Econometric Theory, **40**, 2024, 1364–1390.
doi:10.1017/S0266466623000117

NEW ROBUST INFERENCE FOR PREDICTIVE REGRESSIONS

RUSTAM IBRAGIMOV
Imperial College Business School and Center for Econometrics and Business Analytics, St. Petersburg University

JIHYUN KIM
Sungkyunkwan University and Toulouse School of Economics

ANTON SKROBOTOV
Russian Presidential Academy of National Economy and Public Administration and St. Petersburg University

We propose a robust inference method for predictive regression models under heterogeneously persistent volatility as well as endogeneity, persistence, or heavy-tailedness of regressors. This approach relies on two methodologies, nonlinear instrumental variable estimation and volatility correction, which are used to deal with the aforementioned characteristics of regressors and volatility, respectively. Our method is simple to implement and is applicable both in the case of continuous and discrete time models. According to our simulation study, the proposed method performs well compared with widely used alternative inference procedures in terms of its finite sample properties in various dependence and persistence settings observed in real-world financial and economic markets.

1. INTRODUCTION

Many papers in the literature have focused on econometric analysis of predictive regressions for stock returns (see Phillips, 2015, for an up-to-date review). Predictive regression data are known to have several problematic characteristics, especially in statistical inference of stock return predictability. First, it is widely believed that the popular regressors, such as dividend-price and earnings-price ratios, used in the predictive regressions have near unit roots and their innovations

We thank the Editor (Peter C.B. Phillips), the Co-Editor (Anna Mikusheva), and three anonymous referees for many helpful comments and suggestions. We are also grateful to Walter Distaso, Jean-Marie Dufour, Siyun He, Nour Meddahi, Mikkel Plagborg-Møller, Aleksey Min, Ulrich K. Müller, Rasmus S. Pedersen, Artem Prokhorov, Rogier Quaedvlieg, Robert Taylor, Alex Maynard, Neil Shephard, Yang Zu, and the participants at the Center for Econometrics and Business Analytics (CEBA, St. Petersburg University) and University of Nottingham seminar series, the session on Econometrics of Time Series at the 12th World Congress of the Econometric Society, and iCEBA-2021, 2022 conferences for helpful discussions and comments. Financial support from a British Academy/Leverhulme Small Research Grant (R. Ibragimov, reference SG2122\211040), the French Government and the ANR' Investissements d'Avenir program (J. Kim, Grant ANR-17-EURE-0010) and the Russian Science Foundation (A. Skrobotov, Project No. 20-78-10113) for various and non-overlapping parts of this research is gratefully acknowledged. Address correspondence to Jihyun Kim, School of Economics, Sungkyunkwan University, Seoul 03063, South Korea; e-mail: kim.jihyun@skku.edu.

are correlated with stock returns in the long run. The characteristics of the regressors, which are persistence and endogeneity, jointly cause standard hypothesis tests to become substantially biased (see Stambaugh, 1999). Second, there is some evidence that supporting volatility of stock returns is stochastic and highly persistent (see, e.g., Jacquier, Polson, and Rossi, 2004; Hansen and Lunde, 2014). Cavaliere (2004) shows that persistent stochastic volatility may cause substantial size distortions on standard tests developed mostly under the assumption that a volatility process is stationary with a constant unconditional mean, such as stationary GARCH-type models. Lastly, there are several other characteristics of predictive regression data that include heavy-tailedness of regressors as well as jumps, structural breaks, and regime switching in volatility. These characteristics may also yield jointly or individually a significant distortion of standard hypothesis tests for predictive regressions.

In this paper, we propose a new method for robust inference on parameters of predictive regression models under the aforementioned characteristics of predictive regression data. Our approach relies on a simple nonlinear instrumental variable (IV) estimation and a nonparametric volatility correction. The nonlinear IV estimator in our approach is an IV estimator with the instrument being the sign transformation of the regressor. This particular IV estimator was first proposed by Cauchy (1836), and is called the *Cauchy estimator*. As is known in the literature, the use of the instrument can effectively eliminate the problems caused by the persistent endogeneity, heavy-tailedness, and other problematic characteristics of the regressors (see So and Shin, 1999; Choi, Jacewitz, and Park, 2016; Kim and Meddahi, 2020). On the other hand, volatility correction is used to deal with the problems caused by the presence of heterogeneity and persistence in stock return's volatility. As for the volatility correction, we consider a standard kernel-based nonparametric estimator of volatility.

Many authors have studied the issue of persistent endogeneity of regressors in predictive regressions. Among many of them, Campbell and Yogo (2006), Chen and Deo (2009), and Phillips and Magdalinos (2009) have proposed tests of return predictability, which are aimed at dealing with persistent and endogenous regressors. Though their tests perform well under the presence of persistent endogeneity of regressors, they are not expected to deal with other problematic characteristics of predictive regression data effectively. Our simulation study shows that they have serious size distortions under the null of no predictability when volatility is persistent or incorporates structural breaks or regime switching. In contrast, the robustness of our approach is quite evident. Our approach always yields almost exact sizes in a variety of designs considered in our simulation study. Moreover, the robustness of our approach is obtained with no significant loss of power. The discriminatory powers of our test are comparable to the tests by Campbell and Yogo (2006) and Chen and Deo (2009), which are optimal for the basic Gaussian model.

Our work is closely related to Choi et al. (2016), who propose an inference approach for predictive regressions. Similar to our method, their approach relies on the Cauchy estimator to eliminate the problems caused by the problematic characteristics of the regressors. They also use a nonparametric volatility correction. However, their approach for volatility correction is quite different from ours, and its applicability is limited to a predictive regression equipped with appropriate high-frequency data since their method and theory are developed in a continuous time framework. More precisely, their volatility correction, called the *time change*, requires uniformly consistent estimation of a quadratic variation of a stock price for which high-frequency observations of the stock price are necessary. Consequently, their approach requires the assumption that the sampling interval decreases to zero, and applications of their method on relatively low-frequency data, i.e., monthly or quarterly data, are largely restricted. However, predictive regressions are often estimated using monthly or quarterly data. In contrast, our method can be applied to a discrete time model and a discrete sample collected from an underlying continuous time model as in Choi et al. (2016). Our simulation study shows that both our method and the method by Choi et al. (2016) perform well and have good size and power performances under continuous time settings considered in this paper. However, unlike our method, the Choi et al. (2016) method is not applicable under discrete time settings. Therefore, we may say that our method is more flexible and widely applicable since it can be applied to both high- and low-frequency data.

The rest of the paper is organized as follows: Section 2 introduces the predictive regression models, persistent volatility, and the Cauchy estimator. Section 3 proposes the robust inference method and presents its asymptotic properties. Section 4 generalizes the baseline predictive regression models, which have one persistent volatility factor, to have a two-factor volatility, where one factor is persistent and the other is transient such as a stationary GARCH process. Section 5 provides numerical results on finite sample properties of the proposed robust inference approach. Section 6 makes some concluding remarks.

The Supplementary Material provides a discussion of the Cauchy estimator and general nonlinear IV estimators with the relevant asymptotic results that, in particular, point to the importance and usefulness of the Cauchy estimator (Appendix A), useful auxiliary results (Appendix B), the proofs of the main results in the paper (Appendix C), and some additional simulation results on finite sample performance of inference approaches dealt with (Appendix D).

2. PREDICTIVE REGRESSIONS

2.1. Research Problems and Models

Throughout the paper, we consider $(\mathcal{F}_t)$-adapted processes defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ equipped with an increasing filtration $(\mathcal{F}_t)$ of sub-σ-fields of $\mathcal{F}$. We consider a test for no predictability of the process (y_t) (e.g., the time series of excess stock returns) based on some covariate process (x_t) (e.g., the time series of price-to-dividend ratios). We consider the linear predictive regression model

$$y_t = \alpha + \beta x_{t-1} + u_t, \ \ t = 1, ..., T, \tag{1}$$

where (u_t) is a martingale difference sequence (MDS) with respect to $(\mathcal{F}_t)$. In particular, (u_t) is conditionally heteroskedastistic. Following the usual specification for a volatility model, we assume that

$$u_t = v_t \varepsilon_t,$$

where (v_t) is a volatility process and (ε_t) is an MDS with respect to $(\mathcal{F}_t)$.

Assumption 2.1. (a) (v_t) is $(\mathcal{F}_{t-1})$-adapted and is defined on $[\underline{v}, \bar{v}]$ for some $0 < \underline{v} < \bar{v} < \infty$, (b) $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = 1$, and (c) $\sup_{t \geq 1} E(\varepsilon_t^4 | \mathcal{F}_{t-1}) < \infty$.

The conditions (a) and (b) in Assumption 2.1 are not stringent, and are required for the identification of the conditional variance of u_t. In particular, the conditional variance of u_t given $\mathcal{F}_{t-1}$ is well identified, and we have $E(u_t^2 | \mathcal{F}_{t-1}) = v_t^2$. Our test relies on uniform convergence results for a nonparametric estimator of the volatility process (v_t). The condition (c) is used to obtain a uniform convergence rate of the nonparametric estimator of the volatility process (v_t). Note that Assumption 2.1 implies $\sup_{t \geq 1} E(u_t^4) < \infty$, and hence it rules out a predictive regression model having a heavy-tailed regression error (u_t).

As for a nontrivial example, we let $v_t = f(z_{t-1})$ and $\varepsilon_t \sim iid\mathbb{N}(0,1)$, where f is a positive function and z_t is an $\mathcal{F}_t$-adapted process. Then (v_t, ε_t) satisfies Assumption 2.1. If we assume that f is bounded above, then $u_t = v_t \varepsilon_t$ satisfies $\sup_{t \leq 1} E(|u_t|^4 | \mathcal{F}_{t-1}) < \infty$ a.s. for any $\mathcal{F}_t$-adapted process (z_t) since $\varepsilon_t \sim iid\mathbb{N}(0,1)$. Moreover, u_t is not uniformly bounded, i.e., there does not exist M such that $|u_t| \leq M < \infty$ with probability one even if f is bounded above, since the standard normal random variable ε_t is not uniformly bounded. Further examples of martingales with bounded conditional moments of MDS summands are provided by more general martingale transforms and randomly stopped sums of independent r.v.'s (see Remark 3.3 in de la Peña, Ibragimov, and Sharakhmetov, 2003).

The hypothesis of no predictability of (y_t) corresponds to the hypothesis $\beta = 0$ in predictive model (1). It is well known that the standard ordinary least squares (OLS)-based t-test is not robust with respect to a wide range of statistical problems in predictive regression data. For instance, the standard OLS estimator of β is not asymptotically Gaussian under $H_0 : \beta = 0$ if (x_t) is endogenous and (nearly) nonstationary (see Elliott and Stock, 1994; Phillips, 1987b; Phillips and Magdalinos, 2007) or is stationary with infinite second moment (e.g., Granger and Orr, 1972; Embrechts, Klüppelberg, and Mikosch, 1997; Ibragimov, Ibragimov, and Walden, 2015, and the references therein), even when there is no heteroskedasticity and $v_t = \sigma$ is constant.[1]

In the case of predictive regressions for stock returns, the returns process (y_t) is widely believed to have time-varying stochastic volatility (see Choi et al., 2016 and

[1] The endogeneity of the covariate x_t refers to the existence of nonzero long-run covariance between innovations of u_t and x_t.

the references therein). Moreover, the volatility process is typically very persistent. For example, many authors have found that the autoregressive parameter for the dynamics of the volatility process is close to one under some appropriate functional transformations. In particular, Jacquier et al. (2004) and Hansen and Lunde (2014) provide convincing evidence that the logarithm of the volatility process follows a near unit-root process for a wide range of equity and foreign exchange rate time series. It is well known that the presence of persistent volatility may cause the distribution of the standard t-statistic to be far from standard normal, yielding a substantial distortion in testing relying on standard normal critical values (see, e.g., Chung and Park, 2007; Choi et al., 2016; Kim and Park, 2017).

2.2. The Cauchy Estimator

Our inference method is based on the Cauchy estimator. To effectively explain the main idea, we consider model (1) with no intercept term, i.e., $\alpha = 0$, and introduce the Cauchy estimator $\check{\beta}$ for β, which is given by

$$\check{\beta} = \left(\sum_{t=1}^{T} |x_{t-1}| \right)^{-1} \sum_{t=1}^{T} sign(x_{t-1}) y_t,$$

where $sign(\cdot)$ is the sign function defined as $sign(x) = 1$ for $x \geq 0$, and $sign(x) = -1$ for $x < 0$. Thus, $\check{\beta}$ is an IV estimator with the instrument $sign(x_{t-1})$. This particular IV estimator was first proposed by Cauchy (1836). See, among others, So and Shin (1999), Phillips, Park, and Chang (2004), Choi et al. (2016), and Kim and Meddahi (2020) for econometric applications of the Cauchy estimator.

Under Assumption 2.1(b), not only ε_t, but also $sign(x_{t-1})\varepsilon_t$, hereafter denoted by ξ_t, is an MDS with respect to the filtration $(\mathcal{F}_t)$ with $E(\xi_t^2|\mathcal{F}_{t-1}) = 1$. Let us define a continuous time partial sum process $(W_T(r), 0 \leq r \leq 1)$ by

$$W_T(r) = \frac{1}{T^{1/2}} \sum_{t=1}^{[Tr]} \xi_t. \tag{2}$$

The stochastic process $(W_T(r))$ takes values in $\mathbf{D}_{\mathbb{R}}[0,1]$, where $\mathbf{D}_E[0,1]$ denotes the space of càdlàg functions from $[0,1]$ to $E \subset \mathbb{R}^d$ for some positive integer d. Under Assumption 2.1(b) and (c), the partial sum process $(W_T(r))$ follows the usual functional central limit theorem (CLT) for martingales (see, e.g., Billingsley, 1986, Thm. 18.2), that is,

$$W_T \to_d W$$

in $\mathbf{D}_{\mathbb{R}}[0,1]$, where W is a standard Brownian motion. The convergence $W_T \to_d W$ is to be interpreted as the weak convergence in the probability measures on $\mathbf{D}_{\mathbb{R}}[0,1]$. In our context, it is more convenient, and so is assumed, to endow $\mathbf{D}_E[0,1]$ with the uniform topology rather than the usual Skorohod topology (see Billingsley, 1986, pp. 150–152).

The use of the Cauchy estimator in our inference method is motivated by the above functional CLT for $(W_T(r))$. To convey the main idea, assume that the volatility process (v_t) is observable. Recall that the numerator of $\check{\beta}$ is $\sum_{t=1}^{T} sign(x_{t-1})y_t$, and it becomes $\sum_{t=1}^{T} v_t\xi_t$ under $\beta = 0$, where $\xi_t = sign(x_{t-1})\varepsilon_t$. One then may construct a robust test for the null hypothesis $H_0 : \beta = 0$ against the alternative $H_1 : \beta \neq 0$ using the following statistic:

$$\tau(v) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} sign(x_{t-1}) \frac{y_t}{v_t}. \tag{3}$$

In particular, for $\beta = 0$,

$$\tau(v) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} \xi_t = W_T(1) \to_d W(1) = \mathbb{N}(0,1).$$

In practice, however, the volatility process (v_t) is not observable, and hence the above inference procedure using $\tau(v)$ is not feasible. In Section 3, a feasible version of the Cauchy-based inference method above will be fully addressed under our construction of the persistent volatility introduced in Section 2.3.

2.3. Persistent Volatility

This subsection presents a time-varying and persistent volatility model, which is a well-known stylized fact for many financial returns. We define a stochastic process σ_T on $\mathbf{D}_{\mathbb{R}+}[0,1]$ as $\sigma_T(r) = v_{[Tr]}$. We assume that σ_T has a limiting process σ defined over $0 \leq r \leq 1$ such that (W_T, σ_T) converges to (W, σ) jointly, where (W_T) is defined as in (2). Specifically, we consider the following assumption.

Assumption 2.2. There exists a positive process σ on $\mathbf{D}_{\mathbb{R}+}[0,1]$ such that

$$(W_T, \sigma_T) \to_d (W, \sigma)$$

in $\mathbf{D}_{\mathbb{R}\times\mathbb{R}+}[0,1]$, where W is a standard Brownian motion with respect to the filtration to which W and σ are adapted.

The above assumptions hold for wide classes of models, such as models with nonstationary volatility, regime switching, and structural breaks in volatility. It also holds for the processes with $v_t = \sigma(t/T)$, where $\sigma(s)$ is a deterministic function on $[0,1]$, considered by Cavaliere and Taylor (2007), Xu and Phillips (2008), and Harvey, Leybourne, and Zu (2019), among others.[2] The assumptions also hold for processes with nonstationary volatilities considered by Hansen (1995) and Chung and Park (2007), who assume that v_t^2 is a smooth positive transformation of a (near) unit-root process, i.e., $v_t^2 = \sigma^2(T^{-1/2}z_{t-1})$ for a unit-root process z_t.

[2] Assumption 2.2 is a simplified version of the condition $v_{[Tr]}/a_T \to_d \sigma_r$ considered by Assumption 2 of Cavaliere and Taylor (2009). We rule out the explosive volatility settings with $a_T \to \infty$, and consider the stable volatility processes with $a_T = 1$ for simplicity. The results in the paper can be obtained under the explosive volatility assumption with $a_T \to \infty$ at the cost of a more involved analysis.

One should note that Assumption 2.2 is more general than the volatility models considered in the aforementioned literature and, in particular, allows the volatility to be stochastically discontinuous, which are desirable properties for modeling financial volatility having structural breaks or regime switching.

Assumptions 2.1 and 2.2 rule out some cases of globally homoskedastic processes, such as stationary GARCH processes. In Section 4, we generalize the model to have a two-factor volatility, one for a nonstationary long-run component and the other one for a stationary short-run component, and show the validity of our robust method introduced in Section 3 for the generalized model with the two-factor volatility.

Under our construction of the persistent volatility, the asymptotic behavior of the Cauchy estimator can be obtained immediately. The asymptotics of the Cauchy estimator $\check{\beta}$ are mainly determined by $\sum_{t=1}^{T} v_t \xi_t$ since $\check{\beta} = \beta + \sum_{t=1}^{T} v_t \xi_t / \sum_{t=1}^{T} |x_{t-1}|$. Note that $T^{-1/2} \sum_{t=1}^{[Tr]} v_t \xi_t = \int_0^r \sigma_T(s) dW_T(s)$ for $r \in [0,1]$, and the weak convergence of the stochastic integral $\int \sigma_T(r) dW_T(r)$ is well documented in the literature (see, e.g., Hansen, 1992, Thm. 2.1; Kurtz and Protter, 1991, Thm. 4.6), and we have $(\int \sigma_T(r) dW_T(r)) \to_d (\int \sigma(r) dW(r))$.

LEMMA 2.1. *Under Assumptions 2.1 and 2.2,*

$$\left(\sum_{t=1}^{T} |x_{t-1}| / \sqrt{T} \right) \left(\check{\beta} - \beta \right) \to_d \int_0^1 \sigma(r) dW(r).$$

Two main implications of Lemma 2.1 are (i) the limit distribution of the Cauchy estimator is generally non-Gaussian and (ii) the rate of convergence of the Cauchy estimator is nonstandard and unknown. These asymptotic properties of the Cauchy estimator subsequently imply that the usual t-test relying on the standard normal table becomes an invalid testing procedure for the null hypothesis of $\beta = 0$. The limit $\int_0^1 \sigma(r) dW(r)$ is Gaussian if and only if the limiting volatility process σ is independent of W. In this case, $\int \sigma(r) dW(r)$ has a mixed normal distribution and $\int_0^s \sigma(r) dW(r) =_d \mathbb{MN}(0, \int_0^s \sigma^2(r) dr)$. If the independence condition is violated, then $\int \sigma(r) dW(r)$ becomes a non-Gaussian martingale in general.

Clearly, $\check{\beta}$ requires an extremely mild condition for consistency, that is, $\sum_{t=1}^{T} |x_{t-1}| / \sqrt{T} \to_p \infty$. For example, if there exists a sequence p_T of positive numbers such that

$$\left(p_T^{-1} \sum_{t=1}^{T} |x_{t-1}| \right)^{-1} = O_p(1),$$

then $\check{\beta} - \beta = O_p(T^{1/2}/p_T)$ by Lemma 2.1. For a wide class of time series, the consistency condition $T^{1/2}/p_T \to 0$ is satisfied since $p_T \geq T$ unless $x_t \approx 0$ for most $t = 1, \ldots, T$. Though it is not necessary in our subsequent theory, one may explicitly obtain the sequence p_T for some time series satisfying required regularity conditions.

Example 2.1. (a) For weakly stationary processes (x_t) with $E|x_t| < \infty, p_T = T$.

(b) For stationary α-stable (x_t) with $0 < \alpha < 1$, $p_T = T^{1/\alpha}\ell(T)$ for some slowly varying function ℓ (see Embrechts et al., 1997; Phillips and Solo, 1992; and the references therein).

(c) For the case of unit-root and near unit-root time series (x_t), $p_T = T^{3/2}$ (see Phillips, 1987a, 1987b; Phillips and Magdalinos, 2007; Ibragimov and Phillips, 2008, and the references therein).

(d) For fractionally integrated $I(d)$ processes (x_t) with $\frac{1}{2} < d < \frac{3}{2}, p_T = T^{d+1/2}\ell(T)$ for some slowly varying function ℓ (see Baillie, 1996; Phillips, 1999, Lem. 3.4; Kim and Phillips, 2006; Wang, Lin, and Gulati, 2003; Chan and Wang, 2015 and the references therein).

3. NEW ROBUST INFERENCE APPROACH

Now we introduce our test for no predictability in the regression (1). The test is motivated by $\tau(v)$ in (3). Since (v_t) is not observable, we replace v_t by its consistent estimator $\hat{\sigma}((t-1)/T)$, and we consider the test statistic $\tau(\hat{\sigma})$ defined as

$$\tau(\hat{\sigma}) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} sign(x_{t-1}) \frac{y_t}{\hat{\sigma}((t-1)/T)}, \tag{4}$$

where

$$\hat{\sigma}^2(r) = \frac{\sum_{t=1}^{T} \hat{u}_t^2 K_h(r - t/T)}{\sum_{t=1}^{T} K_h(r - t/T)}, \quad h \le r \le 1; \qquad \hat{\sigma}^2(r) = \hat{\sigma}^2(h), \quad 0 \le r < h, \tag{5}$$

where $\hat{u}_t$ are the OLS residuals given as $\hat{u}_t = y_t - \hat{\beta} x_{t-1}$ with the OLS estimator $\hat{\beta}$. Here, $K_h(t) = K(t/h)$ with a kernel function K and bandwidth h.

The validity of our approach requires that $\hat{\sigma}(r)$ is close enough to $\sigma_T(r) = v_{[Tr]}$ for most $r \in [0,1]$. We first establish a uniform convergence result

$$\sup_{r \in \mathcal{C}_h} \left| \hat{\sigma}^2(r) - \sigma_T^2(r) \right| = o_p(1) \tag{6}$$

for some $\mathcal{C}_h \subset [0,1]$. Invoking the convergence in Assumption 2.2, $\sigma_T \to_d \sigma$ is interpreted as the weak convergence in the probability measures on $\mathbf{D}_{\mathbb{R}+}[0,1]$ endowed with the uniform topology. By virtue of the so-called Skorohod representation theorem (e.g., Pollard, 1984, pp. 71–72), it is indeed possible to construct σ_T and σ on a common probability space, up to the distributional equivalence, so that $\sigma_T \to_{a.s.} \sigma$ uniformly on $[0,1]$. For our development of the uniform convergence results (6), we assume that σ_T is defined up to the distributional equivalence such that $\sigma_T \to_{a.s.} \sigma$ uniformly on $[0,1]$. This assumption is not restrictive since we are interested in the convergence of $\hat{\sigma}^2$ to σ_T^2 rather than σ^2.

For the nonparametric estimator $\hat{\sigma}$, we assume that the kernel function K satisfies the following assumption.

Assumption 3.1. (a) A nonnegative kernel K has a compact support $[0,1]$ with $\int_0^1 K(r)dr = 1$, (b) $|K(r)-K(s)| \le \bar{K}|r-s|$ for all $r,s \in \mathbb{R}$, and $\sup_r K(r) < \bar{K}$ for some $0 < \bar{K} < \infty$.

The condition (b) in Assumption 3.1 is standard in the investigation of uniform consistency. In Assumption 3.1(a), we assume a nonstandard assumption that K is a one-sided kernel, which is unnecessary in developing the uniform consistency (6). When we establish $\tau(\hat{\sigma}) \to_d \mathbb{N}(0,1)$ under $\beta = 0$, however, it is important to make $\hat{\sigma}(t/T)$ measurable with respect to $\mathcal{F}_{t+1}$ so that we can apply a martingale CLT to $\tau(\hat{\sigma})$.[3] For a more precise explanation, we write

$$\hat{\sigma}^2(r) = \hat{\sigma}_1^2(r) + \hat{\sigma}_2^2(r) + \hat{\sigma}_3^2(r) + \hat{\sigma}_4^2(r), \tag{7}$$

where

$$\hat{\sigma}_1^2(r) = \frac{\sum_{t=1}^T E(u_t^2|\mathcal{F}_{t-1})K_h(r-t/T)}{\sum_{t=1}^T K_h(r-t/T)}, \qquad \hat{\sigma}_2^2(r) = \frac{\sum_{t=1}^T (u_t^2 - E(u_t^2|\mathcal{F}_{t-1}))K_h(r-t/T)}{\sum_{t=1}^T K_h(r-t/T)},$$

$$\hat{\sigma}_3^2(r) = (\hat{\beta}-\beta)^2 \frac{\sum_{t=1}^T x_{t-1}^2 K_h(r-t/T)}{\sum_{t=1}^T K_h(r-t/T)}, \quad \hat{\sigma}_4^2(r) = 2(\hat{\beta}-\beta)\frac{\sum_{t=1}^T x_{t-1}u_t K_h(r-t/T)}{\sum_{t=1}^T K_h(r-t/T)}.$$

In the decomposition (7), $\hat{\sigma}_1^2(r) - \sigma_T^2(r)$ is a bias term since $E(u_t^2|\mathcal{F}_{t-1}) = v_t^2 = \sigma_T^2(t/T)$, whereas $\hat{\sigma}_2^2(r)$ is a variance term involving a martingale. On the other hand, $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$ are error components induced by using $\hat{u}_t$, instead of u_t, in the kernel estimation of σ_T^2. Under Assumption 3.1(a), $\hat{\sigma}_1^2(t/T)$ is $\mathcal{F}_{t-1}$-adapted, whereas $\hat{\sigma}_2^2(t/T)$ is $\mathcal{F}_t$-adapted. Consequently, $\tilde{\sigma}^2(t/T)$, where $\tilde{\sigma}^2 = \hat{\sigma}_1^2 + \hat{\sigma}_2^2$, is $\mathcal{F}_t$-adapted, from which one may show that $\tau(\tilde{\sigma}) \to_d \mathbb{N}(0,1)$, where $\tau(\tilde{\sigma})$ is defined as $\tau(\hat{\sigma})$ but with $\hat{\sigma}$ replaced $\tilde{\sigma}$, by a martingale CLT as long as $|\tilde{\sigma}^2(r) - \sigma_T^2(r)| = o_p(1)$ for most $r \in [0,1]$. However, $\hat{\sigma}_3^2(t/T)$ and $\hat{\sigma}_4^2(t/T)$ are not $\mathcal{F}_t$-measurable since $\hat{\beta}$ is not $\mathcal{F}_t$-measurable for any $t < T$. Consequently, we cannot directly apply a martingale CLT to show $\tau(\hat{\sigma}) \to_d \mathbb{N}(0,1)$. Alternatively, for these two terms, it is shown that they have negligible effects in the test statistic $\tau(\hat{\sigma})$, and we have $\tau(\hat{\sigma}) = \tau(\tilde{\sigma})(1+o_p(1))$ as long as $\hat{\beta} \to_p \beta$ sufficiently quickly. For the asymptotic negligibilities of $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$, we assume the following.

Assumption 3.2. For any deterministic sequence $(c_t)_{t=1}^T$ such that $0 \le c_t \le 1$ for all t, $\sum_{t=1}^T c_t x_{t-1} u_t = O_p\left(T^p \left(\sum_{t=1}^T x_{t-1}^2\right)^{1/2}\right)$, for some $p \in [0, 1/8)$.

Assumption 3.2 is very general, and many time series models satisfy the condition. In particular, it holds with $p = 0$ if (x_t) is either (near) unit root or stationary with finite variance. Moreover, if (x_t) is stationary with unbounded variance, then the condition holds with $p = 0$ under some additional conditions on (x_t) and (u_t) (see, e.g., Samorodnitsky et al., 2007).

[3]For the same reason, Hansen (1995) considered a one-sided kernel.

LEMMA 3.1. *If Assumptions 2.1 and 3.2 hold, then* $|\hat{\beta}-\beta| = O_p\left(T^p\left(\sum_{t=1}^T x_{t-1}^2\right)^{-1/2}\right)$.

We will show below that the rate of convergence of $\hat{\beta}$ in Lemma 3.1 is enough to obtain the required uniform convergences of $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$ as well as their asymptotic negligibility in the test relying on the statistic (4).

On the other hand, the convergence $|\hat{\sigma}_1^2(r)-\sigma_T^2(r)| \to_p 0$ requires that σ_T^2 be left-continuous at r due, in particular, to the fact that K is a one-sided kernel having support $[0,1]$. However, σ_T may have countably many jumps since $\sigma_T \in \mathbf{D}[0,1]$. In particular, at a discontinuity point r with $\sigma(r) \neq \sigma(r-)$, we have $|\hat{\sigma}_1^2(r)-\sigma_T^2(r-)| \to_p 0$ instead of $|\hat{\sigma}_1^2(r)-\sigma_T^2(r)| \to_p 0$. Therefore, the set $\mathcal{C}_h$ in (6) should effectively exclude a set of discontinuity points as well as its neighborhoods so that the uniform convergence result holds. Under our convention of $\sigma_T \to_{a.s.} \sigma$ uniformly on $[0,1]$, we only need to consider σ's discontinuity points, and we define

$$\mathcal{C}_h = [h,1]\setminus \mathcal{J}_h, \quad \text{where} \quad \mathcal{J}_h = \{[r,r+h) \subset [0,1] | \sigma(r) \neq \sigma(r-)\}. \tag{8}$$

Clearly, $\mathcal{C}_h$ is a set of left-continuity points, and we establish the uniform convergence result (6) over $\mathcal{C}_h$.

A martingale exponential inequality can be used to show the asymptotic negligibility of the variance component $\hat{\sigma}_2^2(r)$ uniformly in r (see, e.g., de la Peña, 1999; Bercu and Touati, 2008). In this paper, we use the two-sided exponential inequality in Bercu and Touati (2008) under which we can relax the moment condition for (ε_t) at the cost of an assumption on the stochastic order of the extremal process of (ε_t).

Assumption 3.3. For some $q \in [0,1/8)$, $\max_{1\le t\le T}|\varepsilon_t| = O_p(T^q)$.

Assumption 3.3 is not stringent, and a wide class of time series models for ε_t satisfies the condition.[4] For instance, if ε_t is a Gaussian process with $cov(\varepsilon_1,\varepsilon_T)\log T \to 0$, then $\max_{0\le t\le T}|\varepsilon_t| = O_p(\sqrt{\log T})$ and the condition (a) holds for any $q>0$ (see, e.g., Leadbetter and Rootzén, 1988, Thm. 2.5.2).

Assumption 3.4. As $h \to 0$ and $T\to\infty$, (a) $hT^{1/2-2p} \to \infty$ where $p \in [0,1/8)$ is defined as in Assumption 3.2, and (b) $hT^{1-4q}\to\infty$ and $hT^{2q}\to 0$, where $q \in [0,1/8)$ is defined as in Assumption 3.3.

Assumption 3.4 provides the connections among the stochastic orders in Assumptions 3.2 and 3.3 and the bandwidth h. If we let $h = cT^{-\alpha}$ for $c,\alpha>0$ as in the typical situation, then Assumption 3.4 holds for $2q<\alpha<\min\{1/2-2p, 1-4q\}$. Note that such α always exists for $p,q \in [0,1/8)$. In particular, if

[4] Instead of Assumption 3.3, one may obtain the subsequent results by assuming an additional moment condition, i.e., $E|\varepsilon_t|^{4r}<\infty$, for some $r\ge 1$. For a relevant approach, the reader is referred to, e.g., Theorem 2.1 of Wang and Chan (2014).

$p = q = 0$, then $h = cT^{-\alpha}$ satisfies Assumption 3.4 for $0 < \alpha < 1/2$. We note that the condition (a) is used to guarantee $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$ being asymptotically negligible in our inference method. In condition (b), $hT^{1-4q} \to \infty$ is needed for the uniform convergence of $\hat{\sigma}_2^2$, whereas $hT^{2q} \to 0$ is used to effectively handle discontinuity points of σ^2 at which $\hat{\sigma}^2$ becomes inconsistent.

PROPOSITION 3.2. *Let Assumptions 2.1, 2.2, and 3.1–3.4 hold. As $h \to 0$ and $T \to \infty$, we have*

$$(a) \quad \sup_{r \in \mathcal{C}_h} |\hat{\sigma}_1^2(r) - \sigma_T^2(r)| = o_p(1), \qquad (b) \quad \sup_{h \le r \le 1} |\hat{\sigma}_2^2(r)| = O_p\left(T^{2q}\left(\log(hT)/(hT)\right)^{1/2}\right),$$

$$(c) \quad \sup_{h \le r \le 1} |\hat{\sigma}_3^2(r)| = O_p\left(T^{2p}/(hT)\right), \qquad (d) \quad \sup_{h \le r \le 1} |\hat{\sigma}_4^2(r)| = O_p\left(T^{2p}/(hT)\right),$$

and the uniform convergence result (6) holds.

Under Assumption 3.4, we have

$$T^{2p}/(hT) = o\left(T^{2q}\left(\log(hT)/(hT)\right)^{1/2}\right),$$

from which we can see that the error components $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$ have smaller orders than $\hat{\sigma}_2^2$. Indeed, it is shown in the proof of Theorem 3.3 that $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$ have negligible effects in the test statistic $\tau(\hat{\sigma})$, and we have $\tau(\hat{\sigma}) = \tau(\tilde{\sigma})(1 + o_p(1))$, where $\tilde{\sigma}^2 = \hat{\sigma}_1^2 + \hat{\sigma}_2^2$. However, the orders of $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are not sufficiently small to show directly that $\tau(\tilde{\sigma}) = \tau(\sigma_T)(1 + o_p(1))$ even though $\tilde{\sigma}^2$ converges uniformly to σ_T^2. That is because the convergence rate of $\hat{\sigma}_2^2(r) \to_p 0$ is not fast enough for the direct approximation $\tau(\tilde{\sigma}) = \tau(\sigma_T)(1 + o_p(1))$, and the convergence rate of $|\hat{\sigma}_1^2(r) - \sigma_T^2(r)| \to 0$ depends on the degrees of left continuity of σ_T^2 which are unknown in general.

Alternatively, we use the weak convergence of the stochastic integral $\int_0^1 (\sigma_T(r)/\tilde{\sigma}(r))dW_T(r)$, as in Lemma S.8, jointly with the facts that $\tilde{\sigma}(t/T)$ is $\mathcal{F}_t$-adapted and $\sup_{r \in \mathcal{C}_h} |\tilde{\sigma}^2(r) - \sigma_T^2(r)| = o_p(1)$. Here, in particular, we require $\mathcal{C}_h \to_{a.s.} [0,1]$, which holds when σ has finitely many jumps almost surely.

Assumption 3.5. σ has finitely many jumps almost surely.[5]

THEOREM 3.3. *Let Assumptions 2.1, 2.2, and 3.1–3.5 hold. As $h \to 0$ and $T \to \infty$, we have the following.*

(a) *Under $\beta = 0$, $\tau(\hat{\sigma}) \to_d \mathbb{N}(0,1)$.*
(b) *If $\beta \neq 0$ and $\sum_{t=1}^{T-1} |x_t|/\sqrt{T} \to_p \infty$, then*

$$|\tau(\hat{\sigma})| \ge \frac{|\beta|}{\underline{v}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} |x_t| + O_p(1) \to_p \infty,$$

and hence $P\left[|\tau(\hat{\sigma})| > c\right] \to 1$ for any positive constant c.

[5]In other words, σ is of finite activity, so that the probability measure of any set $\{\omega : r \mapsto \sigma(r,\omega)$ has finitely many jumps in $r \in [0,1]\}$ is one.

Remark 3.1. The asymptotic power result in Theorem 3.3(b) implies that the testing power is mainly determined by the asymptotic behavior of $\sum_{t=1}^{T-1}|x_t|$. As an illustration, assume that there exists a diverging sequence p_T such that $p_T/T^{1/2}\to\infty$ and

$$\frac{1}{p_T}\sum_{t=1}^{T}|x_{t-1}|\to_d P \tag{9}$$

for a random variable $P>0$. Clearly, a wide class of time series satisfies (9) (see, e.g., Example 2.1). It then follows immediately from Theorem 3.3(b) that $|\tau(\hat\sigma)|\to_p\infty$ and the speed of divergence is no slower than $p_T/\sqrt{T}$.

Heuristically, we may consider the power of the proposed test by considering local alternatives in which $\beta\neq 0$, but $\beta\to 0$ at an appropriate rate. For our purpose, let $(P_T(r), 0\le r\le 1)$ be a continuous time process defined as

$$P_T(r)=\frac{1}{p_T}\sum_{t=1}^{[Tr]}\frac{|x_{t-1}|}{\sigma_T(t/T)}$$

for a diverging sequence p_T such that $p_T/T^{1/2}\to\infty$. We then assume instead of (9) that

$$P_T\to_d P \tag{10}$$

for a stochastic process $(P(r), 0\le r\le 1)$ having a positive support. If the convergence (10) holds jointly with the convergence in Assumption 2.2, then we may develop the power of the proposed test by considering the following alternative hypothesis:

$$\beta=\bar\beta\times\frac{\sqrt{T}}{p_T} \tag{11}$$

for a constant $\bar\beta\in\mathbb{R}\setminus\{0\}$. Clearly, the hypothesis (11) can be interpreted as a local alternative since $\beta\neq 0$ and $\beta\to 0$. Our construction of the local alternative (11) is useful to develop the asymptotic power result in a unified framework, especially when the covariate (x_t) is general, but satisfies (10). Under the local alternative (11), one can easily deduce from the proof of Theorem 3.3 with (10) that

$$\tau(\hat\sigma)\to_d\bar\beta P(1)+\mathbb{N}(0,1).$$

Clearly, $\tau(\hat\sigma)$ under (11) is not Gaussian asymptotically unless $P(1)$ is either constant or Gaussian.

Example 3.1. Let $\sigma_T(r)=\sigma$ for all $r\in[0,1]$.

(a) (x_t) be a stationary process such that $T^{-1}\sum_{t=1}^{T}|x_t|\to_p E|x_1|<\infty$. Under (11) with $p_T=T$, $\tau(\hat\sigma)\to_d\mathbb{N}\big((\bar\beta/\sigma)E|x_1|,\ 1\big)$.

(b) (x_t) be a unit-root (or near unit-root) process such that $T^{-3/2}\sum_{t=1}^{T}|x_t|\to_d\int_0^1|X(r)|dr$, where $(X(r))$ is the limiting Brownian motion (or

Ornstein–Uhlenbeck process) of (x_t) such that $X_T \to_d X$ for $X_T(r) = T^{-1/2}x_{[Tr]}$.[6] Under (11) with $p_T = T^{3/2}$, $\tau(\hat{\sigma}) \to_d (\bar{\beta}/\sigma)\int_0^1 |X(r)|dr + \mathbb{N}(0,1)$.

When $\sigma_T(r) = \sigma$, for all $r \in [0,1]$, and (x_t) is stationary with $E(x_t^2) < \infty$, the usual t-test procedure is a valid hypothesis testing procedure for the model (1). In this case, the asymptotic power property of the usual t-test is also well known under the local alternative hypothesis (11) with $p_T = T$, and is given by

$$t\text{-statistic} \to_d \mathbb{N}\big((\bar{\beta}/\sigma)(E(x_t^2))^{1/2},\ 1\big).$$

The ratio of the asymptotic biases of the t-test and our test, obtained in Example 3.1(a), is given by $(E(x_t^2))^{1/2}/E|x_t|$. Importantly, the ratio is always greater than one as long as $E(x_t^2) < \infty$ due to Jensen's inequality. This implies that the usual t-test is more powerful than our test under the ideal assumptions, even though the statistics in both tests diverge at the same rate $T^{1/2}$ under a fixed alternative hypothesis. However, when one or more of the ideal assumptions are violated, our test remains valid, whereas the usual t-test becomes invalid. This is another example of the traditional issue of trade-off between efficiency and robustness.

Remark 3.2. A number of works in statistics and econometrics have focused on robust inference using sign tests applied to different models, including time series regressions (see, among others, Dufour and Hallin, 1993; Campbell and Dufour, 1995; So and Shin, 2001; de la Peña and Ibragimov, 2017; Brown and Ibragimov, 2019; Kim and Meddahi, 2020, and the references therein). For instance, Campbell and Dufour (1995) propose sign tests for testing independence of a zero median time series Y_t with $P(Y_t = 0) = 0$, e.g., a time series with continuous distributions symmetric about zero, of past values of Y_t and another time series X_t. The tests in Campbell and Dufour (1995) are based on the observation that, under the above independence/orthogonality hypothesis, for any $T \geq 1$, the sign statistic like $S_0 = 0.5(\sum_{t=1}^{T} sign(Y_t X_{t-1}) + T)$ and its more general analogs follow a Binomial distribution with parameters T and 0.5: $S_0 \sim Bi(T, 0.5)$ (the results in Brown and Ibragimov (2019) imply that sign tests for general zero median or symmetric processes Y_t can be based on similar statistics with randomization over zero values of Y_t). Efron (1969), Edelman (1990), Dufour and Hallin (1993), Pinelis (1994), and de la Peña and Ibragimov (2017) consider related testing procedures based on bounds for tail probabilities of t-statistics of a parameter of interest (e.g., a location parameter or a regression/autoregression coefficient) under symmetry assumptions implied by (sharp) bounds on tail probabilities of weighted sums of i.i.d. symmetric Bernoulli r.v.'s.

Naturally, in the time series regression context, the above sign-based inference approaches are more robust to moment assumptions and heavy tails than the

[6] The reader is referred to Phillips (1987b) and Park (2003) for more discussion about the near unit-root process and its limiting behaviors. Here, in particular, the Ornstein–Uhlenbeck process X follows

$$dX(r) = cX(r)dr + dV(r), \quad X(0) = 0,$$

where V is a Brownian motion.

inference procedures based on the Gaussian asymptotics for the full-sample OLS and Cauchy estimators. Typically, the sign-based tests can be used without any moment conditions on the time series considered, e.g., under infinite variances. However, they usually require symmetry or zero median assumptions on the processes. Such assumptions are often too restrictive in empirical applications, including the analysis of financial markets due to the stylized fact of gain–loss asymmetry in financial returns (see, among others, Cont, 2001 and the references therein). Further, sign-based tests are less efficient than those on the Gaussian asymptotics for the OLS estimator under the validity of the latter tests.

Remark 3.3. Our method can be applied to a discrete time model and a discrete sample collected from an underlying continuous time model as in Choi et al. (2016). The main difference between our approach to Choi et al. (2016) is that we do not require the assumption $\delta \to 0$, where δ is the sampling interval of the discrete samples. Clearly, the method of Choi et al. (2016) is applicable to high-frequency data. Therefore, we may say that our method is more flexible since it can be applied to both high- and low-frequency data. The price we have to pay for the flexibility is the persistent volatility assumption $\sigma_T \to_d \sigma$ in Assumption 2.2. Persistent volatility is a well-known stylized fact of financial time series and, in our view, is best considered within the model formulation.

Our method is also comparable to the IVX approach proposed by Phillips and Magdalinos (2009). The IVX approach is based on a self-generated instrument obtained by differencing the predictor x_t and using an autoregressive filter to construct the instrument. As is shown in Phillips and Magdalinos (2009), the IVX approach is robust to a (near) unit-root or mildly explosive predictor. The Cauchy estimator approaches to inference, including ours and Choi et al. (2016), are restricted to a single regressor.[7] Unlike the Cauchy-based inferences, the IVX approach is applicable to predictive regressions with multiple regressors. We also note that the IVX approach allows for conditional heteroskedasticity. However, to the best of our knowledge, it is not known whether the IVX approach is valid when the volatility is persistent or the predictor is heavy-tailed with infinite second moments, and a continuous time extension of the IVX approach is not available in the literature. Therefore, our method and the IVX may be regarded as complementing each other.

Hansen (1995) provides a nonparametric generalized least squares (GLS) method for regression models with nonstationary volatility using the estimator $\hat{\sigma}$ to correct the heteroskedasticity. One should note that the assumptions on the limiting volatility σ are more general than those in Hansen (1995) and other work in the

[7] In general, a test relying on a single regressor exhibits size distortion when some relevant regressors are omitted. To overcome the issue induced by a single regressor in our approach, one may extend our approach to a multivariate setting based on the recent paper by Shephard (2020) in which a multivariate extension of the Cauchy estimator is proposed. An alternative extension is to use the parsimonious system approach (see Ghysels, Hill, and Motegi, 2020; Xu and Guo, 2022), which is based on a set of misspecified regression models with only one group of regressors, allowing a single regressor for each regression. We leave these extensions for future research.

literature on the topic. In particular, the assumptions in Hansen (1995) do not allow for structural changes or regime switching in the volatility process as the limiting volatility is assumed to have continuous sample paths almost surely. In contrast, the limiting volatility is allowed to have an arbitrary number of jumps in this paper, and hence structural changes or regime switching are allowed. Moreover, we further extend our model to have a two-factor volatility in Section 4.

4. AN EXTENSION TO TWO-FACTOR VOLATILITY MODELS

In this section, we generalize the model (1) to have a two-factor volatility in the regression error (u_t). More specifically, we assume that (ε_t) is conditionally heteroskedastic, rather than conditional homoskedastic as is assumed in Assumption 2.1(b).

Assumption 4.1. (a) $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = w_t^2$ and $E(w_t^2) = 1$. (b) $\max_{t\geq 1} E(|w_t|^{2\eta_1}) < \infty$, for some $\eta_1 > 2$. (c) (w_t) is α-mixing such that the mixing coefficient α satisfies $\alpha(k) \leq Ak^{-\eta_2}$, for some $A < \infty$ and $\eta_2 > (2\eta_1 + 2)/(\eta_1 - 2)$. (d) $hT^{\eta_3} \to \infty$ for some $\eta_3 > (\eta_2(1-2/\eta_1) - 2/\eta_1 - 2)/(\eta_2 + 2)$.

Under Assumptions 2.1(a) and 4.1, the regression error (u_t) in the model (1) can be written as $u_t = v_t w_t e_t$, where (e_t) is an MDS with respect to $(\mathcal{F}_t)$ such that $E(e_t^2|\mathcal{F}_{t-1}) = 1$. Clearly, (u_t) has two volatility factors, (v_t) and (w_t), where (v_t) is the long-run component by Assumption 2.2 and (w_t) is the short-run component by Assumption 4.1(c). Moreover, under Assumption 2.2 and the condition $E(w_t^2) = 1$ in Assumption 4.1, we can identify and estimate the persistent volatility component v_t by the nonparametric estimator (5). In particular, under Assumption 4.1, we may show that

$$\sup_{h\leq r\leq 1}\left|\frac{1}{hT}\sum_{t=1}^{T}(w_t^2-1)K_h(r-t/T)\right| = O_p\left((\log T/(hT))^{1/2}\right)$$

using an exponential inequality for a strongly mixing process (see, e.g., Liebscher, 1996, Thm. 2.1; Kristensen, 2009, Thm. 1; Vogt, 2012, Thm. 4.1). The above uniform convergence result for the mixing process is sufficient to develop the required uniform convergence of the volatility estimator as well as the validity of the inference procedure proposed in Section 3. We also note that the conditions for h and T in Assumptions 3.4 and 4.1(d) hold simultaneously for any $p,q \in [0,1/8)$ as long as Assumption 4.1(d) holds for some $\eta_3 > 1/4$, which is not stringent. For instance, if (w_t) is a stationary GARCH(1,1) process and β-mixing with exponential decay, which hold under some mild conditions (see, e.g., Carrasco and Chen, 2002; Francq and Zakoïan, 2006), then both Assumptions 3.4 and 4.1(d) hold for any $p,q \in [0,1/8)$ since Assumption 4.1(d) holds for any $\eta_3 > 1$.

For our purpose, we again consider the decomposition (7) of the nonparametric estimator (5), and write $\hat{\sigma}^2 = \sum_{k=1}^4 \hat{\sigma}_k^2$. We then can obtain the uniform convergence rate of each component $\hat{\sigma}_k^2$, for $k = 1,2,3,4$, as in Proposition 3.2, and

TABLE 1. Size for the continuous time models

T		$\bar{\kappa}=0$			$\bar{\kappa}=5$			$\bar{\kappa}=20$		
		5	20	50	5	20	50	5	20	50
CNST	OLS	42.2	42.0	43.0	19.5	19.5	19.7	11.1	11.2	10.9
	BQ	8.6	4.9	4.3	7.5	4.5	4.2	8.6	4.1	3.2
	RLRT	8.5	7.7	8.1	5.4	5.9	5.6	4.8	5.2	5.3
	Cauchy RT	5.3	4.9	5.3	5.2	5.4	4.7	5.5	5.1	5.1
	$\tau(\hat{\sigma})$	5.6	5.0	5.3	5.4	5.0	5.1	5.4	5.0	4.8
SB	OLS	38.3	38.8	39.9	29.6	30.8	31.2	24.3	26.4	26.0
	BQ	18.1	12.9	11.9	17.0	15.1	14.1	17.4	14.8	14.3
	RLRT	23.8	22.8	23.6	21.0	21.9	21.8	22.4	24.5	23.6
	Cauchy RT	5.6	5.0	5.1	5.2	5.3	5.0	5.4	5.0	4.9
	$\tau(\hat{\sigma})$	8.0	6.7	6.3	7.8	6.5	6.0	7.9	6.4	6.0
RS	OLS	42.9	43.6	44.6	22.0	23.4	24.5	14.9	18.9	19.5
	BQ	8.8	6.3	6.0	9.8	7.2	6.8	12.6	8.9	8.4
	RLRT	9.3	10.0	10.7	7.5	9.4	9.6	9.6	13.0	14.2
	Cauchy RT	5.0	4.8	5.2	4.9	4.9	4.9	5.4	5.1	4.8
	$\tau(\hat{\sigma})$	5.2	5.4	6.1	5.2	5.1	5.8	5.6	5.8	5.8
GBM	OLS	52.2	53.7	53.1	28.6	30.2	30.9	23.2	26.0	27.0
	BQ	16.8	12.5	11.3	13.9	12.4	13.2	15.8	11.7	12.0
	RLRT	21.7	22.3	21.9	16.3	17.8	19.0	21.4	23.3	23.5
	Cauchy RT	4.4	4.7	4.4	4.3	4.5	4.4	4.6	4.5	4.5
	$\tau(\hat{\sigma})$	5.4	5.5	6.1	5.7	5.7	5.9	5.7	5.9	6.5

Notes: The parameter $\bar{\kappa}$ measures the degree of persistence in the predictor. The sample size corresponds to T yearly observations (total $12T$ observations). CNST, SB, GBM, and RS denote, respectively, constant volatility, structural break, geometric Brownian motion, and regime switching in volatility.

establish the validity of the inference method relying on the test statistic (4) as in Theorem 3.3.

COROLLARY 4.1. *Let Assumptions 2.1(a) and (c), 2.2, 3.1–3.5, and 4.1 hold. As $h \to 0$ and $T \to \infty$, Proposition 3.2 and Theorem 3.3 remain valid.*

5. MONTE CARLO SIMULATIONS

This section provides the numerical results on finite sample performance of the proposed robust test based on $\tau(\hat{\sigma})$. We present the comparisons of the finite sample properties of the test with the test proposed by Choi et al. (2016) (denoted by Cauchy RT; RT for random time) and also two other tests considered in Choi et al. (2016): the Bonferroni Q-test of Campbell and Yogo (2006) (denoted by BQ) and the restricted likelihood ratio test of Chen and Deo (2009) (denoted by RLRT).

TABLE 2. Size for the discrete time models with CNST and SB

		$\bar{\kappa}=0$			$\bar{\kappa}=5$			$\bar{\kappa}=20$		
T		5	20	50	5	20	50	5	20	50
CNST	OLS	43.9	43.8	44.7	19.4	19.8	20.1	9.7	11.2	10.8
	BQ	8.4	5.2	4.8	7.8	4.9	4.5	9.2	4.1	3.4
	RLRT	8.3	8.0	8.1	5.2	5.4	5.3	4.1	5.4	5.3
	$\tau(\hat{\sigma})$	5.5	5.1	5.0	5.5	4.8	5.1	5.1	5.2	5.2
SB	OLS	38.0	39.6	40.0	29.1	31.1	31.4	22.1	26.1	26.8
	BQ	17.2	12.8	12.3	16.5	15.1	14.5	17.7	15.0	15.2
	RLRT	23.1	23.5	24.2	19.9	21.8	21.4	21.2	24.6	25.1
	$\tau(\hat{\sigma})$	8.0	6.7	6.3	7.9	6.2	5.8	7.5	6.5	6.2
ARCH(1)	OLS	45.0	44.1	43.5	23.5	22.5	21.2	17.2	17.0	15.2
$\alpha=0.5773$	BQ	9.7	5.4	4.8	9.5	5.8	4.6	13.1	6.3	4.6
$\xi=4$	RLRT	9.6	8.8	8.7	9.0	7.9	6.7	13.1	11.3	9.1
	$\tau(\hat{\sigma})$	6.1	5.4	6.0	6.1	5.2	5.4	6.0	5.9	6.1
ARCH(1)	OLS	45.8	44.0	43.6	24.4	24.1	22.6	19.7	19.8	18.1
$\alpha=0.7325$	BQ	10.2	6.2	5.2	10.4	7.1	5.8	14.7	8.3	6.8
$\xi=3$	RLRT	10.7	10.0	9.2	10.6	10.0	8.1	15.9	14.9	12.9
	$\tau(\hat{\sigma})$	5.9	5.8	6.5	6.2	5.6	6.0	6.4	6.1	6.1
IGARCH(1,1)	OLS	44.9	45.8	45.6	20.1	21.8	24.3	11.1	14.9	17.3
$\alpha=0.9$	BQ	8.9	5.8	6.0	7.8	6.0	6.9	9.3	5.3	6.4
$\beta=0.1$	RLRT	9.1	10.1	11.5	5.7	8.3	9.8	6.2	9.0	11.5
	$\tau(\hat{\sigma})$	6.2	5.5	5.5	5.8	5.6	6.0	5.9	5.8	5.7
IGARCH(1,1)	OLS	46.0	46.5	45.1	26.9	28.5	28.0	21.6	26.1	26.2
$\alpha=0.1$	BQ	11.7	8.3	8.1	12.7	10.4	10.9	16.4	12.2	13.4
$\beta=0.9$	RLRT	13.3	13.0	12.7	13.7	14.7	15.1	20.2	23.7	23.2
	$\tau(\hat{\sigma})$	6.3	6.4	7.4	6.9	6.7	7.2	6.6	6.9	6.9

Notes: The parameter $\bar{\kappa}$ measures the degree of persistence in the predictor. The sample size corresponds to T yearly observations (total $12T$ observations). CNST, SB, GBM, and RS denote, respectively, constant volatility, structural break, geometric Brownian motion, and regime switching in volatility.

We consider two different settings for simulation models: continuous time and discrete time data generating processes (DGPs). As for the continuous time DGPs, we follow the simulation designs of Choi et al. (2016). The data are generated in the continuous time setting using the following DGP:

$$dY_t = \frac{\bar{\beta}}{T} X_t dt + dU_t, \qquad dU_t = \sigma_t \left(dW_{1t} + \int_{\mathbb{R}} x \Lambda(dt, dx) \right), \tag{12}$$

$$dX_t = -\frac{\bar{\kappa}}{T} X_t dt + \sigma_t dW_{2t}, \tag{13}$$

where W_{1t} and W_{2t} are Brownian motions with $E(W_{1t}W_{2t}) = -0.98t$. We set the constant term in the predictive regression to be zero and use recursive de-meaning.

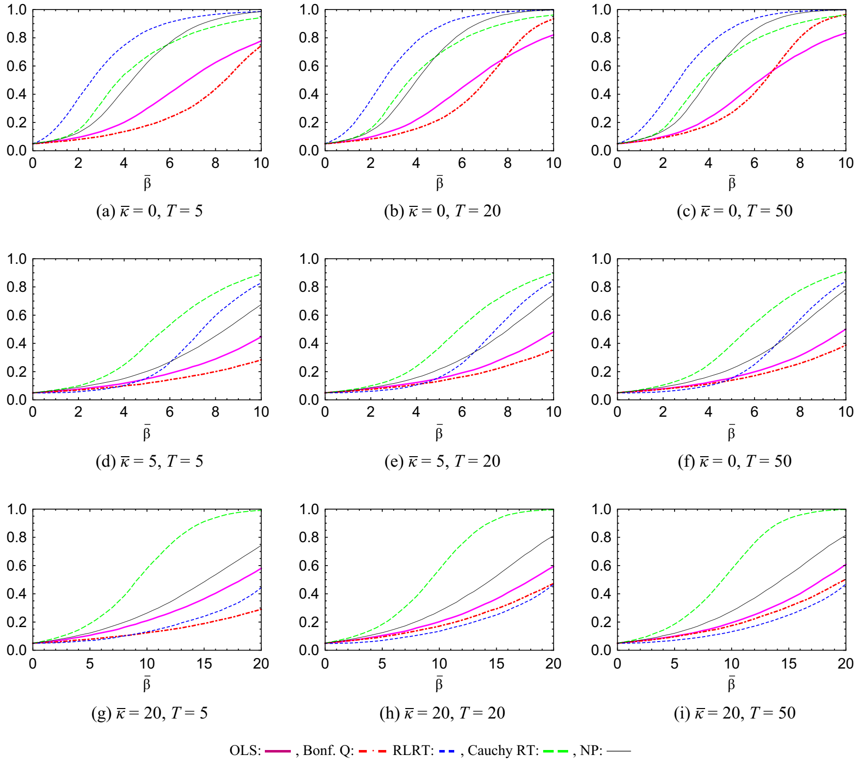


FIGURE 1. Power for SB (continuous time).

We assume that the continuous time models are observed at δ-intervals over T years with $\delta = 1/252$, which corresponds to daily observations of size $252T$.

The volatility process considered in the numerical results is assumed to follow one of the following models:

- Model CNST. *Constant volatility*: $\sigma_t^2 = \sigma_0^2$, $\sigma_0 = 1$.
- Model SB. *Structural break in volatility*: $\sigma_0 + (\sigma_1 - \sigma_0)1\{t/T \geq 4/5\}$ with $\sigma_0 = 1$ and $\sigma_1 = 4$.
- Model GBM. *Geometric Brownian motion*: $d\sigma_t^2 = \frac{1}{2}\frac{\bar{\omega}^2}{T}\sigma_t^2 dt + \frac{\bar{\omega}^2}{\sqrt{T}}\sigma_t^2 dZ_t$, where Z_t is a Brownian motion with $E(W_{1t}Z_t) = -0.4t$, and $\bar{\omega} = 9$.
- Model RS. *Regime switching*: $\sigma_t = \sigma_0(1 - s_t) + \sigma_1 s_t$, where s_t is a homogeneous Markov process indicating the current state of the world which is independent of both Y_t and X_t with the state space $\{0, 1\}$ and the transition matrix

$$P_t = \begin{pmatrix} 0.8 & 0.2 \\ 0.8 & 0.2 \end{pmatrix} + \begin{pmatrix} 0.2 & -0.2 \\ -0.8 & 0.8 \end{pmatrix} \exp\left(-\frac{\bar{\lambda}}{T}t\right),$$

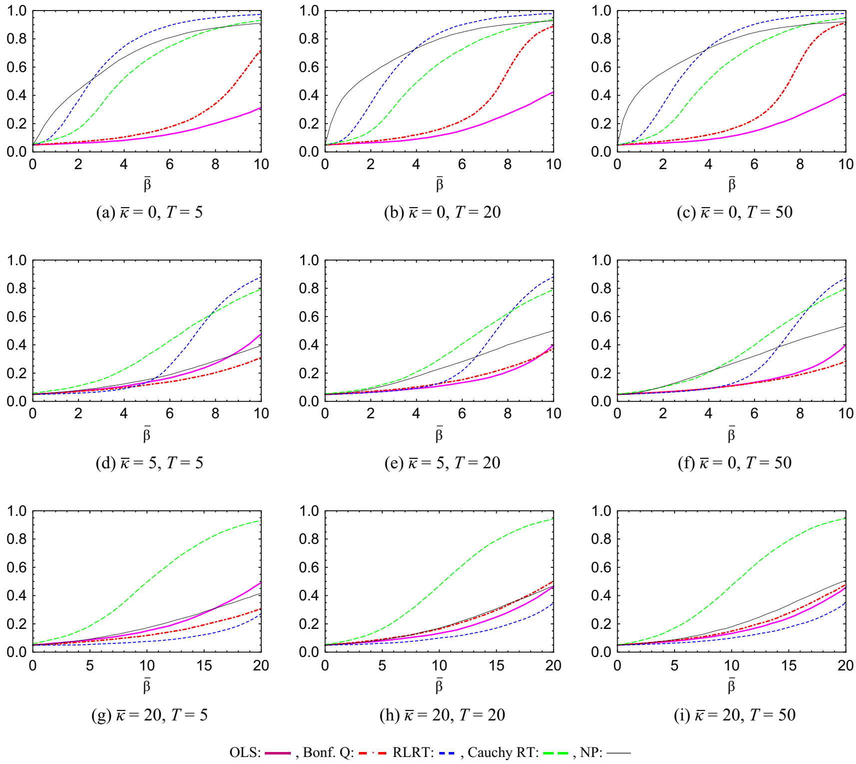


FIGURE 2. Power for GBM (continuous time).

where $\bar{\lambda} = 60$, $\sigma_0 = 1$, and $\sigma_1 = 4$. The process s_t is initialized by its invariant distribution.

We set the number of years $T \in \{5, 20, 50\}$ (which corresponds to 60, 240, and 600 monthly data) and consider the values $\bar{\kappa} \in \{0, 5, 10\}$ for the persistence parameter $\bar{\kappa}$ of X_t in (13).

As indicated before, in contrast to the Cauchy RT test in Choi et al. (2016), our test is applicable, not only in the continuous time models, but also in the discrete time framework. We consider the following discrete time models in the analysis of the finite sample performance of the tests:

$$y_t = \frac{\bar{\beta}}{T} x_{t-1} + \sigma_{\varepsilon,t} \varepsilon_t, \qquad x_t = \left(1 - \frac{\bar{\kappa}}{T}\right) x_{t-1} + \sigma_{\eta,t} \eta_t, \tag{14}$$

for $t = 2, \ldots, T$, where $T \in \{60, 240, 600\}$ (the same number of monthly observations as in continuous time simulations) and the same values of $\bar{\beta}$ and $\bar{\kappa}$. Here, the innovations (ε_t, η_t) are assumed to be multivariate normal with the correlation coefficient -0.98.

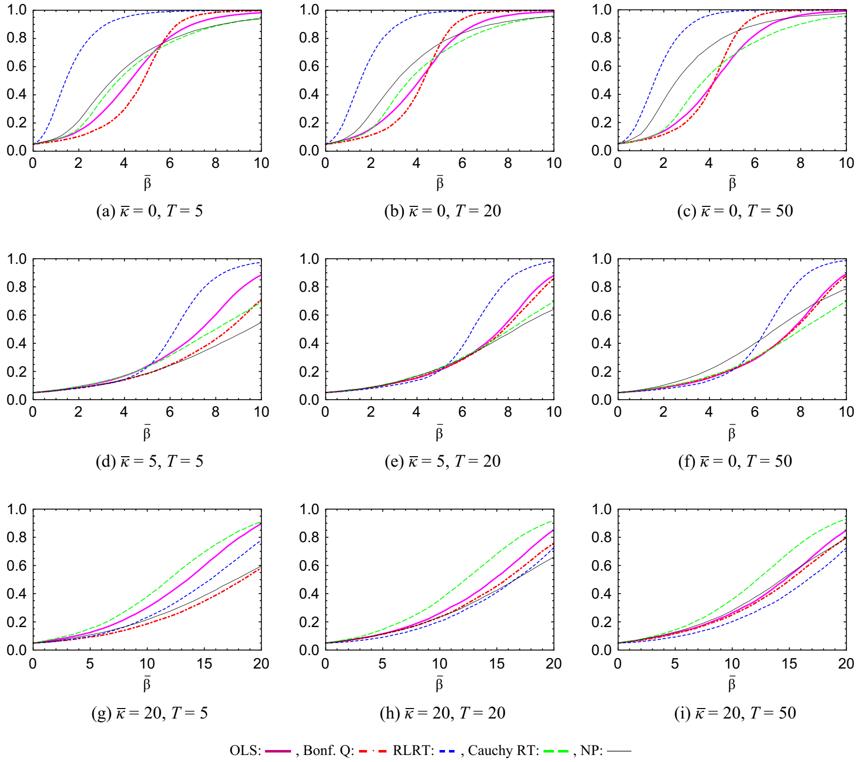


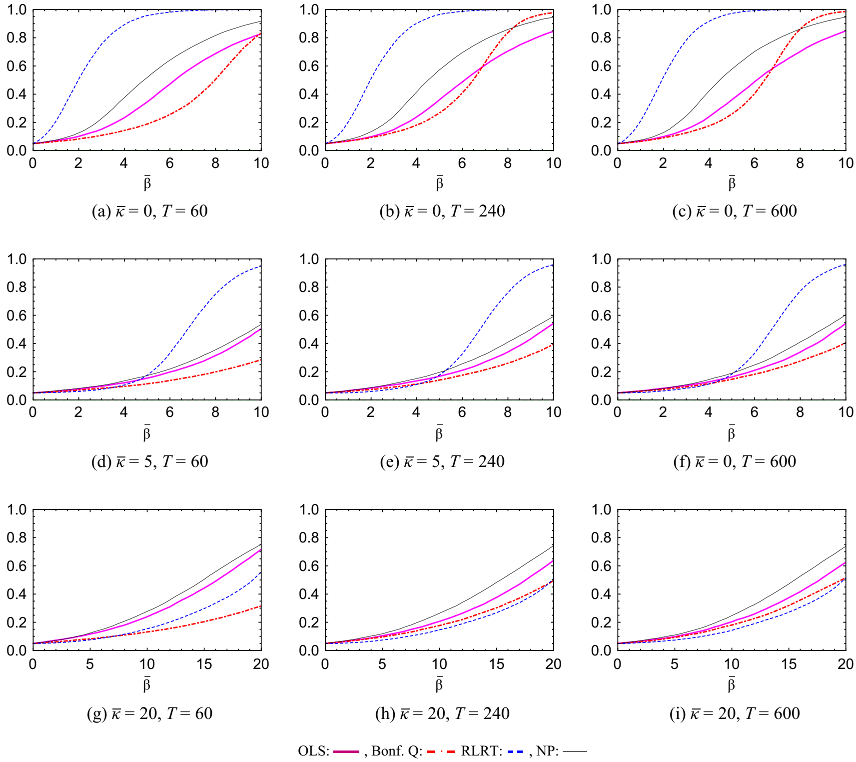
FIGURE 3. Power for RS (continuous time).

For the volatility processes in the discrete time setting, we consider three specifications: Model CNST and Model SB as in the continuous time setup, and GARCH volatility dynamics with

$$\sigma^2_{\varepsilon,t} = 1 + \alpha\varepsilon^2_{t-1} + \theta\sigma^2_{\varepsilon,t-1}, \qquad \sigma^2_{\eta,t} = 1 + \alpha\eta^2_{t-1} + \theta\sigma^2_{\eta,t-1}.$$

In the numerical analysis, we consider the ARCH(1) processes with $\theta = 0$, $\alpha = 0.5773$ (stationary with infinite fourth moment); $\theta = 0$, $\alpha = 0.7325$ (stationary with infinite third moment); and IGARCH(1,1) models with $\alpha = 0.9$, $\theta = 0.1$ and $\alpha = 0.1$, $\theta = 0.9$ (nonstationary). Note that the ARCH(1) processes in our simulations violate the moment conditions in Assumption 4.1. As shown in our simulation results below, our approach has reliable size and power properties even though the required moment conditions are violated.[8]

[8]See, among others, Davis and Mikosch (1998), Mikosch and Stărică (2000), Ibragimov, Pedersen, and Skrobotov (2021), and the references therein for the results on moment properties of GARCH processes and their importance in robust econometric inference.



(a) $\bar{\kappa} = 0$, $T = 60$ (b) $\bar{\kappa} = 0$, $T = 240$ (c) $\bar{\kappa} = 0$, $T = 600$

(d) $\bar{\kappa} = 5$, $T = 60$ (e) $\bar{\kappa} = 5$, $T = 240$ (f) $\bar{\kappa} = 0$, $T = 600$

(g) $\bar{\kappa} = 20$, $T = 60$ (h) $\bar{\kappa} = 20$, $T = 240$ (i) $\bar{\kappa} = 20$, $T = 600$

FIGURE 4. Power for SB (discrete time).

5.1. Finite Sample Size Properties

In this section, we analyze finite sample size properties of the no predictability tests by setting $\bar{\beta} = 0$ in the regression models (12) and (14). The numerical results on the finite sample size properties are presented in Tables 1 and 2.

Table 1 provides the finite sample size results for models CNST, SB, GBM, and RS in the continuous time setting. The finite sample size values for the OLS, BQ, RLRT, and Cauchy RT tests are exactly the same as those reported in Choi et al. (2016). These numerical results show that the size of the OLS, BQ, and RLRT tests is highly distorted for most of the time-varying volatility models considered. In contrast, the rejection probabilities of the proposed test are very close to their nominal levels, such as the Cauchy RT test, regardless of the values of $\bar{\kappa}$ and T, and the volatility models we consider in our simulations. For the 5% test, rejection probabilities stay between 4% to 8% without any exception.

As mentioned before, the Cauchy RT test is inapplicable in the discrete time settings. Table 2 provides the numerical results on finite sample size properties of all the tests except Cauchy RT under the discrete time settings. The quantitative

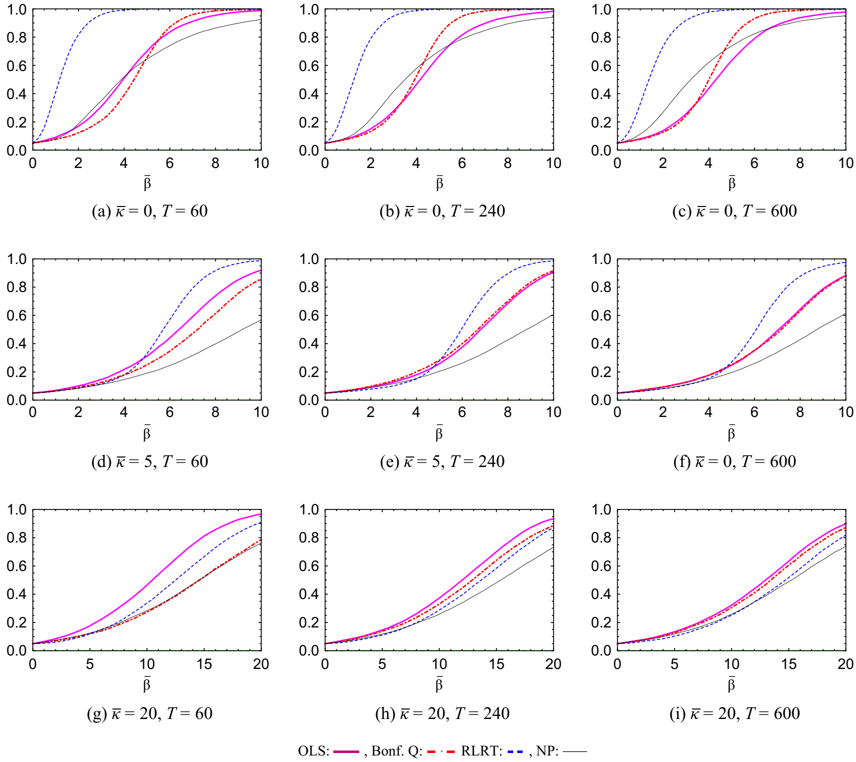


FIGURE 5. Power for GARCH with $\alpha = 0.9$ and $\theta = 0.1$ (discrete time).

and qualitative comparisons of the size properties of the tests are similar to the continuous time case. In summary, the finite sample size properties reported in Tables 1 and 2 show that the proposed test has a reliable size performance and is widely applicable for both discrete and continuous time settings.

5.2. Finite Sample Power Properties

Figures 1–6 present the results on finite sample power properties of the tests considered.[9] In our simulations, we consider the DGPs in (12) in continuous time and (14) in discrete time with $\bar{\beta}$ ranging from 0 to 20. All the power curves presented in the figures are size-adjusted. Taking into account the results of finite sample size performance of the tests and their comparisons, we mainly focus on two tests: the Cauchy RT and our test, in the analysis of finite sample power

[9]We report the power properties for Models SB, GBM, and RS (continuous time) as well as Models SB and GARCH (discrete time) in Section 5.2. The power properties for the other models, Model CNST (continuous time) as well as Models CNST and ARCH (discrete time), are presented in the Supplementary Material.

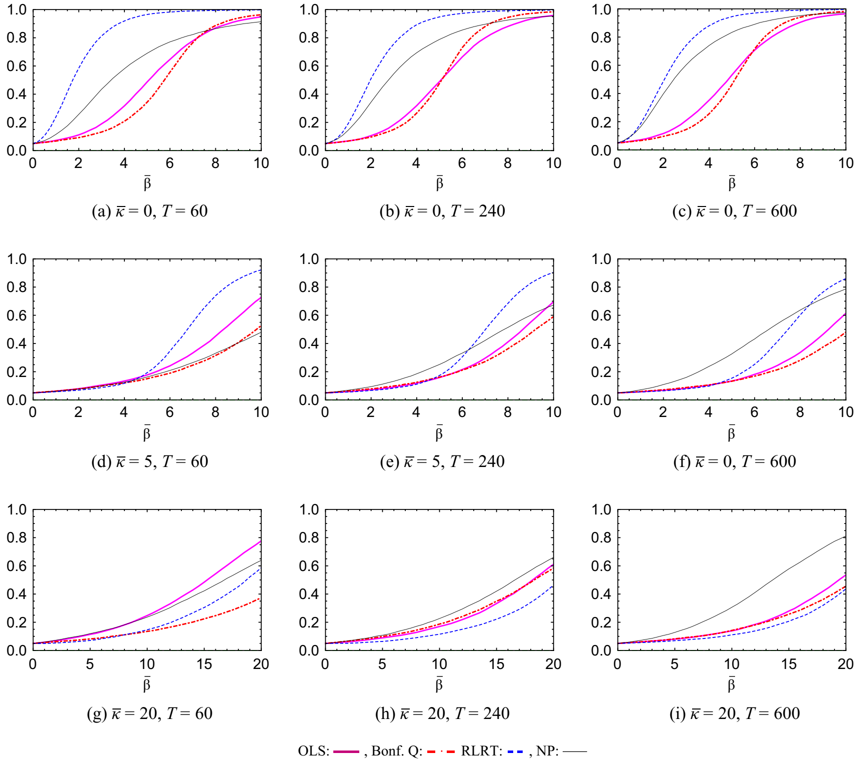


FIGURE 6. Power for GARCH with $\alpha = 0.1$ and $\theta = 0.9$ (discrete time).

properties. For comparison, the analysis also provides the numerical results on the finite sample power of the OLS, BQ, and RLRT tests.

In Figure 1, for the case of the structural break in volatility, one observes that the Cauchy RT test appears to be superior to other testing approaches (except in the cases with $\bar{\kappa} = 0$ and large $\bar{\beta}$). At the same time, the proposed test $\tau(\hat{\sigma})$ also appears to have good finite sample power properties especially in the case of highly persistent predictors.

Figure 2 provides the numerical results on finite sample properties of the tests in the geometric Brownian motion case. For the case of a unit-root regressor, the power properties of the proposed test based on $\tau(\hat{\sigma})$ appear to outperform those of the Cauchy RT test which in turn outperforms other tests considered. However, the finite sample power performance of the Cauchy RT test improves in the case of near unit-root regressor with $\bar{\kappa} = 5$ and $\bar{\kappa} = 20$.

The power curves for the regime switching case presented in Figure 3 demonstrate that the test based on the proposed test $\tau(\hat{\sigma})$ has better power properties than other tests in the case $\bar{\kappa} = 0$. For the case of the near unit-root persistence in the regressor, the power properties of the Cauchy RT test appear to be better than those of the test based on $\tau(\hat{\sigma})$ for relatively small sample sizes (small values of T). However, as the sample size increases, the test based on $\tau(\hat{\sigma})$ becomes more powerful than the Cauchy RT test (see, e.g., Figure 3 for the case $\bar{\kappa} = 5$ and $T = 50$). For large deviations from a unit-root regressor ($\bar{\kappa} = 20$), the Cauchy RT is more powerful than other tests, but the power curves appear to be very similar.

Figures 4–6 present the numerical results on power properties under discrete time settings for all the tests considered except Cauchy RT, which is inapplicable in discrete time settings. Results in the figures are provided for the cases of the structural break in volatility (Figure 4); the GARCH cases with $(\alpha, \theta) = (0.9, 0.1)$ (Figure 5) and $(\alpha, \theta) = (0.1, 0.9)$ (Figure 6). For all the cases, the conclusions on power properties of the tests and their comparisons are virtually the same as in the continuous time framework.

Overall, the numerical results on finite sample properties of the tests indicate good performance of the test based on $\tau(\hat{\sigma})$ in comparison to the Cauchy RT. Again, the latter test is inapplicable in the discrete time settings. Their relative finite sample performances vary across different models. Which test should be used in practice depends on the availability of high-frequency data as well as the size-power trade-off for a specific model. Therefore, the test proposed in this paper and the Cauchy RT complement rather than substitute one another.

6. CONCLUSION

Endogenously persistent regressors have been extensively analyzed in the predictive regression literature. A widely believed characteristic of stock returns is heteroskedastic and persistent volatility, which is often ignored in the predictive regression literature except for Choi et al. (2016). These two characteristics cause standard hypothesis tests to become substantially biased and often over-reject the null of no predictability. The main contribution of this paper is to provide an inference method that is designed to be robust to these problematic characteristics of predictive regression data. The proposed method relies on the Cauchy estimator and a kernel-based nonparametric correction of volatility. Its theoretical validity is provided by analyzing the asymptotic size and power properties. Moreover, it is shown through a simulation study that the proposed method has a reliable finite sample performance compared to the most advanced existing inference methods.

Our inference method is comparable to the method proposed by Choi et al. (2016). Similar to our method, their approach relies on the Cauchy estimator and a nonparametric volatility correction. However, their approach to the volatility correction is quite different from ours, and its applicability is limited to a predictive regression equipped with appropriate high-frequency data. In terms of finite sample properties, our method and the method by Choi et al. (2016) perform well and have good size and power performances under continuous time settings. However, unlike our method, Choi et al. (2016) method is not applicable under discrete time settings. In contrast, our method can be applied to a discrete time model as well as a discrete sample collected from an underlying continuous time model. Therefore, our method is more flexible and widely applicable since it can be applied to both high- and low-frequency data.

A further approach to robust inference in predictive regressions under heterogeneous and persistent volatility as well as endogenous, persistent, or heavy-tailed regressors is provided by the simple to implement robust t-statistic inference approach (see Ibragimov and Müller, 2010) based on asymptotically normal group Cauchy estimators of a regression parameter of interest. This approach will be explored in a companion paper now in preparation.

SUPPLEMENTARY MATERIAL

Ibragimov, R., Kim, J, and Skrobotov, A. (2022): Supplement to "New robust inference for predictive regressions," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466623000117

REFERENCES

Baillie, R.T. (1996) Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 5–59.

Bercu, B. & A. Touati (2008) Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability* 18(5), 1848–1869.

Billingsley, P. (1986) *Convergence of Probability Measures*. Wiley.

Brown, D. & R. Ibragimov (2019) Sign tests for dependent observations. *Econometrics and Statistics* 10, 1–8.

Campbell, B. & J.-M. Dufour (1995) Exact nonparametric orthogonality and random walk tests. *Review of Economics and Statistics* 77, 1–16.

Campbell, J. & M. Yogo (2006) Efficient tests of stock return predictability. *Journal of Financial Econometrics* 81, 27–60.

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Cauchy, A. (1836) On a new formula for solving the problem of interpolation in a manner applicable to physical investigations. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 8(49), 459–468.

Cavaliere, G. (2004) Testing stationarity under a permanent variance shift. *Economics Letters* 82, 403–408.

Cavaliere, G. & A.R. Taylor (2007) Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics* 140, 919–947.

Cavaliere, G. & A.R. Taylor (2009) Heteroskedastic time series with a unit root. *Econometric Theory* 25, 1228–1276.

Chan, N. & Q. Wang (2015) Nonlinear regressions with nonstationary time series. *Journal of Econometrics* 185, 182–195.

Chen, W. & R. Deo (2009) Bias reduction and likelihood-based almost exactly sized hypothesis testing in predictive regressions using the restricted likelihood. *Econometric Theory* 25, 1143–1179.

Choi, Y., S. Jacewitz, & J.Y. Park (2016) A reexamination of stock return predictability. *Journal of Econometrics* 192, 168–189.

Chung, H. & J.Y. Park (2007) Nonstationary nonlinear heteroskedasticity in regression. *Journal of Econometrics* 137, 230–259.

Cont, R. (2001) Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1, 223–236.

Davis, R.A. & T. Mikosch (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Annals of Statistics* 26, 2049–2080.

de la Peña, V.H. (1999) A general class of exponential inequalities for martingales and ratios. *Annals of Probability* 27(1), 537–564.

de la Peña, V.H. & R. Ibragimov (2017) Sharp probability inequalities for random polynomials, generalized sample cross-moments, and Studentized processes. In I. Pinelis (ed.), *Inequalities and Extremal Problems in Probability and Statistics: Selected Topics*, pp. 159–187. Elsevier.

de la Peña, V.H., R. Ibragimov, & S. Sharakhmetov (2003) On extremal distributions and sharp L_p-bounds for sums of multilinear forms, *Annals of Probability* 31, 630–675.

Dufour, J.-M. & M. Hallin (1993) Improved Eaton bounds for linear combinations of bounded random variables, with statistical applications. *Journal of the American Statistical Association* 88, 1026–1033.

Edelman, D. (1990) An inequality of optimal order for the tail probabilities of the t statistic under symmetry. *Journal of the American Statistical Association* 85, 120–122.

Efron, B. (1969) Student's t test under symmetry conditions. *Journal of the American Statistical Association* 64, 1278–1302.

Elliott, G. & J.H. Stock (1994) Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory* 10, 672–700.

Embrechts, P., C. Klüppelberg, & T. Mikosch (1997) *Modelling Extremal Events for Insurance and Finance*. Springer.

Francq, C. & J.-M. Zakoïan (2006) Mixing properties of a general class of GARCH (1, 1) models without moment assumptions on the observed process. *Econometric Theory* 22, 815–834.

Ghysels, E., J.B. Hill, & K. Motegi (2020) Testing a large set of zero restrictions in regression models, with an application to mixed frequency Granger causality. *Journal of Econometrics* 218, 633–654.

Granger, C.W.J. & D. Orr (1972) Infinite variance and research strategy in time series analysis. *Journal of the American Statistical Association* 67, 275–285.

Hansen, B.E. (1992) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.

Hansen, B.E. (1995) Regression with nonstationary volatility. *Econometrica* 63, 1113–1132.

Hansen, P.R. & A. Lunde (2014) Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory* 30, 60–93.

Harvey, D.I., S.J. Leybourne, & Y. Zu (2019) Testing explosive bubbles with time-varying volatility. *Econometric Reviews* 38, 1131–1151.

Ibragimov, M., R. Ibragimov, & J. Walden (2015) *Heavy-Tailed Distributions and Robustness in Economics and Finance*. Lecture Notes in Statistics, vol. 214. Springer.

Ibragimov, R. & U. Müller (2010) t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics* 28, 453–468.

Ibragimov, R., R.S. Pedersen, & A. Skrobotov (2021) New Approaches to Robust Inference on Market (Non-)Efficiency, Volatility Clustering and Nonlinear Dependence. Working paper, Imperial College Business School and the University of Copenhagen. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3580916.

Ibragimov, R. & P.C.B. Phillips (2008) Regression asymptotics using martingale convergence methods. *Econometric Theory* 24, 888–947.

Jacquier, E., N.G. Polson, & P.E. Rossi (2004) Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122, 185–212.

Kim, C.S. & P.C.B. Phillips (2006) Log Periodogram Regression: The Nonstationary Case. Cowles Foundation Discussion Paper no. 1587. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=937915.

Kim, J. & N. Meddahi (2020) Volatility regressions with fat tails. *Journal of Econometrics* 218, 690–713.

Kim, J. & J.Y. Park (2017) Asymptotics for recurrent diffusions with application to high frequency regression. *Journal of Econometrics* 196, 37–54.

Kristensen, D. (2009) Uniform convergence rates of kernel estimators with heterogeneous dependent data. *Econometric Theory* 25, 1433–1445.

Kurtz, T.G. & P. Protter (1991) Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035–1070.

Leadbetter, M. & H. Rootzén (1988) Extremal theory for stochastic processes. *Annals of Probability* 16, 431–478.

Liebscher, E. (1996) Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and their Applications* 65, 69–80.

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

Park, J.Y. (2003) Weak Unit Roots. Unpublished manuscript. http://www.ruf.rice.edu/econ/papers/2003papers/17park.pdf.

Phillips, P.C.B. (1987a) Time-series regression with a unit root. *Econometrica* 55, 277–301.

Phillips, P.C.B. (1987b) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.

Phillips, P.C.B. (1999) Discrete Fourier Transforms of Fractional Processes. Cowles Foundation Discussion paper no. 1243.

Phillips, P.C.B. (2015) Halbert White Jr. Memorial JFEC Lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics* 13, 521–555.

Phillips, P.C.B. & T. Magdalinos (2007) Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.

Phillips, P.C.B. & T. Magdalinos (2009) Econometric Inference in the Vicinity of Unity, Working paper, Yale University and University of Nottingham.

Phillips, P.C.B., J.Y. Park, & Y. Chang (2004) Nonlinear instrumental variable estimation of an autoregression. *Journal of Econometrics* 118, 219–246.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Pinelis, I. (1994) Extremal probabilistic problems and Hotelling's t test under a symmetry condition. *Annals of Statistics* 22, 357–368.

Pollard, D. (1984) Convergence of Stochastic Processes, *Springer Series in Statistics*. Springer.

Samorodnitsky, G., S.T. Rachev, J.-R. Kurz-Kim, & S.V. Stoyanov (2007) Asymptotic distributions of unbiased linear estimators in the presence of heavy-tailed stochastic regressors and residuals. *Probability and Mathematical Statistics* 27, 275–302.

Shephard, N. (2020) An Estimator for Predictive Regression: Reliable Inference for Financial Economics. Working paper, Harvard University.

So, B. & D. Shin (1999) Cauchy estimators for autoregressive processes with applications to unit root tests and confidence intervals. *Econometric Theory* 15, 165–176.

So, B. & D. Shin (2001) An invariant sign tests for random walks based on recursive median adjustment. *Journal of Econometrics* 102, 197–229.

Stambaugh, R.F. (1999) Predictive regressions. *Journal of Financial Economics* 54, 375–421.

Vogt, M. (2012) Nonparametric regression for locally stationary time series. *Annals of Statistics* 40, 2601–2633.

Wang, Q. & N. Chan (2014) Uniform convergence rates for a class of martingales with application in non-linear cointegrating regression. *Bernoulli* 20, 207–230.

Wang, Q., Y.-X. Lin, & C.M. Gulati (2003) Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory* 19, 143–164.

Xu, K.-L. & J. Guo (2022) A new test for multiple predictive regression. *Journal of Financial Econometrics*, nbac030. https://doi.org/10.1093/jjfinec/nbac030.

Xu, K.-L. & P.C.B. Phillips (2008) Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics* 142, 265–280.