Appl. Math. J. Chinese Univ.
2025, 40(1): 149-168

Robust tests of stock return predictability under heavy-tailed innovations

WONG Hsin-Chieh[1] CHUNG Meng-Hua[2]
FUH Cheng-Der[2] PANG Tian-xiao[3,*]

Abstract. This paper provides a robust test of predictability under the predictive regression model with possible heavy-tailed innovations assumption, in which the predictive variable is persistent and its innovations are highly correlated with returns. To this end, we propose a robust test which can capture empirical phenomena such as heavy tails, stationary, and local to unity. Moreover, we develop related asymptotic results without the second-moment assumption between the predictive variable and returns. To make the proposed test reasonable, we propose a *generalized correlation* and provide theoretical support. To illustrate the applicability of the test, we perform a simulation study for the impact of heavy-tailed innovations on predictability, as well as direct and/or indirect implementation of heavy-tailed innovations to predictability via the unit root phenomenon. Finally, we provide an empirical study for further illustration, to which the proposed test is applied to a U.S. equity data set.

§1 Introduction

One fundamental question in finance is whether future stock returns are predictable using publicly available information such as the dividend-price ratio, the earnings-price ratio, and various measures of the interest rate. The predictive regression model, proposed by [11] and [17], has been used to answer this question in part. Along this line, many studies have been applied to test the predictability based on this model. Take, for example, the papers by [3], [4] and [17], among others.

Under such an econometric method with conventional critical values, one could expect to ask whether there is strong evidence for the predictability of returns. Unfortunately, conventional

Received: 2021-05-16. Revised: 2022-03-07.
MR Subject Classification: 60F05, 62F12.
Keywords: domain of attraction of the normal law, heavy-tailed, least squares estimator, predictive regression, unit root, robust test.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-025-4456-x.
The research of WONG Hsin-Chieh is partially supported by the NSTC(111-2118-M-305-004-MY2), and the research of PANG Tian-xiao is partially supported by the National Social Science Foundation of China(21BTJ067).
*Corresponding author.

tests of the predictability of stock returns could be invalid. More specifically, they reject the null too frequently when the lag of the financial variables is *persistent* and its innovations are *highly correlated* with returns. This concept has attracted much attention in the literature. One celebrated approach is the Bonferroni Q-test proposed by [3].

How the returns of many macroeconomic and financial time series are distributed is an important issue. Since the 1960s, however, empirical evidence has led many to reject the normal assumption in favor of various heavy-tailed alternatives. It is now commonly accepted that financial asset returns are, in fact, heavy-tailed. In other words, it is observed that in many cases, "Normal is not normal", especially for financial data sets. This indicates that the normal assumption might not be suitable for real applications; see, e.g. [1]. For modeling the predictive regression model, there are basically two approaches to overcome this heavy-tailed problem, the heavy-tailed innovation and the time-series innovation. For the time-series innovation, if the innovations are generated by the GARCH model, it exhibits the heavy-tailed feature. However, how the GARCH parameter affects the asymptotic behavior of the estimator is far from clear. A few exceptions can be found in the seminal work of [12], and the references therein. Accordingly, we submit that the heavy-tailed innovations assumption better fits empirical data, which means that heavy-tailed distributions are useful to model certain economic variables and stock price changes; see, e.g. [13].

A typical study in predictive regression is an ordinary least squares (OLS) regression of stock returns onto the lag of the financial variable. Specifically, the model is described as

$$y_t = \beta_0 + \beta_1 x_{t-1} + u_t,$$

where y_t is the predictable variable, say the log excess stock return at time t, and x_{t-1} is a financial variable such as the dividend-price ratio at time $t-1$. In large samples, when u_t is independent and identically distributed (*i.i.d.*) with finite variance or u_t is a martingale difference with $E[u_t^2|\mathcal{F}_{t-1}] < \infty$ ($\mathcal{F}_t \equiv \sigma(u_s; s \leq t)$), CLT (i.e., central limit theorem) based results apply. In this paper, motivated by the empirical fact that the heavy-tailed innovations, one contribution is to derive explicit expressions for the asymptotic behavior of the OLS estimators under the case that the noise distribution exhibits a heavy tail. A similar concern was raised from [13] (in fact, [13] derives explicit finite sample expressions for the tail probabilities of the distribution of the OLS estimator).

On the other hand, most empirically economic/financial data are non-stationary with large volatility. To capture this large volatility, heavy-tailed distributions are often used in the finance and insurance literature, cf. [6] and [10]; while a time-varying volatility framework can be found in [4]. To capture the non-stationary phenomenon under the first-order autoregressive model, one strand of the literature suggests modeling the variables as local-to-unity processes, cf. [3]. These papers assume the form of a first-order auto-regression as

$$x_t = \rho x_{t-1} + e_t$$

with root $\rho = 1 + c/n$, where n denotes the sample size. A prominent application of this theory for empirical study is the construction of confidence intervals (CIs) for autoregressive roots through the inversion of unit root test statistics. Combining local-to-unity processes simultaneously with infinite-variance errors, many literatures raised the concern; see, e.g. [15]. In this paper, this feature attracts attention to us and it can capture the *persistent* issue of the previous discussion.

Local-to-unity asymptotics provide an accurate approximation of the finite-sample distribution of test statistics when the predictor variable is persistent. In fact, it turns out that appropriately centered statistics have limits as $c \to -\infty$ that correspond to the stationary limit theory for fixed $|\rho| < 1$. More specifically, the limit theory in [14] suggests that inversion of appropriately centered test statistics should lead to CIs that correspond to those that apply to the stationary region and are based on stationary asymptotics. Thus, three cases of ρ that need to be considered in this regard are (i) $|\rho| < 1$; (ii) $\rho = \rho_n$, depending on the sample size n, and $n(\rho_n - 1) \to c$, as $n \to \infty$, with $-\infty < c < \infty$; and (iii) $\rho = \rho_n$, depending on the sample size n, and $n(\rho_n - 1) \to -\infty$ as $n \to \infty$.

Motivated by the above observations, this paper considers the following model:

$$y_t = \beta_0 + \beta_1 x_{t-1} + u_t \quad \text{and} \quad x_t = \rho x_{t-1} + e_t, \quad \text{for } t = 1, ..., n. \tag{1}$$

Here parameter ρ is the unknown degree of persistence in the variable x_t. And we will assume that the innovations exhibit the heavy-tailed feature in the sense that they have possibly infinite variances; see Section 3 for details.

As mentioned previously, for testing the predictability of stock returns, Bonferroni Q-test proposed by [3] has solved that the lag of the financial variables is *persistent* and its innovations are *highly correlated* with returns. The key idea of the Bonferroni Q-test is to involve not only the first-order autoregressive root but also the highly correlated correlation with returns. However, the confidence interval of such a Bonferroni procedure is invalid in case (iii), as noted in [16]. Moreover, the correlation coefficient of the predictor variable and returns does not exist when the innovations are heavy-tailed random variables without the finite second-moment assumption. Therefore, the Bonferroni Q-test proposed by [3] is not robust enough and, in the heavy-tailed case, is not even well-defined.

In this paper, motivated by the above observations and seeking to bridge this gap, we study the predictability under possible heavy-tailed innovations and the stationary as well as local-to-unity cases. To be more precise, under the assumption that the innovations have possibly infinite variances, and that the correlation coefficient of the innovations may not exist, we study the limiting distributions of the least squares estimators of β_0 and β_1 under the three cases of ρ defined in (1). On the basis of our method, we study the predictability under possible heavy-tailed innovations assumption and attempt to describe the relationships between heavy-tail, unit root, and predictability. Since the predictive tests using these estimators work well under both heavy-tailed innovations and unit root/near-unit root cases, we term this method a *robust test*.

There are two contributions to this study. First, we investigate the least squares estimators under the predictive regression model. Since the consistency properties of the estimator can be implied by the corresponding results on limiting distributions, we focus on the limiting distributions of the estimators and the corresponding proposed statistics. Although there are three cases of ρ, we can unify the statistics. Theoretically, under the heavy tails, it is difficult to investigate innovations which are highly correlated with returns, because correlation cannot be defined without a finite second-moment assumption. Accordingly, we first introduce a *generalized correlation* without the finite second-moment assumption, and then provide a consistent estimator. To the best of our knowledge, the generalized correlation has not been studied in the literature yet. This paper is a seminal work for considering the correlation without the finite second-moment assumption.

Second, based on the above theoretical results as well as those in [3] and [16], we propose a robust Q-test statistic when the correlation coefficient of innovations does not exist, which we term a *generalized Bonferroni Q-test*. Note that the Bonferroni Q-test proposed by [3] is widely used; however, [16] shows that the confidence interval by such a Bonferroni procedure is invalid in case (iii). To include all three cases, in this paper, we combine these two methods and propose a robust estimator under the heavy-tailed innovations. Our numerical results show that this reformation is precise and useful. Moreover, by using the proposed "generalized Bonferroni Q-test", we study the impact of heavy-tailed innovations on predictability.

The remainder of this paper is organized as follows. Our model setting and theoretical results are given in Section 2. Section 3 presents a generalized Bonferroni Q-test. In Section 4, we conduct simulations to demonstrate the feasibility of our method for finite samples and discuss the interesting issue of predictability. In Section 5, we apply our test procedure to a U.S. equity data set and examine the empirical evidence for predictability and describe their heavy-tailed properties. Section 6 concludes. Due to the reason of space, all proofs of theoretical results are not presented in this paper, but they are available upon request.

§2 Robust inference for predictive regressions with heavy-tailed innovations

In this section, we state the asymptotic results for the least squares estimators of the parameters (β_0, β_1). These results not only support our robust test in Section 3, but also offer some independent contributions. The cases of ρ to be considered in this regard are (i) $|\rho| < 1$; (ii) $\rho = \rho_n$, depending on sample size n, and $n(\rho_n - 1) \to c$, as $n \to \infty$, with $-\infty < c < \infty$; (iii) $\rho = \rho_n$, depending on sample size n, and $n(\rho_n - 1) \to -\infty$ as $n \to \infty$. Note that case (ii) reduces to the unit root case when $\rho = \rho_n \equiv 1$.

2.1 Predictive regressions model with heavy-tailed innovations

Throughout this paper, the innovations $\{(u_t, e_t), t \geq 1\}$ are assumed to be *i.i.d.* bivariate random variables with zero mean and possibly infinite variance. Under the assumption that the innovations have possibly infinite variances, studying the asymptotic distribution of the least squares estimator of $(\beta_0, \beta_1)^T$ is challenging. Therefore, we follow [7] and control the tail behavior to investigate the asymptotic distribution of the parameter estimator.

Assumption 1. *Assume that for all $t \in \mathbb{N} \cup \{0\}$ and $0 \le \eta \le 2 \le \zeta$,*

$$\lim_{x\to\infty} \frac{x^{\zeta-\eta} E\left[|u_t|^\eta I\{|u_t| > x\}\right]}{E|u_t|^\zeta I\{|u_t| \le x\}} = k \quad and \quad \lim_{x\to\infty} \frac{x^{\zeta-\eta} E\left[|e_t|^\eta I\{|e_t| > x\}\right]}{E|e_t|^\zeta I\{|e_t| \le x\}} = k,$$

where $0 \le k = \frac{\zeta-2}{2-\eta} \le \infty$ with $\eta = 2$ if $k = \infty$.

Assumption 1 is a technique to ensure that the tail eventually converges and to allow infinite variances (for example, if Y is a random variable with a Pareto distribution that the density function is given by $f_Y(y) = \frac{2}{y^3} I(y > 1)$, then $EY^2 = \infty$ and $\lim_{x\to\infty} \frac{x^2 E[|Y| I\{|Y|>x\}]}{E|Y|^3 I\{|Y|\le x\}} = 1$). We require merely the form of tail behavior and do not assume any particular global form for the distribution function. This assumption is an important sub-class of heavy-tailed random variables. It is a common technique to analyze the asymptotic properties of the heavy tail. Along this line, by a more general assumption, many studies have been applied to analyze the asymptotic stable distribution, for example, [13], among others. However, in the issue of this paper, we do not have an expression for the location and skewness parameters of the elements of an asymptotic stable distribution. Of course, further research is obviously required, and discussion of these elements of the stable distribution is beyond the scope of this paper. Thus, this paper chooses Assumption 1 to deal with the issue discussed in Section 1. Assumption 1 has a nice property that the random variables $\{u_t\}$ and $\{e_t\}$ are in the domain of attraction of the normal law (DAN, for short). Here let us introduce DAN: a sequence of *i.i.d.* random variables $\{X_n, n \ge 1\}$ belongs to the DAN if there exist two constant sequences $\{A_n, n \ge 1\}$ and $\{B_n, n \ge 1\}$ such that $Z_n := B_n^{-1}(X_1 + \cdots + X_n) - A_n$ converges to a standard normal random variable in distribution, cf. page 172 in [7]. It is known that B_n must take the form $\sqrt{n}c(n)$, where $c(n)$ is a slowly varying function at infinity. This condition holds under appropriate primitive assumptions. For example, if $\{u_t\}$ is a sequence of *i.i.d.* random variables with zero mean and finite second moment $\sigma^2 = E[u_1^2]$, then the condition follows from the classical CLT with $B_n = \sqrt{n}\sigma$. In summary, Assumption 1 allows that the innovations can be either light-tailed (although the impact of heavy-tailed innovations on stock return predictability is inevitable, the innovations are assumed to have finite second and/or higher moments in most literature, and even assumed to be normally distributed in the predictive regressions model. Related works include but are not limited to [3] and [8], among others) or heavy-tailed distribution.

Another difficulty here is the correlation coefficient of u_t and e_t, which does not exist without the finite second-moment assumption. Hence we present a concept of correlation without the second-moment assumption. Before doing so, let us define

$$\begin{cases} l_X(t) = EX^2 I\{|X| \le t\}, \quad b_X = \inf\{t \ge 1 : l_X(t) > 0\}, \\ d_{1j} = \inf\{s : s \ge b_X + 1, \frac{l_X(s)}{s^2} \le \frac{1}{j}\}, \quad \text{for } j = 1, 2, 3, ..., \\ l_Y(t) = EY^2 I\{|Y| \le t\}, \quad b_Y = \inf\{t \ge 1 : l_Y(t) > 0\}, \\ d_{2j} = \inf\{s : s \ge b_Y + 1, \frac{l_Y(s)}{s^2} \le \frac{1}{j}\}, \quad \text{for } j = 1, 2, 3, ..., \\ X^{(n)} = XI\{|X| \le d_{1n}\}, \quad Y^{(n)} = YI\{|Y| \le d_{2n}\}, \quad \text{for } n = 1, 2, 3, ... \end{cases} \tag{2}$$

Definition 1. *For a given bivariate random variable (X, Y) such that each of X and Y satisfies*

Assumption 1 with possibly infinite variance, if

$$\frac{E(X^{(n)}Y^{(n)}) - EX^{(n)}EY^{(n)}}{\sqrt{l_X(d_{1n})l_Y(d_{2n})}} \to \alpha_{XY}, \quad \text{as } n \to \infty, \tag{3}$$

then we call the constant α_{XY} as the generalized correlation of X and Y.

The *generalized correlation* of u_t and e_t is assumed to be negative. Note that this assumption is almost the same as that in [3], but is not restrictive since the sign of β_1 is unrestricted; redefining the predictor variable as x_t flips the signs of both β_1 and the correlation between the innovations. For convenience, the notation δ will denote the generalized correlation of u_t and e_t.

Next, we provide a proposition that the correlation is well-defined.

Proposition 2.1. (a) *When both X and Y have finite variances, the generalized correlation defined in* (3) *reduces to the classical correlation.* (b) *The generalized correlation of α_{XY} satisfies*

$$|\alpha_{XY}| \leq 1,$$

which still holds for the case of possibly infinite variances.

Finally, in the $AR(1)$ model, x_k is obtained from knowing the value of x_{k-1}, and in turn x_{k-1} is obtained from knowing x_{k-2}, and so on. Observe that x_k, for any $k \in \mathbb{N}$, can be obtained from an initial value x_0 that is k time periods prior. Thus, we need an assumption on the initial point x_0 relating to sample size.

Assumption 2. $x_0 = o_p(\sqrt{n})$.

2.2 Asymptotic results

In this subsection, we are going to show asymptotic results for the least squares estimators of the parameters (β_0, β_1). Before stating the procedure of our asymptotic results, we define some notations. Hereafter, we use "$\xrightarrow{d}$" to denote convergence in distribution. Note that, the least squares estimator of $(\beta_0, \beta_1)^T$ can be written, in matrix form, as

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} n & \sum_{t=1}^n x_{t-1} \\ \sum_{t=1}^n x_{t-1} & \sum_{t=1}^n x_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^n y_t \\ \sum_{t=1}^n x_{t-1} y_t \end{pmatrix}. \tag{4}$$

To start with, $|\rho| < 1$ means that the sequence $\{x_n, n \geq 1\}$ is modeled by a stationary $AR(1)$ model. We have the following asymptotic properties for the least squares estimators of the unknown parameters.

Theorem 2.1. *Let $|\rho| < 1$ in model (1). Suppose Assumptions 1–2 hold. As $n \to \infty$, we have*

$$\begin{pmatrix} \sqrt{\frac{n}{\hat{\sigma}_1^2}}(\hat{\beta}_0 - \beta_0) \\ \sqrt{\frac{n\hat{\sigma}_2^2}{(1-\hat{\rho}^2)\hat{\sigma}_1^2}}(\hat{\beta}_1 - \beta_1) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \tag{5}$$

and

$$\sqrt{\frac{\sum_{t=1}^n x_{t-1}^2}{\hat{\sigma}_1^2}}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0,1), \tag{6}$$

where Z_1 and Z_2 are two independent standard normal random variables, and $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are defined in (12) below.

Next, we consider the asymptotic properties of the least squares estimator of $(\beta_0, \beta_1)^T$ when the sequence $\{x_n, n \geq 1\}$ is modeled by a nearly unit root model. That is, in the following theorems, we replace the above ρ by $\rho_n = 1 + c_n/n$, such that $0 > c_n \to c \in [-\infty, \infty)$ as $n \to \infty$.

First, we introduce some notations. Let

$$l_1(t) = Eu_1^2 I\{|u_1| \leq t\}, \quad b_1 = \inf\{t \geq 1 : l_1(t) > 0\},$$
$$\eta_{1j} = \inf\{s : s \geq b_1 + 1, \frac{l_1(s)}{s^2} \leq \frac{1}{j}\}, \quad \text{for } j = 1, 2, 3, ...,$$

and

$$l_2(t) = Ee_1^2 I\{|e_1| \leq t\}, \quad b_2 = \inf\{t \geq 1 : l_2(t) > 0\},$$
$$\eta_{2j} = \inf\{s : s \geq b_2 + 1, \frac{l_2(s)}{s^2} \leq \frac{1}{j}\}, \quad \text{for } j = 1, 2, 3, ...$$

Then we denote

$$u_t^{(1)} = u_t I\{|u_t| \leq \eta_{1n}\} - Eu_t I\{|u_t| \leq \eta_{1n}\}, \quad e_t^{(1)} = e_t I\{|e_t| \leq \eta_{2n}\} - Ee_t I\{|e_t| \leq \eta_{2n}\}.$$

Lemma 2.1. *Under Assumptions 1-2, we have the following weak invariance principle:*

$$\sum_{t=1}^{[nr]} \begin{pmatrix} u_t^{(1)}/\sqrt{nl_1(\eta_{1n})} \\ e_t^{(1)}/\sqrt{nl_2(\eta_{2n})} \end{pmatrix} \Rightarrow \begin{pmatrix} W_1(r) \\ W_2(r) \end{pmatrix} = BM(0, \Sigma(r))$$

as $n \to \infty$, where

$$\Sigma(r) = r \cdot \lim_{n\to\infty} E\left[\left(u_t^{(1)}/\sqrt{l_1(\eta_{1n})}, e_t^{(1)}/\sqrt{l_2(\eta_{2n})}\right)^T \left(u_t^{(1)}/\sqrt{l_1(\eta_{1n})}, e_t^{(1)}/\sqrt{l_2(\eta_{2n})}\right)\right]$$

is the covariance matrix of the bivariate Brownian motion indexed by $r \in [0, 1]$.

Theorem 2.2. *Let $\rho = \rho_n = 1 + c_n/n$ in model (1) such that $0 > c_n \to c \in (-\infty, \infty)$ as $n \to \infty$. Suppose Assumptions 1–2 hold. As $n \to \infty$, we have*

$$\begin{pmatrix} \sqrt{\frac{n}{\hat{\sigma}_1^2}}(\hat{\beta}_0 - \beta_0) \\ n\sqrt{\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2}}(\hat{\beta}_1 - \beta_1) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & \int_0^1 I_c(r)dr \\ \int_0^1 I_c(r)dr & \int_0^1 I_c^2(r)dr \end{pmatrix}^{-1} \begin{pmatrix} W_1(1) \\ \int_0^1 I_c(r)dW_1(r) \end{pmatrix} \tag{7}$$

and

$$\sqrt{\frac{\sum_{t=1}^n (x_{t-1} - \bar{x}_{n-1})^2}{\hat{\sigma}_1^2}}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} \frac{\int_0^1 I_c(r)dW_1(r) - W_1(1)\int_0^1 I_c(r)dr}{\sqrt{\int_0^1 I_c^2(r)dr - (\int_0^1 I_c(r)dr)^2}}. \tag{8}$$

Here and afterwards, $W_1(\cdot)$ and $W_2(\cdot)$ are defined in Lemma 2.1, and $I_c(r) = W_2(r) + c\int_0^r \exp((r-s)c)W_2(s)ds$ is an Ornstein-Uhlenbeck process.

When $c_n \to c = -\infty$, we have the following asymptotic properties for the least squares estimators of the unknown parameters.

Theorem 2.3. *Let $\rho = \rho_n = 1 + c_n/n$ in model (1) such that $0 > c_n \to c = -\infty$ as $n \to \infty$.*

Suppose Assumptions 1–2 hold. As $n \to \infty$, we have

$$\begin{pmatrix} \sqrt{\frac{n}{\hat\sigma_1^2}}(\hat\beta_0 - \beta_0) \\ \sqrt{\frac{n\hat\sigma_2^2}{(1-\hat\rho^2)\hat\sigma_1^2}}(\hat\beta_1 - \beta_1) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_3 \\ Z_4 \end{pmatrix} \tag{9}$$

and

$$\sqrt{\frac{\sum_{t=1}^n x_{t-1}^2}{\hat\sigma_1^2}}(\hat\beta_1 - \beta_1) \xrightarrow{d} N(0,1), \tag{10}$$

where Z_3 and Z_4 are two independent standard normal random variables.

Note that from (6), (8), and (10), we can unify the statistics to have

$$\sqrt{\frac{\sum_{t=1}^n (x_{t-1} - \bar x_{n-1})^2}{\hat\sigma_1^2}}(\hat\beta_1 - \beta_1)$$

$$\xrightarrow{d} \begin{cases} N(0,1), \quad when\ |\rho| < 1, \\ \frac{\int_0^1 W_2(r)dW_1(r) - W_1(1)\int_0^1 W_2(r)dr}{\sqrt{\int_0^1 W_2^2(r)dr - (\int_0^1 W_2(r)dr)^2}}, \quad when\ \rho = \rho_n = 1 + c_n/n\ and\ 0 > c_n \to 0, \\ N(0,1), \quad when\ \rho = \rho_n = 1 + c_n/n\ and\ 0 > c_n \to c = -\infty. \end{cases}$$

Therefore, whichever ρ lies in the above three situations of ρ, one can always use the statistics $\sqrt{\frac{\sum_{t=1}^n (x_{t-1} - \bar x_{n-1})^2}{\hat\sigma_1^2}}(\hat\beta_1 - \beta_1)$ to construct confidence interval and/or hypothesis testing for the parameter β_1.

Theorem 2.4. *Suppose Assumptions 1–2 hold. Denote the generalized correlation of u_1 and e_1 as δ. Then for each of the following three cases—(i) ρ is fixed and $|\rho| < 1$, (ii) $\rho = \rho_n = 1+c_n/n$ such that $0 > c_n \to c \in (-\infty, \infty)$ as $n \to \infty$, (iii) $\rho = \rho_n = 1+c_n/n$ such that $0 > c_n \to c = -\infty$ as $n \to \infty$—one has as $n \to \infty$,*

$$\frac{\hat\sigma_{12}}{\hat\sigma_1\hat\sigma_2} \longrightarrow \delta \ in\ probability,$$

where $\hat\sigma_{12}$ is defined in (12) below.

To study the property of generalized correlation, we propose a consistent estimator in Theorem 2.4, which is of independent interest, and can be applied to the case of an undefined correlation. Specifically, when the second moment of the distribution does not exist, and hence the correlation cannot be calculated directly. For example, if X is a random vector from a t-distribution with location vector 0 and scale matrix Σ, written as $X \sim t_v(0, \Sigma)$. When the degree of freedom v is larger than 2, then $Cov(X) = \Sigma v/(v-2)$, and the correlation matrix $Cor(X)$ equals the scale matrix Σ. When the degree of freedom v is less than or equal to 2, we can apply Theorem 2.4 to define the generalized correlation matrix $Cor(X)$. That is, we extend the scale matrix Σ to the case where the second moment does not exist. To illustrate the applicability, we also use this example in the Monte Carlo simulation study.

Note that in our setting, we do not assume u_t and e_t to be independent for any given t, although this is a common regularity condition in the literature of predictive regressions. In most literature, it is also assumed that the correlation coefficient of u_t and e_t exists and is even given in advance, cf. [3]. In Theorems 2.1–2.4, we do not require the existence of the correlation coefficient of u_t and e_t. Furthermore, the results in Theorems 2.1, 2.3 and 2.4 are irrelevant to the dependence structure of u_t and e_t. Hence, when $|\rho| < 1$ or $\rho = \rho_n$ with

$n(\rho_n - 1) \to -\infty$, the proposed robust statistics in Theorems 2.1 and 2.3 are indeed pivotal, and can be applied directly to test the predictability of stock returns in predictive regression models. If the dependent information of u_n and e_n, $\Sigma(\cdot)$ defined in Lemma 2.1, is known in advance, then Theorem 2.2 shares the same property.

§3 Methodology: A generalized Bonferroni Q-test

Consider the predictive regression model defined in (1). The problem of interest is testing the null hypothesis of no predictability, i.e., $H_0 : \beta_1 = 0$. In this section, we propose a robust test based on Theorems 2.1–2.4 in this paper, [3], and [16] under heavy-tailed distributions. In the case of light-tailed distributions, the reader is referred to [2] for details of the Bonferroni Q-test. Note that the innovations in [2] are assumed to be the normal distribution. Although there is no objection to the plausibility for this hypothesis, little empirical evidence has been given to support it. On the other hand, assuming heavy-tailed innovations has been found to be more appropriate for fitting the empirical data. Next, we remedy the situation in which the Bonferroni procedure is invalid in case (iii). Since the proposed testing statistic captures the phenomena of heavy tails, stationary, and local to unity, we call the proposed test a "generalized Bonferroni Q-test".

Given the observations $\{(x_{t-1}, y_t), t = 1, \ldots, n\}$, denote $\bar{x}_{n-1} = \sum_{t=0}^{n-1} x_t/n$ and $\bar{y}_n = \sum_{t=1}^{n} y_t/n$. Then the least squares estimators of the unknown parameters (β_0, β_1) are

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n}(x_{t-1} - \bar{x}_{n-1})(y_t - \bar{y}_n)}{\sum_{t=1}^{n}(x_{t-1} - \bar{x}_{n-1})^2}, \quad \hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_{n-1}. \tag{11}$$

To test predictability under model (1) with heavy-tailed distributions, we need a confidence interval of β_1. A valid confidence interval can be constructed through the following steps.

1. Run the first regression in (1) to obtain the standard error for $\hat{\beta}_1$, denoted as $\mathrm{SE}(\hat{\beta}_1)$. Run the second regression in (1) to obtain the standard error for $\widehat{\rho}$, denoted as $\mathrm{SE}(\widehat{\rho})$. Compute
$$\begin{cases} \hat{\sigma}_1^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{t-1})^2, \\ \hat{\sigma}_2^2 = \frac{1}{n}\sum_{t=1}^{n}(x_t - \hat{\rho} x_{t-1})^2, \text{ where } \hat{\rho} = \sum_{t=1}^{n} x_t x_{t-1} / \sum_{t=1}^{n} x_{t-1}^2, \\ \hat{\sigma}_{12} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{t-1})(x_t - \hat{\rho} x_{t-1}), \\ \hat{\delta} = \hat{\sigma}_{12}/(\hat{\sigma}_1 \hat{\sigma}_2). \end{cases} \tag{12}$$

2. Compute the nuisance parameter c with $\rho = 1 + c/n$. A simple estimator of c is $\widehat{c} = n(\widehat{\rho} - 1)$, in which $\widehat{\rho}$ is the OLS estimator of ρ. Next, based on our simulation study, we modify the original estimator $\widehat{c}$ of c for having a more precise estimator (the adjustment is based on simulation):
$$c^* = \lceil \widehat{c} + 1 \rceil = \lceil n(\widehat{\rho} - 1) + 1 \rceil. \tag{13}$$
In the following Bonferroni procedure, denote c_0 as a benchmark constant and consider that $|c|$ is large whenever $|c^*| \geq c_0$.

3. When $|c^*| < c_0$, move to Step 4. When $|c^*| \geq c_0$, move to Step 5.

To construct a Bonferroni confidence interval for β_1, we first construct a $100(1-\alpha_1)\%$ confidence interval for ρ, denoted as $C_\rho(\alpha_1)$. For each value of ρ in the confidence interval, we then construct a $100(1-\alpha_2)\%$ confidence interval for β_1 given ρ, denoted as $C_{\beta_1|\rho}(\alpha_2)$. A confidence interval that does not depend on ρ can be obtained by $C_{\beta_1}(\alpha) = \bigcup_{\rho \in C_\rho(\alpha_1)} C_{\beta_1|\rho}(\alpha_2)$. By Bonferroni's inequality, this confidence interval has coverage of at least $100(1-\alpha)\%$, where $\alpha = \alpha_1 + \alpha_2$.

4. When $|c^*| < c_0$, compute the Dickey-Fuller generalized least squares (DF-GLS) statistic of [5]. Given the values of the DF-GLS statistic and $\widehat{\delta}$, we can find an appropriate confidence interval $[c_L, c_U]$ for c from the tables of [2]. Proceed directly to Step 6.

5. When $|c^*| \geq c_0$, we use the unit root t-test, $t_{\widehat{\rho}} = \frac{\widehat{\rho}-1}{\sigma_{\widehat{\rho}}}$. Under heavy-tailed distributions, we have that for any fixed c, as $n \to \infty$,
$$t_{\widehat{\rho}} = \frac{n(\widehat{\rho}-\rho) + n(\rho-1)}{\left\{\widehat{\sigma}_2^2 / \left(n^{-2}\sum_{t=1}^n x_{t-1}^2\right)\right\}^{1/2}} \xrightarrow{d} \frac{\int_0^1 J_c(r)dW(r)}{\left(\int_0^1 J_c(r)^2 dr\right)^{1/2}} + c\left(\int_0^1 J_c(r)^2 dr\right)^{1/2} =: \tau_c, \tag{14}$$
where $J_c(r) = \int_0^r e^{c(r-s)} dW(s)$ is a linear diffusion and W is the standard Brownian motion. The limit representation for τ_c holds for all $c \in \mathbb{R}$. In what follows, we concentrate on the half line $(-\infty, 0)$. By using a simple modification of [14], we can prove that the two components of (14) satisfy
$$\lambda_c = \frac{\int_0^1 J_c(r)dW(r)}{\left(\int_0^1 J_c(r)^2 dr\right)^{1/2}} \xrightarrow{d} \xi \equiv N(0,1) \text{ and } (-2c)\int_0^1 J_c(r)^2 dr \to 1 \text{ in probability}, \tag{15}$$
as $c \to -\infty$. This implies that as $c \to -\infty$, the asymptotic form of τ_c in a suitably expanded probability space is
$$\tau_c = -\frac{|c|^{1/2}}{2^{1/2}} + \frac{1}{2}\xi + O_p(|c|^{-1/2}) \sim N\left(-\frac{|c|^{1/2}}{2^{1/2}}, \frac{1}{4}\right) + O_p(|c|^{-1/2}). \tag{16}$$
After an appropriate transform, the $100(1-\alpha_1)\%$ confidence interval for c is
$$[c_L, c_U] = [-2(\widehat{\tau}^2 - z_{\alpha_1/2}\widehat{\tau}), -2(\widehat{\tau}^2 + z_{\alpha_1/2}\widehat{\tau})], \tag{17}$$
where $\hat{\tau} := t_{\widehat{\rho}}$ is the estimator of τ_c defined in (14).

6. The $100(1-\alpha_1)\%$ confidence interval for ρ is
$$[\rho_L, \rho_U] = [1 + c_L/n, 1 + c_U/n]. \tag{18}$$

7. The Q-statistic is proposed by [3]:
$$\beta_1(\rho) = \frac{\sum_{t=1}^N x_{t-1}^\mu [y_t - \beta_{ue}(x_t - \rho x_{t-1})]}{\sigma_1(1-\delta^2)^{1/2}[\sum_{t=1}^n (x_{t-1}^\mu)^2]^{1/2}}, \tag{19}$$
where $\beta_{ue} = \sigma_{12}/\sigma_2^2$ is negative under our assumption and x_{t-1} has been replaced by its demeaned counterpart x_{t-1}^μ. To be more precise, for given two values $\rho = [\rho_L, \rho_U]$, compute an equal-tailed $100(1-\alpha_2)\%$ confidence interval for β_1 given ρ as follows. Run the first regression, replacing y_t with $y_t - \hat{\sigma}_{12}\hat{\sigma}_2^{-2}(x_t - \rho x_{t-1})$. Let $\hat{\beta}_1(\rho)$ denote the

coefficient on x_{t-1}. The confidence interval for β_1 given ρ is $[\beta_{1L}(\rho), \beta_{1U}(\rho)]$, where $\beta_{1L}(\rho) = \hat{\beta}_1(\rho) - z_{\alpha_2/2}(1-\hat{\delta}^2)^{1/2}SE(\hat{\beta}_1)$, $\beta_{1U}(\rho) = \hat{\beta}_1(\rho) + z_{\alpha_2/2}(1-\hat{\delta}^2)^{1/2}SE(\hat{\beta}_1)$, with $z_{\alpha_2/2}$ being the $1-\alpha_2/2$ quantile of the standard normal distribution. Then the $100(1-\alpha_2)\%$ confidence interval for β_1 is $C_{\beta_1|\rho}(\alpha_2) = [\beta_{1L}(\rho, \alpha_2),\ \beta_{1U}(\rho, \alpha_2)]$.

8. Then the $100(1-\alpha)\%$ Bonferroni confidence interval is given by

$$C_{\beta_1}(\alpha) = [\beta_{1L}(\rho_U), \beta_{1U}(\rho_L)]. \tag{20}$$

Note that the $100(1-\alpha)\%$ Bonferroni confidence interval $C_{\beta_1}(\alpha)$ defined in (20) corresponds to an $\alpha\%$ two-sided test under the null hypothesis $\beta_1 = 0$, but the same idea can be applied to an $\frac{\alpha}{2}\%$ one-sided test.

It is worth mentioning that the estimators in the above Bonferroni procedure are indeed consistent under Assumptions 1 and 2. Hence the proposed generalized Bonferroni Q-test is robust. The term *robust* is either in the sense of heavy-tailed distributions, or in the sense of local to unity, or both.

§4 Finite sample study

To illustrate the performance of the proposed generalized Bonferroni Q-test, in this section, we conduct a Monte Carlo simulation study. Subsection 4.1 checks the robustness of the generalized Bonferroni Q-test under a large nuisance parameter $|c|$ and heavy-tailed innovations. Then we discuss the relationships among heavy tails, unit root, and predictability in Subsection 4.2.

4.1 Rejection rates and powers

To illustrate the applicability of the proposed generalized Bonferroni Q-test for light-tailed and heavy-tailed innovations, we conduct a finite sample performance of the proposed testing statistic by means of an extensive Monte Carlo study. For the testing experiments with null hypothesis $H_0 : \beta_1 = 0$ against alternative $H_1 : \beta_1 > 0$, all tests are evaluated at the 5% significance level. For ease of presentation, the nuisance parameter is normalized as $\beta_0 = 0$ in this section. The simulation studies are based on $10,000$ Monte Carlo draws of the sample path using model (1) with the initial condition $x_0 = 0$. Here $c_0 = -20$ as the threshold for $|c|$ to be large. Note that when $|c| > 20$, the traditional Bonferroni Q-test is not sufficiently robust for practical use. This will be shown in the simulation results later in this subsection.

In what follows, we show the rejection rates using the methods proposed by [3] and our generalized Bonferroni Q-test. Since the theories mentioned in Step 5 of the generalized Bonferroni Q-test procedure are under $n \to \infty$ and $c \to -\infty$ when $|c^*| \geq c_0$, we choose $|c|$ to be proportional to n, and consider different sample sizes n for different degrees of persistence ($\rho = 1 + c/n$) in the large $|c|$ experiment.

Table 1 reports finite-sample rejection rates of one-sided, right-tailed tests of predictability at the 5% significance level under different innovation assumptions. Under the assumption of

Table 1. Rejection rates.

			Method		
c	n	ρ	CY(2006)	Proposed result	$\lvert c\rvert/n$
Panel A: Bivariate normal innovations					
0		1	0.0500	0.0500	0
-2	250	0.992	0.0471	0.0471	0.0080
-20		0.92	0.0620	0.0525	0.0800
-40	550	0.927273	0.0368	0.0496	0.0727
-50	700	0.928571	0.0370	0.0505	0.0714
-60	850	0.929412	0.0528	0.0498	0.0705
-70	1000	0.93	0.6687	0.0499	0.0700
-80	1150	0.930435	0.9950	0.0501	0.0696
-90	1300	0.930769	1	0.0500	0.0692
-100	1450	0.931034	1	0.0498	0.0690
Panel B: Bivariate t-distributed innovations					
0		1	0.0468	0.0468	0
-2	250	0.992	0.0452	0.0452	0.0080
-20		0.92	0.0523	0.0506	0.0800
-40	300	0.866667	0.0301	0.0501	0.1333
-50	400	0.875	0.0366	0.0504	0.1250
-60	500	0.88	0.0522	0.0498	0.1200
-70	650	0.892308	0.7143	0.0502	0.1077
-80	750	0.893333	0.9757	0.0498	0.1067
-90	900	0.9	0.9967	0.0500	0.1000
-100	1050	0.904762	0.9976	0.0499	0.0952
Panel C: Bivariate Pareto innovations					
0		1	0.0464	0.0464	0
-2	250	0.992	0.0447	0.0447	0.0080
-20		0.92	0.0509	0.0496	0.0800
-40	280	0.857143	0.0313	0.0502	0.1429
-50	380	0.868421	0.0358	0.0497	0.1316
-60	480	0.875	0.0518	0.0495	0.1250
-70	600	0.883333	0.6987	0.0501	0.1167
-80	700	0.885714	0.9746	0.0494	0.1143
-90	850	0.894118	0.9955	0.0500	0.1059
-100	1000	0.9	0.9975	0.0500	0.1000

This table reports the finite-sample rejection rates of one-sided, right-tailed tests of predictability at the 5% significance level. The innovations have correlation $\delta = -0.95$. The innovations in Panels A, B, and C are drawn from a bivariate normal distribution with mean zero and unit variance, a bivariate t-distribution with degrees of freedom 2, and a Pareto distribution with shape parameter 2 and scale parameter 1, respectively. The rejection rates are rounded off to the fourth decimal digit. Note: CY(2006) is the results for [3].

bivariate normal innovations, Panel A shows relatively robust performance for the generalized Bonferroni Q-test when $|c|$ is large. In the case where innovations $\{(u_t, e_t), t \geq 1\}$ in (1) are heavy-tailed distributions with possible infinite variances, we conduct an analysis in Panel B for bivariate t-distributed innovations with degrees of freedom 2. Note that when the degree of freedom is less than or equal to 2, the second moment does not exist. The simulation outcomes in Panel B show that the generalized Bonferroni Q-test is robust under heavy-tailed innovations, which provides a finite sample support of our theoretical results. As an additional robustness check, we repeat the simulation experiment under a different distribution. Here we consider the bivariate Pareto innovations with shape parameter 2 and scale parameter 1. These parameters are chosen to have infinite variance. The simulation outcomes are reported in Panel C.

Table 1 presents the simulation results for $|c|$ to be proportional to n. However, in real applications, it is not easy to determine a precise relationship between $|c|$ and n. To study the sensitivity of $|c|$ proportional to n, we conduct an experiment for various $|c|$ with the same sample size. Table 2 shows that the rejection rates of our results are all near 0.05 for a suitable range of $|c|$ with the same sample size $n = 1,000$.

Table 2. Robustness under large $|c|$.

c	δ	n	ρ	Rejection rates	$\|c\|/n$
-20	-0.95	1000	0.98	0.0468	0.02
-40	-0.95	1000	0.96	0.0473	0.04
-50	-0.95	1000	0.95	0.0480	0.05
-60	-0.95	1000	0.94	0.0499	0.06
-70	-0.95	1000	0.93	0.0500	0.07
-80	-0.95	1000	0.92	0.0512	0.08
-90	-0.95	1000	0.91	0.0531	0.09
-100	-0.95	1000	0.90	0.0570	0.10

This table reports the finite-sample rejection rates of one-sided, right-tailed tests of predictability at the 5% significance level. The innovations have correlation $\delta = -0.95$ and are drawn from bivariate normal innovations with zero mean and unit variance. The rejection rates are rounded off to the fourth decimal digit.

In summary, we show that the proposed generalized Bonferroni Q-test is rather precise and robust in the sense of local to unity and heavy-tailed innovations. To be more precise, the simulation outcomes from Tables 1–2 show that if $|c|/n$ is in $[0, 0.15]$, i.e., $\rho \in [0.85, 1]$, then the proposed generalized Bonferroni Q-test is a robust test.

Next, we examine the power properties of the statistics. We follow [3] to consider a sequence of alternatives of the form $\beta_1 = b/n$ for some fixed constant b. Here we present the results for sample size $n = 500$ and correlations $\delta = -0.95$. Figure 1 presents the power of the two tests with three distributed innovations for $c = -20$ and $c = -40$. For these results, our generalized Bonferroni Q-test possesses a similar power to that of the Bonferroni Q-test when $c = -20$; however, it dominates that of the Bonferroni Q-test when $c = -40$. These conclusions correspond to our theorems. One is that the estimators of the predictive regression are robust under heavy-tailed innovations and can be used directly in the test; thus the power properties are the same when we use the same test. The other is that when $|c|$ is large, our generalized Bonferroni Q-test has a higher power than the traditional Bonferroni Q-test. Clearly the proposed generalized Bonferroni Q-test is taking effect.

4.2 Predictability

The issue of predictability is addressed in this subsection, in which we discuss the impact of heavy-tailed innovations and a unit root on predictability, respectively. First, we consider a direct impact of heavy-tailed innovations on predictability via a hypothesis testing point of view. The test of predictability under the null hypothesis $\beta_1 = 0$ means no predictability. We use our generalized Bonferroni Q-test and calculate the p-value, which will be compared with the significance level. If the p-value is smaller than the significance level, we reject the null hypothesis. As shown in Panel A of Table 3, our simulation results show that under the null

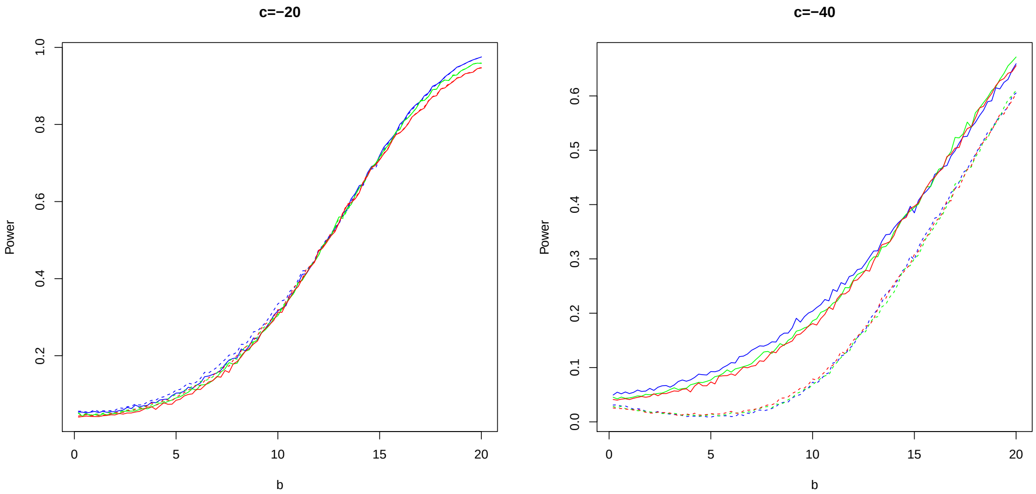

Figure 1. Power plots for sample size $n = 500$ and correlation of innovations $\delta = -0.95$.

Note: dotted lines and solid lines stand for the traditional Bonferroni Q-test and the generalized Bonferroni Q-test, respectively. For the innovations assumption, the blue, green, and red correspond to the normal distribution, the t-distribution with degrees of freedom 2, and the bivariate Pareto with shape parameter 2 and scale parameter 1.

hypothesis $H_0 : \beta_1 = 0$, when the innovations in model (1) have heavier tails, the rejection rates of the predictive test decrease. That is, as the tail distribution becomes heavier, the probability of making wrong rejections decreases. In other words, the heavy-tailed property decreases the predictability.

Next, we discuss an indirect impact of the heavy tail on predictability via the property of unit root ρ. That is, we first use the hypothesis testing point of view to discuss the impact of a heavy tail on the unit root. The DF-GLS statistic is computed as follows: regress $(x_0, x_1 - \rho_{GLS}x_0, ..., x_n - \rho_{GLS}x_{n-1})^T$ onto $(1, 1 - \rho_{GLS}, ..., 1 - \rho_{GLS})^T$, where $\rho_{GLS} = 1 - 7/n$, to obtain the coefficient μ_{GLS}. Let $\overline{x}_t = x_t - \mu_{GLS}$. Run the regression without the intercept: $\Delta\overline{x}_t = \theta\bar{x}_{t-1} + e_t$. The t-statistic for θ is the DF-GLS statistic: $t_{\hat{\theta}} = \hat{\theta}/\hat{\sigma}_{\hat{\theta}}$ with $\hat{\sigma}^2_{\hat{\theta}} = \hat{\sigma}^2_2 / \sum_{t=1}^{n} \bar{x}^2_{t-1}$. The null hypothesis is $\theta = 0$, which means a unit root. If the p-value of the test is smaller than the significance level, we reject the null hypothesis. Under the null hypothesis $\theta = 0$ (or $\rho = 1$), our simulation results in the Panel B of Table 3 show that if the innovations in the model (1) have a heavier tail, the rejection rates of the unit root test decrease. That is, the unit root tests have a smaller probability of making wrong rejections as the tail of distribution becomes heavier. In other words, the heavy-tailed properties are strong evidence of a unit root.

Last, we discuss the impact of unit root ρ on predictability β_1 based on a testing experiment with null hypothesis $H_0 : \beta_1 = 0$ against alternative $H_1 : \beta_1 > 0$. This relation can be seen easily according to the formula of the generalized Bonferroni Q-test. When the persistence factor ρ increases, the value of $\beta_1(\rho)$ decreases. The confidence interval of β_1 moves to the left, and hence has a higher probability to cover the value of $\beta_1 = 0$. The p-value increases and the probability of rejecting the null hypothesis decreases, as shown in Panel C of Table 3. That means the predictability decreases. This phenomenon is also shown in Figure 4 of [3].

In summary, from the direct and/or the indirect impact of heavy-tailed innovations via unit root on predictability, these three factors are strongly related. All in all, they share the same financial intuition: when the data set has a heavy-tailed distribution, the variance becomes larger, and it is more difficult to capture its trend. This makes predictability more difficult.

Table 3. Relationship between heavy tails, unit root, and predictability.

Panel A Heavy tail - predictability			**Panel B Heavy tail - unit root**			**Panel C Unit root - predictability**		
Distribution	ρ	R.R.	Distribution	ρ	R.R.	Distribution	ρ	R.R.
$N(0,1)$		0.0500	$N(0,1)$		0.0522		0.98	0.0468
$t(1000)$		0.0488	$t(1000)$		0.0518		0.96	0.0473
$t(10)$	1	0.0474	$t(10)$	1	0.0502	$N(0,1)$	0.95	0.0480
$t(2)$		0.0461	$t(2)$		0.0473		0.94	0.0499
$t(1)$		0.0427	$t(1)$		0.0308		0.93	0.0500

This table reports how the three factors impact each other. In Panel A, we set $n = 250$ and $\rho = 1$; the null hypothesis is $\beta_1 = 0$ against the alternative $\beta_1 > 0$. In Panel B, we set $n = 1000$ and $\rho = 1$ with a null hypothesis of $\rho = 1$ against the alternative $\rho < 1$. In Panel C, the null hypothesis is $\beta_1 = 0$ against the alternative $\beta_1 > 0$ and we set $n = 1000$ with different ρ as in Table 2. R.R. stands for Rejection rates.

§5 Empirical study

The traditional Bonferroni Q-test in [3] is widely used to test predictability, where involves that the predictor variable is persistent and its innovations are highly correlated with returns. However, this test is based on the normal innovations assumption, which might not be suitable for real applications. To further illustrate the effectiveness of the generalized Bonferroni Q-test under heavy-tailed distributions, in this section we implement the proposed method for predictability on a U.S. equity data set. During the study of this empirical data set, we observe that the heavy-tailed assumption is more appropriate. Moreover, we note that the generalized Bonferroni Q-test is robust.

In this study, we treat monthly CRSP value-weighted log excess returns as dependent variables. Moreover, we use seven variables as potential predictors in the predictability test. There are two time periods in this empirical study. The first time period, from January 1927 to December 2016, contains the dividend-price ratio, earnings-price ratio, dividend-payout ratio, long-term yield, and default yield spread. The second time period, from January 1934 to December 2016, contains the 3-month T-Bill and the term-spread. The monthly S&P 500 dividends and earnings values are available on the Robert Shiller online data set. The dividend-payout ratio is computed as dividends divided by earnings. The long-term yield is the long-term US government bond yield, also from the Robert Shiller online data set. The default yield spread is the difference between Moody's seasoned Baa corporate bond yield and Moody's seasoned Aaa corporate bond yield taken from the economic research database at the Federal Reserve at St. Louis (FRED). The 3-month U.S. Treasury bill rate is also taken from FRED. The difference between the long-term yield and the T-bill rate is the term spread. These variables have been widely used in predictability tests. The reader is referred to [8] for details on the data and examples of studies that have utilized these variables.

On the basis of the data sets, we first show that the heavy-tailed assumption is more appropriate than normal innovations. Numerous papers mention that the innovations of predictive regression are heavy-tailed, typically fitting them using t-distributions with various degrees of freedom. Moreover, by examining the kurtosis of the macroeconomic data set commonly used for diffusion index forecast (for example, [10]), [6] finds that most of these variables have heavier tails than the t-distribution with five degrees of freedom. Here, we use the t-distribution to fit the innovations u and e and obtain the estimated degree of freedom.

Table 4. Estimated distribution of innovations for full sample period.

Series	Variable	Innovations u	Innovations e
Panel A: 1927/01-2016/12			
CRSP value-weighted returns	Dividend-price ratio	$t(4)$	$t(3)$
	Earning-price ratio	$t(4)$	$t(3)$
	Dividend-payout ratio	$t(4)$	$t(2)$
	Long-term yield	$t(4)$	$t(1)$
	Default yield spread	$t(4)$	$t(1)$
Panel B: 1934/01-2016/12			
CRSP value-weighted returns	3-month T-bill	$t(5)$	$t(2)$
	Term spread	$t(5)$	$t(2)$

This table reports the estimated distribution of the innovations u and e from the model (1). The degrees of freedom for the t distribution are rounded to the nearest integer.

As shown in Table 4, all the degrees of freedom obtained are smaller than 5, and some are even much smaller. Note that the second moment does not exist when the degree of freedom is less than or equal to 2. On the basis of our test and the data sets, although the second moment of the innovations mostly exists, there exist some extreme cases that the second moment does not exist. This observation supports the merit of our assumptions that the second moment may or may not exist. This also indicates that this empirical data set is actually better fitted by the heavy-tailed innovations. This finding highlights the applicability of our Theorems 2.1-2.3 that allows the heavy-tail innovations assumption and implies that the proposed method is more robust than the previous method based on the normal assumption of the innovations.

Now we are going to show that the empirical economic data set tends to show a near unit-root phenomenon. In the following tables, we use the variables mentioned above to illustrate our study. Table 5 reports the 95% confidence interval of the autoregressive root ρ (and the corresponding c) for dividend-price ratio, earning-price ratio, dividend-payout ratio, long-term yield, default yield spread, 3-month T-bill, and term spread. The fourth column of Table 5 shows that all series are highly persistent. We also report point estimators of δ in the second column of Table 5. For these variables of dividend-price ratio and earning-price ratio, the correlations for the valuation ratios are large. This finding is similar to [3]. As expected, while there is a small degree of freedom, we can see that all of the correlations are between 0 and -1. This observation supports the merit of our Theorem 2.4 that would be applied to the case of an undefined correlation.

Table 5. Estimates of model parameters.

Variables	δ	DF-GLS	95% CI:ρ	95% CI:c
Panel A: 1927/01-2016/12				
Dividend-price ratio	-0.869	-3.347	(0.971, 0.987)	(-31.773, -13.66)
Earning-price ratio	-0.844	-2.747	(0.979, 0.992)	(-23.032, -8.111)
Dividend-payout ratio	-0.131	-7.732	(0.948, 0.960)	(-55.944, -42.977)
Long-term yield	-0.046	-1.583	(0.994, 0.997)	(-6.738, -3.329)
Default yield spread	-0.407	-4.481	(0.954, 0.970)	(-49.193, -32.555)
Panel B: 1934/01-2016/12				
3-month T-bill	-0.051	-1.976	(0.990, 0.994)	(-10.393, -5.84)
Term spread	-0.012	-4.647	(0.952, 0.961)	(-47.787, -38.545)

This table reports estimates of the parameters for the predictive regression model. δ is the estimated correlation between the innovations to returns and the predictor variable. The last two columns are the 95% confidence intervals for the largest autoregressive root (ρ) and the corresponding local-to-unity parameter (c) for each of the predictor variables. The values are rounded off to the third decimal digit.

As discussed above, the normal assumption might not be suitable for real applications, which implies that the generalized Bonferroni Q-test is robust to the heavy-tailed situation. Here, we also construct valid confidence intervals for β_1 through our generalized Bonferroni Q-test to test the predictability of returns. Table 6 shows stronger evidence of the predictability of earnings-price ratio, long-term yield, 3-month T-bill, and term spread. The evidence for predictability with other predictor variables is weaker. And then, the empirical data set explicitly reveals the properties of non-stationary and heavy-tailed innovations.

Table 6. Tests of predictability.

Series	Variable	$\widehat{\beta}_1$	90% CI:β_1
Panel A: 1927/01-2016/12			
CRSP value-weighted returns	Dividend-price ratio	0.023	[-0.033, 0.114]
	Earning-price ratio	0.045	**[0.006, 0.099]**
	Dividend-payout ratio	-0.001	[-0.005, 0.004]
	Long-term yield	0.000	**[0.000, 0.001]**
	Default yield spread	-0.002	[-0.003, 0.001]
Panel B: 1934/01-2016/12			
CRSP value-weighted returns	3-month T-bill	0.000	**[0.000, 0.001]**
	Term spread	0.001	**[0.000, 0.002]**

This table reports statistics used to infer the predictability of returns. The third and fourth columns report the point estimate $\widehat{\beta}_1$ from an OLS regression of returns onto the predictor variable and the 90% Bonferroni confidence intervals for β_1 using the generalized Q-test. The values are rounded off to the third decimal digit.

The above-reported results are based on the full sample period, from January 1927 to December 2016. However, to test the predictability, it is also common to test subperiods in the literature, examining whether the full sample period carries through the whole information (see [3] and [8] for recent examples). Here, we follow [8], splitting the sample period into two sub-periods: the pre-1994 period (January 1927 to December 1994) and the post-1952 period (January 1952 to December 2016).

Table 7. Tests of predictability in subperiods.

Time period	Variable	$\widehat{\beta}_1$	90% CI:β_1
Panel A: pre-1994			
1927/01–1994/12	Dividend-price ratio	0.045	**[0.006, 0.226]**
	Earning-price ratio	0.055	**[0.018, 0.139]**
	Dividend-payout ratio	-0.012	**[-0.026, 0.000]**
	Long-term yield	0.000	**[0.000, 0.001]**
	Default yield spread	-0.001	[-0.003, 0.002]
1934/01–1994/12	3-month T-bill	0.000	**[0.000, 0.001]**
	Term spread	0.001	**[0.000, 0.003]**
Panel B: post-1952			
1952/01–2016/12	Dividend-price ratio	0.089	[-0.084, 0.044]
	Earning-price ratio	0.041	[-0.024, 0.054]
	Dividend-payout ratio	0.002	**[0.000 0.008]**
	Long-term yield	0.000	[-0.001, 0.001]
	Default yield spread	0.001	[-0.001, 0.004]
	3-month T-bill	0.000	**[0.000, 0.001]**
	Term spread	0.001	**[0.000, 0.002]**

This table reports statistics used to infer the predictability of returns. The third and fourth columns report the point estimate $\widehat{\beta}_1$ from an OLS regression of returns onto the predictor variable and the 90% Bonferroni confidence intervals for β_1 using the generalized Q-test. The values are rounded off to the third decimal digit.

From Table 7, we have the predictability with dividend-price ratio, earnings-price ratio, dividend-payout ratio, long-term yield, 3-month T-bill, and term spread in the pre-1994 period, while only dividend-payout ratio, 3-month T-bill, and term spread have predictability in the post-1952 period. During the pre-1994 period, we found much more significant evidence in favor of predictability compared with the full sample period. Examining the post-1952 period, the predictability evidence almost disappears. Our study is similar to much of the literature: the predictability evidence is generally much weaker post-1952, analyzed for the predictability with the same splitting period (see examples in [3] and [8]). To explain this phenomenon, many papers attempt to determine whether the variables change over time, for example, cf. [9].

§6 Conclusion

In this paper, we study the issue of predictability under a predictive regression model, which allows the regressors to be a non-stationary process and the innovations to be correlated and heavy-tailed. We prove the asymptotic properties of the least squares estimators under this model, which allows us to provide a robust test based on the celebrated Bonferroni Q-test in [3]. Furthermore, our simulation results confirm that the proposed generalized Bonferroni Q-test is robust. We also show that this model can be used to explain the predictability phenomenon observed in empirical studies. That is, we consider the monthly CRSP value-weighted log excess returns with the predictors being the dividend-price ratio, earnings-price ratio, dividend-payout ratio, long-term yield, default yield spread, 3-month T-Bill and the term-spread. Our empirical analysis shows that all these variables have high persistence, and the earnings-price ratio, long-term yield, 3-month T-bill and term spread have predictability in the selected time period. Although predictability is important, we want to find the properties of the heavy-tail. We then use the empirical data to ensure that in our predictive regression model, the assumption of normal distribution innovations is not enough to describe the real economic data. Our tables and figures show that the innovations are heavy-tailed and our theorems are truly needed.

In summary, we show that the generalized Bonferroni Q-test is robust with suitable modifications, both for light-tailed and heavy-tailed innovations. We expect this method to be useful in other cases as well, and to find applications in the study of financial econometrics.

We investigate the relationships between heavy-tail, unit root, and predictability. To this end, we conduct a finite sample performance of the proposed testing statistic by means of an extensive Monte Carlo study. On the basis of our robustness test, we have three conclusions about the aforementioned relationships: (i) The heavy-tailed properties lead to decreased predictability. (ii) The heavy-tailed properties constitute stronger evidence of a unit root. (iii) The unit root situation leads to decreased predictability, which is also shown in Figure 4 of [3]. Regardless of whether heavy tails impact predictability directly or indirectly, these factors are bound together.

It would also be interesting to propose the generalized Bonferroni Q-test for multiple predictors. A valid confidence interval for the high-dimensional predictive regression using the generalized Bonferroni Q-test is certainly important for this literature.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

[1] J Y Campbell, A W Lo, A C McKinley. *The Econometrics of Financial Markets*, Princeton University Press, Princeton, 1997.

[2] J Y Campbell, M Yogo. *Implementing the econometric methods in "Efficient tests of stock return predictability"*, 2005, https://api.semanticscholar.org/CorpusID:67799405.

[3] J Y Campbell, M Yogo. *Efficient tests of stock return predictability*, Journal of Financial Economics, 2006, 81: 27-60.

[4] Y Choi, S Jacewitz, J Y Park. *A reexamination of stock return predictability*, Journal of Econometrics, 2016, 192: 168-189.

[5] G Elliott, T J Rothenberg, J H Stock. *Efficient tests for an autoregressive unit root*, Econometrica, 1996, 64: 813-836.

[6] J Fan, Y Ke, Y Liao. *Augmented factor models with applications to validating market risk factors and forecasting bond risk premia*, Journal of Econometrics, 2021, 222: 269-294.

[7] W Feller. *An Introduction to Probability Theory and Its Applications*, 2nd ed John Wiley and Sons, New York, 1971.

[8] A Kostakis, T Magdalinos, M P Stamatogiannis. *Robust econometric inference for stock return predictability*, Review of Financial Studies, 2015, 28: 1506-1553.

[9] M Lettau, S V Nieuwerburgh. *Reconciling the return predictability evidence*, Review of Financial Studies, 2008, 21: 1607-1652.

[10] S Ludvigson, S Ng. *Macro factors in bond risk premia*, Review of Financial Studies, 2009, 22: 5027-5067.

[11] N G Mankiw, M Shapiro. *Do we reject too often? Small sample properties of tests of rational expectation models*, Economics Letters, 1986, 20: 139-145.

[12] T Mikosch, C Stărică. *Limit theory for the sample autocorrelations and extremes of a* $GARCH(1,1)$ *process*, The Annals of Statistics, 2000, 28: 1427-1451.

[13] T Mikosch, C G de Vries. *Heavy tails of OLS*, Journal of Econometrics, 2013, 172: 205-221.

[14] P C B Phillips. *Towards a unified asymptotic theory for autoregression*, Biometrika, 1987, 74: 535-547.

[15] P C B Phillips. *Time series regression with a unit root and infinite-variance errors*, Econometric Theory, 1990, 6: 44-62.

[16] P C B Phillips. *On confidence intervals for autoregressive roots and predictive regression*, Econometrica, 2014, 82(3): 1177-1195.

[17] R F Stambaugh. *Predictive regressions*, Journal of Financial Economics, 1999, 54: 375-421.

[1]Department of Statistics & Fintech and Green Finance Center, Taipei University, New Taipei City 23741, China.

Email: hcwong@gm.ntpu.edu.tw

[2]Graduate Institute of Statistics, Central University, Taoyuan County 32049, China.

Emails: jessie0720hao@gmail.com, cdffuh@gmail.com

[3]School of Mathematical Sciences, Zhejiang University, Hangzhou 310058, China.

Email: txpang@zju.edu.cn