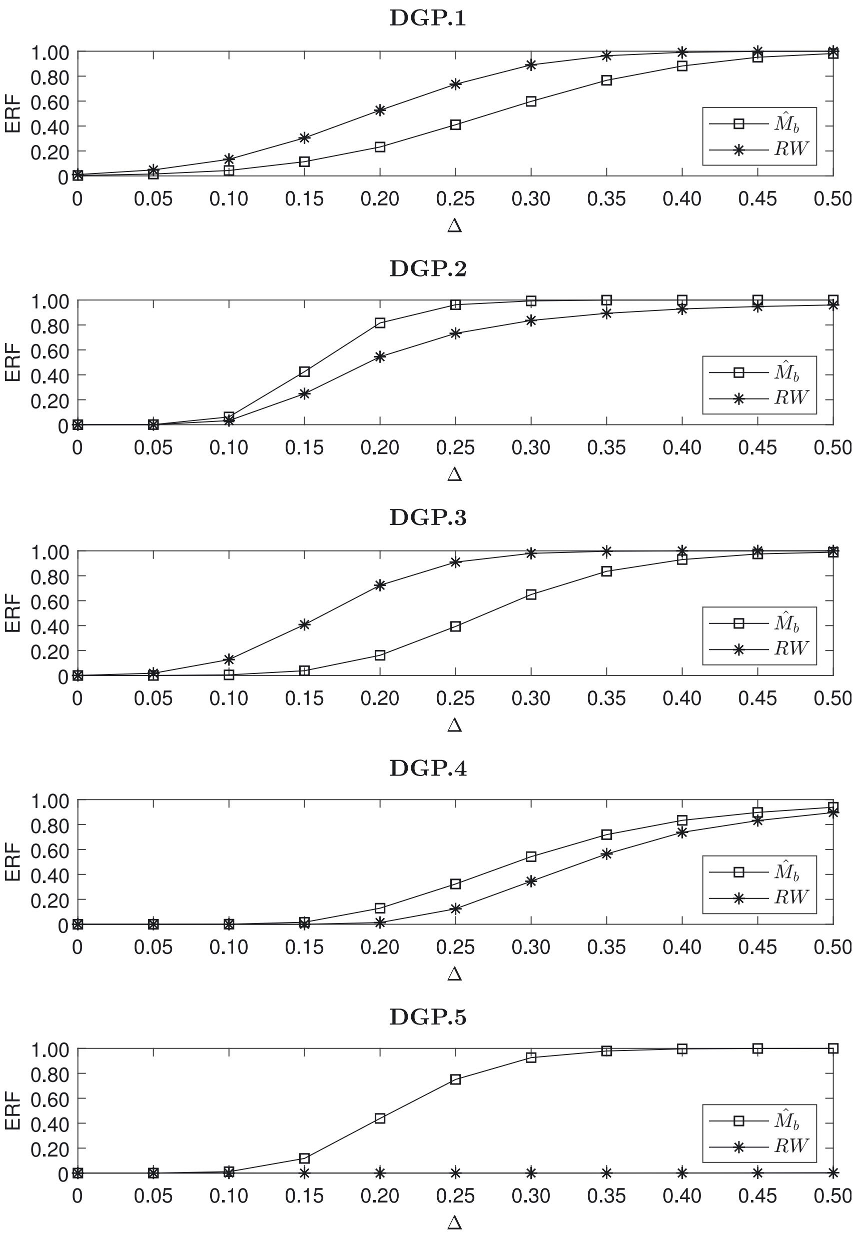

Figure 3: Empirical rejection frequencies under H_A (Exp. H_A) for many tests with $N = 5$ return categories and different values of Δ. We perform 10,000 Monte Carlo repetitions.

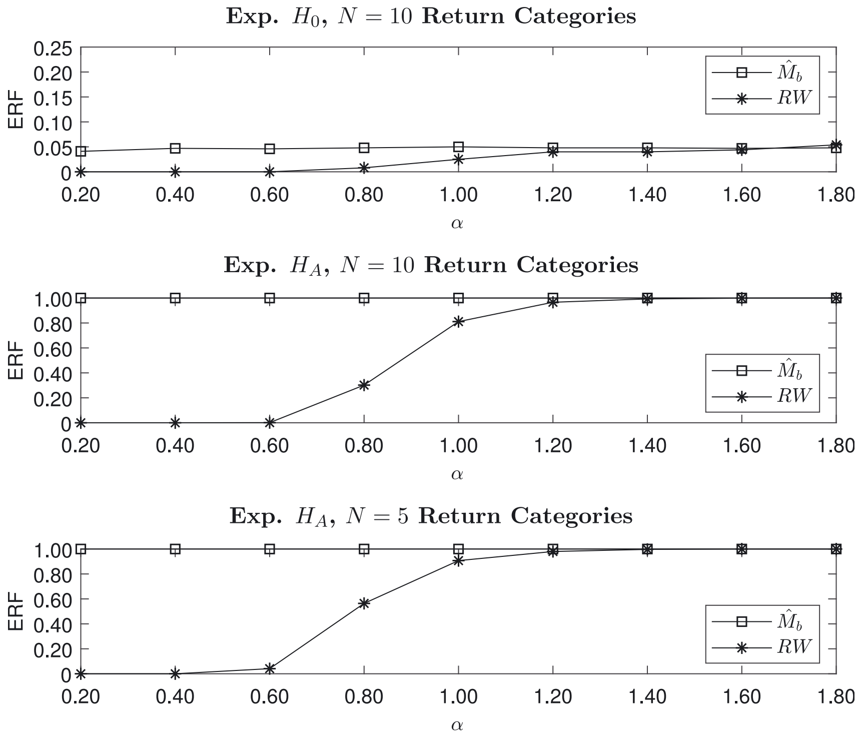

Figure 4: Empirical rejection frequencies for many tests under DGP.6 with $\Delta = 0.50$ and different values of the stability index α. We perform 10,000 Monte Carlo repetitions.

when the sample size is larger than $T = 200$. Our test also outperforms other tests when $T < 200$ under conditional heteroskedasticity (DGP 2) or heavy-tailed distributions with undefined mean (DGP 5). Overall, our test presents a proper finite-sample performance.

4 Empirical Application: Portfolio Sorts

To highlight the usefulness of our procedure, we revisit one of the empirical applications of Patton and Timmermann (2010). We analyze returns on portfolios ranked by firm characteristics such as market equity (ME), book-to-market value (B-M), cashflow-price ratio (CF-P), earnings-price ratio (E-P), dividend-price ratio (D-P), investment (INV), short-term reversal (STR), and long-term reversal (LTR). Monotonicity tests for returns on these categories are important to verify

Table 1: Empirical rejection frequencies under H_0 (Exp. H_0).

DGP	T	$\hat{M}_{\lambda_1}$	$\hat{M}_{\lambda_2}$	$\hat{M}_{\lambda_3}$	RW	MR	MR$_{\text{all}}$
				DGP.1			
	100	0.049	0.050	0.050	0.049	1.000	0.257
	200	0.048	0.048	0.048	0.048	1.000	0.253
	300	0.045	0.048	0.048	0.054	1.000	0.256
	400	0.046	0.048	0.048	0.053	1.000	0.268
	500	0.047	0.044	0.044	0.048	1.000	0.247
				DGP.2			
	100	0.051	0.051	0.051	0.046	1.000	0.710
	200	0.052	0.052	0.052	0.051	1.000	0.727
	300	0.047	0.047	0.047	0.048	1.000	0.720
	400	0.043	0.043	0.043	0.047	1.000	0.717
	500	0.048	0.048	0.048	0.049	1.000	0.725
				DGP.3			
	100	0.132	0.093	0.093	0.289	1.000	0.637
	200	0.109	0.083	0.083	0.287	1.000	0.660
	300	0.096	0.061	0.061	0.292	1.000	0.665
	400	0.086	0.043	0.043	0.295	1.000	0.685
	500	0.090	0.054	0.031	0.295	1.000	0.684
				DGP.4			
	100	0.045	0.044	0.044	0.011	1.000	0.603
	200	0.047	0.047	0.047	0.012	1.000	0.587
	300	0.043	0.043	0.043	0.011	1.000	0.598
	400	0.046	0.046	0.046	0.010	1.000	0.573
	500	0.046	0.046	0.045	0.010	1.000	0.617
				DGP.5			
	100	0.052	0.052	0.052	0.000	0.539	0.148
	200	0.051	0.049	0.049	0.000	0.527	0.152
	300	0.046	0.045	0.045	0.000	0.523	0.152
	400	0.047	0.046	0.046	0.000	0.519	0.151
	500	0.048	0.050	0.050	0.000	0.526	0.154

This table displays the empirical rejection frequencies under H_0 (Exp. H_0) for many tests with $N = 10$ return categories and $\Delta = 0.50$. $\hat{M}_{\lambda_k}$ denotes our test statistic in (9) with three different subsample sizes $b_k = \kappa_i T^{0.3}$, where $\kappa_k = \sum_{j=0}^{\lambda_k} |\text{Corr}(\omega_{i,t}, \omega_{i,t-j})|$ and $\lambda_k \in \{T^{0.05}, T^{0.10}, T^{0.15}\}$ for $k = 1, 2, 3$. RW is the conservative test of Romano and Wolf (2013). MR and MR$_{\text{all}}$ are the proposed tests of Patton and Timmermann (2010) based on the minimal set of possible inequalities and on all possible inequalities implied by monotonicity, respectively. We applied $B = 999$ bootstrap replications and a block length of $b = \lfloor T^{1/3} \rfloor$ for the MR, MR$_{\text{all}}$, and RW tests. We performed 10,000 Monte Carlo simulations.

Table 2: Empirical rejection frequencies under H_A (Exp. H_A).

DGP	T	$\hat{M}_{\lambda_1}$	$\hat{M}_{\lambda_2}$	$\hat{M}_{\lambda_3}$	RW	MR	MR_{all}
				DGP.1			
	100	0.933	0.933	0.933	0.998	1.000	1.000
	200	0.999	0.999	0.999	1.000	1.000	1.000
	300	1.000	1.000	1.000	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000
				DGP.2			
	100	1.000	1.000	1.000	0.895	1.000	0.987
	200	1.000	1.000	1.000	0.959	1.000	0.997
	300	1.000	1.000	1.000	0.978	1.000	0.998
	400	1.000	1.000	1.000	0.984	1.000	0.997
	500	1.000	1.000	1.000	0.988	1.000	1.000
				DGP.3			
	100	0.834	0.720	0.725	0.980	1.000	0.999
	200	0.994	0.980	0.981	1.000	1.000	1.000
	300	1.000	0.990	0.999	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	0.999	1.000	1.000	1.000
				DGP.4			
	100	0.829	0.827	0.827	0.753	1.000	1.000
	200	0.989	0.989	0.989	0.921	1.000	1.000
	300	1.000	1.000	1.000	0.978	1.000	1.000
	400	1.000	1.000	1.000	0.992	1.000	1.000
	500	1.000	1.000	1.000	0.998	1.000	1.000
				DGP.5			
	100	0.999	0.999	0.999	0.000	0.312	0.155
	200	1.000	1.000	1.000	0.000	0.311	0.164
	300	1.000	1.000	1.000	0.000	0.303	0.154
	400	1.000	1.000	1.000	0.000	0.297	0.154
	500	1.000	1.000	1.000	0.000	0.306	0.160

This table reports the empirical rejection frequencies under H_A (Exp. H_A) for many tests with $N = 10$ return categories and $\Delta = 0.50$. We used the same specifications of Table 1.

their validity as unobserved risk factors. Thus, we test whether there is a strictly increasing relationship between portfolio rank and returns for portfolios ranked on B-M, CF-P, E-P, and D-P. In addition, we test for a strictly decreasing relationship for returns on portfolios sorted on ME, INV, STR, and LTR. Since we are testing whether these relationships are strictly decreasing, we reorder the asset returns as in Patton and Timmermann (2010).

Our data consist of monthly stock returns on value-weighted decile portfolios obtained from Kenneth French's website, made up of common stocks listed on NYSE, AMEX, and NASDAQ. The data cover the earliest starting date for each sorting variable: ME from 1926.07 to 2018.09, B-M from 1926.07 to 2018.09, CF-P from 1951.07 to 2018.09, E-P from 1951.07 to 2018.09, D-P from 1927.07 to 2018.09, INV from 1963.07 to 2018.09, STR from 1926.02 to 2018.09, and LTR from 1931.07 to 2018.09. Figure 5 displays the returns on decile portfolios ranked by each one of the eight characteristics.

We first verify whether the returns on portfolio sorts have heavy-tailed distributions. We estimate the kurtosis of the returns for each decile of the portfolio sorts. If we reject normality, we may apply the maximum likelihood method proposed by DuMouchel (1973) to estimate the index of stability ($\hat{\alpha}$) of a stable distribution. We apply this procedure since Nolan (2001) demonstrated that the

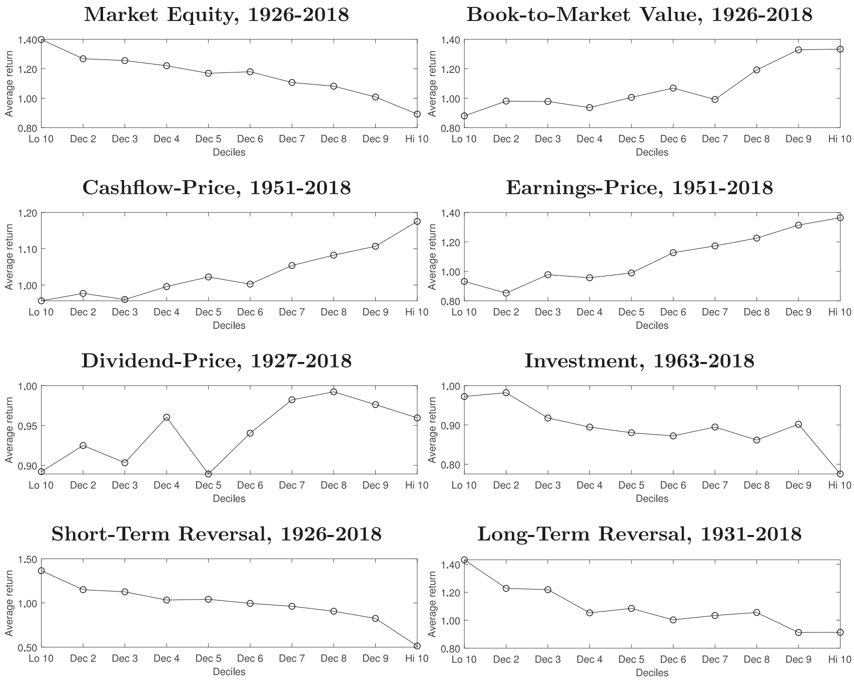

Figure 5: Monthly average returns on decile portfolios ranked by firm characteristics.

estimator of DuMouchel (1973) is the most efficient estimator of α when the true distribution is not normal.

Table 3 presents the estimated kurtosis and index of stability ($\hat{\alpha}$) of the returns on decile portfolios for all sorts and for each decile. The kurtosis values of the returns on each decile suggest that their distributions are leptokurtic. In addition, unreported Jarque–Bera normality tests indicate that all decile portfolios' returns

Table 3: Estimated kurtosis and index of stability ($\hat{\alpha}$): Decile portfolios.

	ME	B-M	CF-P	E-P	D-P	INV	STR	LTR
				Kurtosis				
Low	44.782	9.144	4.526	4.518	6.523	5.348	9.371	23.931
2	25.653	7.194	5.052	5.104	7.676	4.806	9.440	33.324
3	21.671	9.946	4.937	4.890	7.206	4.792	9.834	25.351
4	19.415	18.426	5.050	5.190	11.202	4.970	8.288	22.334
5	14.800	16.366	6.141	5.109	14.162	4.892	22.598	27.623
6	16.789	23.521	4.786	5.079	14.622	5.124	9.883	17.732
7	13.706	19.966	5.350	5.070	11.021	5.410	19.605	21.808
8	14.731	21.873	4.948	5.084	20.839	5.128	24.612	12.742
9	13.501	24.969	5.508	5.475	25.983	4.720	26.855	10.153
High	9.987	25.454	5.187	5.273	32.679	4.607	15.942	6.731
				Index of stability ($\hat{\alpha}$)				
Low	1.535	1.740	1.868	1.837	1.722	1.857	1.538	1.578
2	1.619	1.749	1.876	1.837	1.768	1.830	1.596	1.563
3	1.644	1.738	1.805	1.835	1.657	1.847	1.623	1.578
4	1.692	1.653	1.832	1.862	1.744	1.829	1.650	1.644
5	1.677	1.641	1.828	1.845	1.689	1.802	1.663	1.598
6	1.692	1.637	1.880	1.922	1.683	1.779	1.673	1.650
7	1.683	1.613	1.859	1.856	1.703	1.837	1.674	1.668
8	1.693	1.617	1.868	1.842	1.645	1.851	1.724	1.696
9	1.680	1.567	1.845	1.789	1.639	1.796	1.662	1.747
High	1.695	1.525	1.835	1.867	1.556	1.855	1.701	1.758

This table displays the estimated kurtosis and index of stability ($\hat{\alpha}$) of monthly returns on decile portfolios sorted on firm characteristics. We estimated the index of stability ($\hat{\alpha}$) of a stable distribution by applying the maximum likelihood method proposed by DuMouchel (1973) and Nolan (2001). We obtained the data from Kenneth French's website. The data cover the earliest starting date for each sorting variable: market equity (ME) from 1926.07 to 2018.09, book-to-market value (B-M) from 1926.07 to 2018.09, cashflow-price ratio (CF-P) from 1951.07 to 2018.09, earnings-price ratio (E-P) from 1951.07 to 2018.09, dividend-price ratio (D-P) from 1927.07 to 2018.09, investment (INV) from 1963.07 to 2018.09, short-term reversal (STR) from 1926.02 to 2018.09, and long-term reversal (LTR) from 1931.07 to 2018.09.

do not follow a normal distribution. Finally, all estimated indices of stability ($\hat{\alpha}$) lie in the interval $\hat{\alpha} \in (1.53; 1.93)$, indicating that the distributions of returns on decile portfolios are heavy-tailed.

Table 4 presents the results of monotonicity tests. The $\hat{M}_b$ test does not reject the null hypothesis of no systematic relationship, at the 5% significance level, for all portfolio sorts. The RW test also fails to reject the null hypothesis for each one of the portfolio sorts at the 5% significance level. Both MR and MR_{all} tests do not reject the null hypothesis for returns on portfolios ranked by B-M, E-P, D-P, and INV at the 5% significance level. Conversely, the MR and MR_{all} tests are supportive of monotonicity in expected returns on the portfolios ranked by ME, CF-P, and STR. The MR p-value also reports evidence of monotonicity in returns on portfolios ranked by LTRs.

By controlling the type 1 error under heavy-tailed distributions, our $\hat{M}_b$ test rejects a monotonic relationship between asset returns and eight portfolio sorts. Thus, we find no significant evidence of monotonicity in asset returns for these portfolios.

Table 4: Monotonicity tests in asset returns: Decile portfolios.

	ME	B-M	CF-P	E-P	D-P	INV	STR	LTR
$\hat{M}_b$ p-value	1.000	0.992	0.919	0.848	0.994	0.906	0.795	0.904
MR p-value	0.037	0.298	0.026	0.138	0.893	0.140	0.004	0.032
MR_{all} p-value	0.035	0.234	0.018	0.068	0.761	0.068	0.004	0.064
RW test stat.	−0.007	−0.035	−0.019	−0.031	−0.067	−0.035	−0.003	−0.015
RW 5%-c.v.	0.049	0.049	0.058	0.058	0.050	0.064	0.049	0.051

This table reports the results of tests for monotonicity in monthly average returns for stocks sorted into eight value-weighted decile portfolios. We obtained the data from Kenneth French's website. The data cover the earliest starting date for each sorting variable: market equity (ME) from 1926.07 to 2018.09, book-to-market value (B-M) from 1926.07 to 2018.09, cashflow-price ratio (CF-P) from 1951.07 to 2018.09, earnings-price ratio (E-P) from 1951.07 to 2018.09, dividend-price ratio (D-P) from 1927.07 to 2018.09, investment (INV) from 1963.07 to 2018.09, short-term reversal (STR) from 1926.02 to 2018.09, and long-term reversal (LTR) from 1931.07 to 2018.09. $\hat{M}_b$ p-value is the p-value of our test statistic in (9) with subsample of size $b = \kappa T^{0.3}$, where $\kappa = \sum_{j=0}^{\lambda} |\text{Corr}(\omega_{i,t}, \omega_{i,t-j})|$ and $\lambda = T^{0.15}$. MR p-value and MR_{all} p-value are the p-values of the proposed tests of Patton and Timmermann (2010) based on the minimal set of possible inequalities and on all possible inequalities implied by monotonicity, respectively. RW test stat. is the conservative test statistic proposed by Romano and Wolf (2013), whose 5%-critical values are displayed in RW 5%-c.v. We applied $B = 999$ bootstrap replications and a block length of $b = \lfloor T^{1/3} \rfloor$ for the MR, MR_{all}, and RW tests.

5 Conclusions

Many postulated theories in economics and finance assume monotonic relationships between expected asset returns and certain underlying characteristics of an asset. For instance, certain asset pricing models suggest that expected returns should increase monotonically with the asset's book-to-market ratio, cash-flow price, or earnings-price ratio. Patton and Timmermann (2010) proposed a monotonicity test that postulates a weakly decreasing relationship under the null hypothesis versus the alternative hypothesis of a strictly increasing relationship. Since their test rules out a weakly increasing relationship under the null hypothesis, it fails to control the type 1 error. Romano and Wolf (2013) developed monotonicity tests that also consider a weakly increasing relationship under the null hypothesis that provide a correct size. However, their tests can have low power and incorrect finite-sample size under conditional heteroskedasticity or heavy-tailed distributions of the return differentials.

In this paper, we propose a sign-based test for monotonicity in asset returns that is valid under a general setting. We develop a test that has correct size under conditional heteroskedasticity and heavy-tailed distributions of return differentials. Monte Carlo simulations illustrate that our test statistic has a correct finite-sample size under all DGPs, together with similar power to other tests when the sample size is larger than $T = 200$. Conversely, the MR and $\mathrm{MR}_{\mathrm{all}}$ tests fail to control the probability of a type 1 error for all DGPs, while the RW test is conservative only for DGPs 1–2. Therefore, our method allows for testing for monotonicity in a variety of situations, extending the procedures of Patton and Timmermann (2010) and Romano and Wolf (2013). One can apply our test as a summary statistic for monotonic relationships in asset returns, controlling the possibility of falsely addressing strict monotonicity under dependent data, conditional heteroskedasticity, or heavy-tailed distributions.

Acknowledgments: The authors acknowledge helpful comments received from an anonymous reviewer. The authors also thank Carlos Velasco and Abderrahim Taamouti for providing helpful suggestions to improve the paper.
Research Funding: This work was supported by Spain's Ministerio de Ciencia e Innovación [grant number PID2020-114646RB-C43].

Appendix

Proof of Theorem 1. To prove part (i), we need to show that the conditions of Theorem 1 of Politis and Romano (1993) (Assumptions A0–A2 and conditions (i)–(iii)) are satisfied. Under Assumptions (iv) and (vi), we only need to analyze $\hat{\Omega}^*$ so that these assumptions allow us to consider our test statistic as univariate. Assumption (vi) establishes H_0^* of the worst-case scenario where all Ω_i are greater than zero except for Ω^*, implying a weakly monotonic relationship that falls into the R_2 region of (2). Besides, Assumptions (ii) and (iii) imply that $\hat{\Omega}^*$ has bounded moments.

Under Assumption (A2) of Politis and Romano (1993), both M and L are equal to 1, and T is a function that in our case is described by $\hat{\Omega}^* \equiv \min_{i=1,\ldots,N} \hat{\Omega}_i$, where $\hat{\Omega}_i \equiv (1/T)\sum_{t=1}^{T} \omega_{i,t} = \bar{\omega}_{i,t}$. Moreover, under H_0^* of Assumption (vi), $\hat{\Omega}^* = 0$. Then, Assumption (A0) and condition (iii) of Theorem 1 of Politis and Romano (1993) are satisfied by our Assumption (i), whereas Assumptions (A1), (A2), and (A3) follow by our Assumptions (ii), (iii) and (vi), and (iii), respectively. Finally, condition (i) of Theorem 1 of Politis and Romano (1993) is satisfied since our framework is univariate, and condition (ii) of Theorem 1 of Politis and Romano (1993) follows from our Assumption (v). Therefore, part (i) of Theorem 1 follows from an application of Theorem 1 of Politis and Romano (1993).

Then, by part (i) of Theorem 1, under H_0^* of Assumption (vi), we have that $\sqrt{T}\hat{M}_b \xrightarrow{d} N(0,1)$. □

References

Boldin, M., G. Simonova, and Y. N. Tyurin. 1997. *Sign-Based Methods in Linear Statistical Models*. *Translations of Mathematical Monographs*. Providence: American Mathematical Society.

Brandt, M. W., and C. S. Jones. 2006. "Volatility Forecasting with Range-Based EGARCH Models." *Journal of Business & Economic Statistics* 24: 470–86.

Campbell, B., and J. M. Dufour. 1991. "Over-Rejections in Rational Expectations Models: A Nonparametric Approach to the Mankiw-Shapiro Problem." *Economics Letters* 35: 285–90.

Campbell, B., and J. M. Dufour. 1995. "Exact Nonparametric Orthogonality and Random Walk Tests." *The Review of Economics and Statistics* 77: 1–16.

Campbell, B., and J. M. Dufour. 1997. "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter." *International Economic Review* 38: 151–73.

Capanu, M., G. A. Jones, and R. H. Randles. 2006. "Testing for Preference Using a Sum of Wilcoxon Signed Rank Statistics." *Computational Statistics & Data Analysis* 51: 793–6.

Cont, R. 2001. "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues." *Quantitative Finance* 1: 223–36.

Coudin, E., and J.-M. Dufour. 2009. "Finite-Sample Distribution-Free Inference in Linear Median Regressions under Heteroscedasticity and Non-linear Dependence of Unknown Form." *The Econometrics Journal* 12: S19–49.
Delgado, M. A., and C. Velasco. 2005. "Sign Tests for Long-Memory Time Series." *Journal of Econometrics* 128: 215–51.
DuMouchel, W. H. 1973. "On the Asymptotic Normality of the Maximum-Likelihood Estimate when Sampling from a Stable Distribution." *Annals of Statistics* 1: 948–57.
Embrechts, P., C. Kluppelberg, and T. Mikosch. 1997. *Modelling Extremal Events: With a View towards Insurance and Finance*. New York: Springer.
Fama, E. F. 1965. "The Behavior of Stock-Market Prices." *Journal of Business* 38: 34–105.
Fama, E. F. 1984. "Term Premiums in Bond Returns." *Journal of Financial Economics* 13: 529–46.
Gabaix, X. 2009. "Power Laws in Economics and Finance." *Annual Review of Economics* 1: 255–94.
Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley. 2003. "A Theory of Power-Law Distributions in Financial Market Fluctuations." *Nature* 423: 267–70.
Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley. 2006. "Institutional Investors and Stock Market Volatility." *Quarterly Journal of Economics* 121: 461–504.
Gerard, P. D., and W. R. Schucany. 2007. "An Enhanced Sign Test for Dependent Binary Data with Small Numbers of Clusters." *Computational Statistics & Data Analysis* 51: 4622–32.
Guillaume, D. M., M. M. Dacorogna, R. R. Davé, U. A. Müller, R. B. Olsen, and O. V. Pictet. 1997. "From the Bird's Eye to the Microscope: A Survey of New Stylized Facts of the Intra-Daily Foreign Exchange Markets." *Finance and Stochastics* 1: 95–129.
Hansen, P. R. 2005. "A Test for Superior Predictive Ability." *Journal of Business & Economic Statistics* 23: 365–80.
Ibragimov, R., and J. Walden. 2007. "The Limits of Diversification when Losses May Be Large." *Journal of Banking & Finance* 31: 2551–69.
Ibragimov, R., D. Jaffee, and J. Walden. 2009. "Nondiversification Traps in Catastrophe Insurance Markets." *Review of Financial Studies* 22: 959–93.
Ibragimov, M., R. Ibragimov, and P. Kattuman. 2013. "Emerging Markets and Heavy Tails." *Journal of Banking & Finance* 37: 2546–59.
Jansen, D. W., and C. G. de Vries. 1991. "On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective." *The Review of Economics and Statistics* 73: 18–24.
Künsch, H. R. 1989. "The Jackknife and the Bootstrap for General Stationary Observations." *Annals of Statistics* 3: 1217–41.
Loretan, M., and P. C. B. Phillips. 1994. "Testing the Covariance Stationarity of Heavy-Tailed Time Series: An Overview of the Theory with Applications to Several Financial Datasets." *Journal of Empirical Finance* 1: 211–48.
Lux, T. 1996. "The Stable Paretian Hypothesis and the Frequency of Large Returns: An Examination of Major German Stocks." *Applied Financial Economics* 6: 463–75.
Mandelbrot, B. 1963. "The Variation of Certain Speculative Prices." *Journal of Business* 36: 394–419.
McCulloch, J. H. 1996. "Financial Applications of Stable Distributions." In *Handbook of Statistics*, 393–425. Amsterdam: Elsevier.
McCulloch, J. H. 1997. "Measuring Tail Thickness to Estimate the Stable Index α: A Critique." *Journal of Business & Economic Statistics* 15: 74–81.

Nelson, D. B. 1991. “Conditional Heteroskedasticity in Asset Returns: A New Approach.” *Econometrica* 59: 347–70.

Nolan, J. P. 2001. “Maximum Likelihood Estimation and Diagnostics for Stable Distributions.” In *Lévy Processes*, 379–400. Boston: Birkhäuser.

Patton, A. J., and A. Timmermann. 2010. “Monotonicity in Asset Returns: New Tests with Applications to the Term Structure, the CAPM, and Portfolio Sorts.” *Journal of Financial Economics* 98: 605–25.

Politis, D. N., and J. P. Romano. 1993. “On the Sample Variance of Linear Statistics Derived from Mixing Sequences.” *Stochastic Processes and their Applications* 45: 155–67.

Politis, D., J. Romano, and M. Wolf. 1999. *Subsampling*. New York: Springer-Verlak.

Rachev, S., and S. Mittnik. 2000. *Stable Paretian Models in Finance*. New York: Wiley.

Rachev, S. T., C. Menn, and F. J. Fabozzi. 2005. *Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*. Hoboken, NJ: Wiley.

Romano, J. P., and M. Wolf. 2005. “Stepwise Multiple Testing as Formalized Data Snooping.” *Econometrica* 73: 1237–82.

Romano, J. P., and M. Wolf. 2013. “Testing for Monotonicity in Expected Asset Returns.” *Journal of Empirical Finance* 23: 93–116.

Sakov, A., and P. Bickel. 2000. “An Edgeworth Expansion for the M Out of N Bootstrapped Median.” *Statistics & Probability Letters* 49: 217–23.

Scherer, F. M., D. Harhoff, and J. Kukies. 2000. “Uncertainty and the Size Distribution of Rewards from Innovation.” *Journal of Evolutionary Economics* 10: 175–200.

Silverberg, G., and B. Verspagen. 2007. “The Size Distribution of Innovations Revisited: An Application of Extreme Value Statistics to Citation and Value Measures of Patent Significance.” *Journal of Econometrics* 139: 318–39.

Van der Vaart, A. W. 2000. *Asymptotic Statistics*, Vol. 3. Cambridge: Cambridge University Press.

Wolak, F. A. 1987. “An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model.” *Journal of the American Statistical Association* 82: 782–93.

Wolak, F. A. 1989. “Testing Inequality Constraints in Linear Econometric Models.” *Journal of Econometrics* 31: 205–35.