Economics Letters 148 (2016) 10–16

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

A nonparametric approach to test for predictability☆

Zhiyuan Pan [a,b], Yudong Wang [c,*], Chongfeng Wu [d]

[a] Institute of Chinese Financial Studies, Southwestern University of Finance and Economics, China
[b] Collaborative Innovation Center of Financial Security, China
[c] School of Economics and Management, Nanjing University of Science and Technology, China
[d] Antai College of Economics & Management, Shanghai Jiao Tong University, China

HIGHLIGHTS

- We propose a nonparametric method to test for predictability.
- Linear and nonlinear DGPs are considered in Monte Carlo analysis.
- Our test displays more corrected size than three existing methods.
- Our test has higher power than two nonparametric causality tests.

ARTICLE INFO

Article history:
Received 30 July 2016
Received in revised form 18 August 2016
Accepted 8 September 2016
Available online 20 September 2016

JEL classification:
C12
G17
Q47

Keywords:
Nonparametric test
Predictability
Size
Power

ABSTRACT

Predictability of macroeconomic and financial variables is an important issue in economics. In this paper, we propose a nonparametric test for the predictability of the direction of price changes. The Monte Carlo simulation results show that our method displays better finite-sample property than the traditional parametric Granger causality test (Granger, 1969) and two nonparametric causality tests of Hiemstra and Jones (1994) and Diks and Panchenko (2006).

1. Introduction

This paper develops a nonparametric test for the predictability of direction of price changes. We compare the size and power property of our test with three existing methods including the traditional parametric Granger causality test (Granger, 1969) and two nonparametric causality tests of Hiemstra and Jones (1994) and Diks and Panchenko (2006). The linear and nonlinear models are employed to generate simulation data. We find that for both linear and nonlinear data generating processes (DGP), our statistic displays more corrected size and higher power than two existing nonparametric causality statistics. Although the parametric Granger test presents slightly higher power than our method, it has much higher sizes than the nominal significance levels, suffering from the problem of over-rejection.

The remainder of this paper is organized as follows: Section 2 shows the details about our proposed nonparametric test for predictability. Section 3 concludes the paper.

2. A nonparametric test

Let $\{R_t^A, R_t^B\}_{t=1}^T$ denote the logarithmic returns on two financial assets, where T is the sample size. Following Hong et al. (2007) and Li (2014), R_t^A and R_t^B are standardized to have mean zero and unit variance for the simplicity of statistical inference and computation. Given an exceedance level c and $R_t^A > c^A$ or $R_t^A < -c^A$ at time t,

☆ This work was supported by the Chinese National Science Foundation through grant number 71501095 (Yudong Wang), 71320107002 (Chongfeng Wu), 71601161 (Zhiyuan Pan) and the Fundamental Research Funds for the Central Universities through grant number 3301100004005040042 (Zhiyuan Pan).

* Correspondence to: Xiaolinwei 200, Xuanwu District, Nanjing 210094, China.
E-mail address: wangyudongnj@126.com (Y. Wang).

http://dx.doi.org/10.1016/j.econlet.2016.09.006
0165-1765/

the probability of k step ahead prediction is defined as

$$pred(c,k) = Pr(R^B_{t+k} > c^B | R^A_t > c^A) + Pr(R^B_{t+k} < -c^B | R^A_t < -c^A) \equiv pred^+(c,k) + pred^-(c,k), \quad (1)$$

where k and $c = (c^A, c^B)$ are both nonnegative constant. Eq. (1) calculates the probability that the normalized return of asset B is higher than the threshold c^B or lower than $-c^B$ when past return of asset A is higher than c^A or lower than $-c^A$.

Similarly, the probability of k step ahead non-prediction can be defined as

$$npred(c,k) = Pr(R^B_{t+k} < c^B | R^A_t > c^A) + Pr(R^B_{t+k} > c^B | R^A_t < -c^A) \equiv npred^+(c,k) + npred^-(c,k). \quad (2)$$

We are interested in whether the probability of prediction is significantly greater than the probability of non-prediction. Thus, our test is for a composite hypothesis. The null hypothesis is

$$\boldsymbol{H}_0 : pred(c,k) = npred(c,k), \quad \text{for all} \quad c \geq 0, \quad (3)$$

and the alternative hypothesis is

$$\boldsymbol{H}_1 : pred(c,k) > npred(c,k), \quad \text{for some} \quad c \geq 0. \quad (4)$$

Intuitively, if the null hypothesis in (3) is true, then the difference vector below

$$pred(c,k) - npred(c,k) = [pred(c_1,k) - npred(c_1,k), \ldots, pred(c_m,k) - npred(c_m,k)]', \quad (5)$$

must be equal to zero, where $c_1 = (c^A_1, c^B_1), \ldots, c_m = (c^A_m, c^B_m)$ are m chosen nonnegative exceedance levels.

To construct a feasible test statistic, we need to estimate both $pred(c_i,k)$ and $npred(c_i,k)$ for $1 \leq i \leq m$. According to Eqs. (1) and (2), we should estimate $Pr(R^B_{t+k} > c^B_i | R^A_t > c^A_i)$, $Pr(R^B_{t+k} < -c^B_i | R^A_t < -c^A_i)$, $Pr(R^B_{t+k} < c^B_i | R^A_t > c^A_i)$ and $Pr(R^B_{t+k} > -c^B_i | R^A_t < -c^A_i)$. The corresponding estimators are given as

$$\frac{\sum_{t=1}^{T-k} I[R^B_{t+k} > c^B_i] I[R^A_t > c^A_i]}{\sum_{t=1}^{T-k} I[R^A_t > c^A_i]}, \quad (6)$$

$$\frac{\sum_{t=1}^{T-k} I[R^B_{t+k} < -c^B_i] I[R^A_t < -c^A_i]}{\sum_{t=1}^{T-k} I[R^A_t < -c^A_i]}, \quad (7)$$

$$\frac{\sum_{t=1}^{T-k} I[R^B_{t+k} < c^B_i] I[R^A_t > c^A_i]}{\sum_{t=1}^{T-k} I[R^A_t > c^A_i]}, \quad (8)$$

and

$$\frac{\sum_{t=1}^{T-k} I[R^B_{t+k} > -c^B_i] I[R^A_t < -c^A_i]}{\sum_{t=1}^{T-k} I[R^A_t < -c^A_i]}, \quad (9)$$

where $I[\cdot]$ is an indicator function which takes the value of 1 when the condition in the square bracket is satisfied and 0 otherwise.

Substituting (6) and (7) into (1) yields the estimator of $pred(c_i,k)$,

$$\widehat{pred}(c_i,k) = \frac{\sum_{t=1}^{T-k} I[R^B_{t+k} > c^B_i] I[R^A_t > c^A_i]}{\sum_{t=1}^{T-k} I[R^A_t > c^A_i]} + \frac{\sum_{t=1}^{T-k} I[R^B_{t+k} < -c^B_i] I[R^A_t < -c^A_i]}{\sum_{t=1}^{T-k} I[R^A_t < -c^A_i]}, \quad (10)$$

and we obtain the estimator of $npred(c_i,k)$ by substituting (8) and (9) into (2),

$$\widehat{npred}(c_i,k) = \frac{\sum_{t=1}^{T-k} I[R^B_{t+k} < c^B_i] I[R^A_t > c^A_i]}{\sum_{t=1}^{T-k} I[R^A_t > -c^A_i]} + \frac{\sum_{t=1}^{T-k} I[R^B_{t+k} > -c^B_i] I[R^A_t < -c^A_i]}{\sum_{t=1}^{T-k} I[R^A_t < -c^A_i]}. \quad (11)$$

It can be proven that under null hypothesis (3) and the regular assumptions in Appendix, $\sqrt{T}(\widehat{pred}(c,k) - \widehat{npred}(c,k))$ has an asymptotic normal distribution with mean zero and a positive definite variance–covariance matrix, $\Omega(k)$. That is $\sqrt{T}(\widehat{pred}(c,k) - \widehat{npred}(c,k)) \xrightarrow{d} \mathcal{N}(0, \Omega(k))$. $\Omega(k)$ can consistently be estimated by

$$\hat{\Omega}(k) = \sum_{l=1-T}^{T-1} K(l/p)\hat{\delta}_l(k), \quad (12)$$

where the element (i,j) of the matrix, $\hat{\delta}_l(k)$, is given as

$$\hat{\delta}_l(c_i, c_j, k) = \frac{1}{T} \sum_{t=|l|+1}^{T} \hat{\eta}_t(c_i,k)\hat{\eta}_{t-|l|}(c_j,k), \quad (13)$$

and

$$\begin{aligned}\hat{\eta}_t(c_i,k) = & \frac{(I[R^B_{t+k} > c^B_i] - \widehat{pred}^+(c_i,k)) I[R^A_t > c^A_i]}{\sum_{t=1}^{T-k} I(R^A_t > c^A_i)/(T-k)} \\ & + \frac{(I[R^B_{t+k} < -c^B_i] - \widehat{pred}^-(c_i,k)) I[R^A_t < -c^A_i]}{\sum_{t=1}^{T-k} I(R^A_t < -c^A_i)/(T-k)} \\ & - \frac{(I[R^B_{t+k} < c^B_i] - \widehat{npred}^+(c_i,k)) I[R^A_t > c^A_i]}{\sum_{t=1}^{T-k} I(R^A_t > c^A_i)/(T-k)} \\ & - \frac{(I[R^B_{t+k} > -c^B_i] - \widehat{npred}^-(c_i,k)) I[R^A_t < -c^A_i]}{\sum_{t=1}^{T-k} I(R^A_t < -c^A_i)/(T-k)} \\ \equiv & \hat{\eta}^{p+}_t(c_i,k) + \hat{\eta}^{p-}_t(c_i,k) - \hat{\eta}^{np+}_t(c_i,k) - \hat{\eta}^{np-}_t(c_i,k), \end{aligned} \quad (14)$$

and $K(\cdot)$ is a kernel function, l denotes a suitable weight to each lag of order, p means the lag truncation order or smoothing parameter. In this paper, we adopt the Bartlett kernel as

$$K(x) = (1 - |x|)I(|x| < 1), \quad (15)$$

which is famous and is used by Newey and West (1994), Hong et al. (2007) and many others.

According to (10)–(12), we are ready to form a statistic testing for the null hypothesis as

$$J(k) = T\left(\widehat{pred}(c,k) - \widehat{npred}(c,k)\right)\hat{\Omega}^{-1}(k) \times \left(\widehat{pred}(c,k) - \widehat{npred}(c,k)\right). \quad (16)$$

Our test is based on a sample average. In this context, it is not surprising that we can obtain the asymptotic normality of the proposed statistic. The following theorem gives the asymptotic distribution theory for our statistic under the null hypothesis.

Theorem 1. *Suppose Assumptions hold and $p = p(T) \to \infty$, $p/T \to 0$ as $T \to \infty$. Under the null hypothesis and when $T \to \infty$, we have*

$$J(k) = T\left(\widehat{pred}(c,k) - \widehat{npred}(c,k)\right)\hat{\Omega}^{-1}(k)\left(\widehat{pred}(c,k) - \widehat{npred}(c,k)\right) \xrightarrow{d} \chi^2(m), \quad (17)$$

where $\chi^2(m)$ denotes a chi-square distribution with the degrees of freedom m.

3. Monte Carlo studies

To investigate the performance of the proposed $J(k)$ statistics in the finite sample, we conduct an experiment using the Monte Carlo method. For comparison, we also compute the size and power of Granger (1969)'s statistics, Hiemstra and Jones (1994)'s statistics (HJ thereafter) and Diks and Panchenko (2006)'s statistics (DP thereafter). Two models are used as data generating processes (DGP) for $\{R_t^A, R_t^B\}_{t=1}^T$. The first linear model is expressed as

$$\begin{aligned} R_t^A &= (h_t^A)^{1/2} z_t^A \\ h_t^A &= 0.005 + 0.05(R_{t-1}^A)^2 + 0.9h_{t-1}^A \\ R_t^B &= 0.1R_{t-1}^A + \epsilon_t^B = 0.1R_{t-1}^A + (h_t^B)^{1/2} z_t^B \\ h_t^B &= 0.005 + 0.05(\epsilon_{t-1}^B)^2 + 0.9h_{t-1}^B, \end{aligned} \tag{18}$$

where z_t^A and z_t^B are standard normal distribution. In this model, the conditional volatility is assumed to follow a GARCH process, consistent with the stylized fact of volatility clustering in financial markets. From Eq. (18), we can see that R_t^A will help to improve the prediction of R_t^B. In our design, the null hypothesis is that $\{R_t^A\}$ cannot predict $\{R_t^B\}$ for the analysis of size performance.

The second model used to generate data is the symmetric Joe-Clayton copula, which is modified by Patton (2006) and given as

$$\begin{aligned} C^{sjc}(u_t, v_t; \tau_u, \tau_l) = 0.5\big(&C^{jc}(u_t, v_t; \tau_u, \tau_l) + C^{jc}(1 - u_t, 1 \\ &- v_t; \tau_u, \tau_l) + u_t + v_t - 1\big), \end{aligned} \tag{19}$$

where $C^{jc}(\cdot)$ denotes the Joe-Clayton copula,

$$\begin{aligned} C^{jc}(u_t, v_t; \tau_u, \tau_l) &= 1 - (1 - \{[1 - (1 - u_t)^\kappa]^{-\gamma} \\ &\quad + [1 - (1 - v_t)^\kappa]^{-\gamma} - 1\}^{-1/\gamma})^{1/\kappa} \\ \kappa &= 1/\log_2(2 - \tau_u) \\ \gamma &= -1/\log_2(\tau_l) \\ \tau_u, \tau_l &\in (0, 1). \end{aligned}$$

The two parameters τ_u and τ_l measure the upper and lower tail dependence, rather than the linear correlation (see Patton, 2006 for more details). In our setting, we assign $(\tau_u, \tau_l) = (0.05, 0.05), (0.1, 0.1)$ for the symmetric case and $(\tau_u, \tau_l) = (0.05, 0.1)$ for asymmetric case, respectively. The asymmetric phenomenon is also viewed as a stylized fact in financial markets (Ang and Chen, 2002; Hong et al., 2007). In order to investigate the predictive effect, we make $R_t^B = u_t$ and $R_{t-1}^A = v_t$. Again, the null hypothesis is that R^A cannot predict R^B.

In the simulation study, we choose $k = 1$ and $c = 0$ to match with the HJ and DP statistics. Furthermore, $c^A = Q5$, $c^A = Q10$ and $c^A = Q15$ are adopted to examine the behavior of our statistics in tail area, where Q5, Q10 and Q15 mean the quantiles of the returns for the cumulative probabilities of 5%, 10% and 15%, respectively. In these cases, the exceedance levels are asymmetric because $c^B = 0$. Certainly, we can impose more asymmetric exceedance levels. Note that the implementation of exceedance level reduces the effective sample size because of the conditioning, and so it is harder to get to the asymptotic properties. We perform 10,000 Monte Carlo simulations for different sample lengths and record times for each statistic. The rejection times divided by 10,000 is our empirical proportion of rejections.

Table 1 shows the size performances of *Granger*, HJ, DP and our J tests for different sample sizes. We consider the nominal significance levels of 5% and 10%. The empirical sizes show that the Granger statistics[1] are severely over-rejected while HJ and DP statistics tend to be slightly under-rejected. As a comparison, the sizes of our J statistics are closer to the corresponding nominal significance levels than the other three statistics, indicating that our statistic is more reliable under the null hypothesis.

Table 2 reports the power of these statistics. Similar to Diks and Panchenko (2006), we use the same DGP when doing power analysis but interchange the positions of $\{R^A\}$ and $\{R^B\}$ in Eqs. (18) and (19). Then the null hypothesis tested becomes $\{R^A\}$ cannot predict $\{R^B\}$ accordingly. As expected, *Granger* test has highest power since there exists over-rejection. Comparing with the other two nonparametric tests, we can see that our J statistic has higher power than both DP and HJ statistics regardless of whether linear or nonlinear DGPs are used.

4. Conclusions

This paper develops a nonparametric test for the predictability based on the accuracy of direction prediction. The simulation results show that our method displays better size and power property than existing three causality tests.

Appendix. Proof

Let C be a large constant, which may take different values at different places. To derive the asymptotic distribution under the null hypothesis, we need some regularity conditions given below:

Assumption 1. $\{R_t^A, R_t^B\}$ has zero mean and is a bivariate stationary α-mixing process with $\sum_{j=-\infty}^{\infty} j^2 \alpha(j)^{\kappa/(\kappa-1)} < \infty$ and $E(|R_t^A|^{4\kappa}) + E(|R_t^B|^{4\kappa}) < \infty$ for some $\kappa > 1$.

Assumption 2. The kernel function $K(\cdot) : \mathbb{R} \to [-1, 1]$ is symmetric about 0 and is continuous at all points on support $\mathbb{R}$ except a finite number of points, with $K(0) = 1$ and $\int_{-\infty}^{\infty} |K(x)|dx < \infty$.

Assumption 3. The nonstochastic smoothing parameter $p = p(T) \to \infty$, $p/T \to 0$ as the sample size T goes to infinity.

Assumption 4. The kernel function, $K(x)$, should satisfies: (1) for some $b > 1$, $|K(x)| \leq C|x|^b$ as $x \to \infty$; and (2)$|K(x_1) - K(x_2)| \leq C|x_1 - x_2|$, for any $x_1, x_2 \in R$.

Assumption 5. Let $\hat{p}$ be a data-dependent bandwidth and satisfies $\hat{p}/p = 1 + O_p(p^{1+b}/T^{\kappa(1+b)})$ where $\kappa \in (0, 1/2)$. Further, the nonstochastic smoothing parameter p satisfies $p = p(T) \to \infty$, $p/T^\kappa \to 0$.

The assumptions given above are regular and commonly used in financial econometrics (Hong et al., 2007; Li, 2014). Assumption 1 takes into account the existence of volatility clustering and weak serial correlation, which are famous stylized facts in most financial data. Assumptions 2 and 3 are sufficient on the kernel function $K(\cdot)$ and bandwidth p when we use nonstochastic smoothing parameter. Assumption 4 gives some extra conditions on the kernel function when we consider data-dependent bandwidth $\hat{p}$. Assumption 5 rules out the truncated and Daniell kernels (see Andrews, 1991; Newey and West, 1994 for more details on data driven bandwidths).

Proof of Theorem 1. We first show that $\sqrt{T}(\widehat{pred}(c, k) - \widehat{npred}(c, k))$ converge to $\mathcal{N}(0, \Omega(k))$ in distribution. We consider

$$\begin{aligned} &\widehat{pred}(c, k) - \widehat{npred}(c, k) \\ &= \widehat{pred}^+(c, k) + \widehat{pred}^-(c, k) - \widehat{npred}^+(c, k) - \widehat{npred}^-(c, k) \\ &= \frac{1}{T-k} \sum_{t=1}^{T-k} \sum_{i=1}^{m} \lambda_i \{I[R_{t+k}^B > c_i^B] \\ &\quad - pred^+(c_i, k)\} I[R_t^A > c_i^A] \Big/ \left\{ \frac{1}{T-k} \sum_{t=1}^{T-k} I[R_t^A > c_i^A] \right\} \end{aligned}$$

[1] The optimal lag is selected using the Bayesian Information Criterion (BIC).

Table 1
Empirical proportion of rejections (size) for the statistics.

DGP	T	Granger stat		HJ stat		DP stat		J stat		J stat(Q5)		J stat(Q10)		J stat(Q15)	
		5% level	10% level	5% level	10% level	5% level	10% level	5% level	10% level	5% level	10% level	5% level	10% level	5% level	10% level
	250	0.182	0.282	0.035	0.080	0.025	0.066	0.061	0.111	0.084	0.138	0.067	0.121	0.062	0.117
Linear	500	0.171	0.276	0.040	0.086	0.033	0.076	0.055	0.104	0.068	0.125	0.060	0.114	0.057	0.108
model	750	0.162	0.274	0.045	0.094	0.039	0.083	0.054	0.108	0.062	0.116	0.056	0.109	0.058	0.110
	1000	0.165	0.272	0.045	0.098	0.039	0.088	0.056	0.111	0.062	0.114	0.056	0.109	0.054	0.107
SJC	250	0.167	0.299	0.031	0.067	0.016	0.046	0.054	0.115	0.120	0.177	0.067	0.121	0.063	0.123
with	500	0.133	0.289	0.033	0.079	0.017	0.055	0.061	0.101	0.069	0.131	0.065	0.119	0.057	0.112
$\tau_l = 0.05$	750	0.131	0.283	0.041	0.085	0.026	0.065	0.050	0.110	0.067	0.118	0.063	0.109	0.057	0.101
$\tau_u = 0.05$	1000	0.135	0.269	0.033	0.080	0.020	0.065	0.052	0.101	0.057	0.114	0.058	0.110	0.046	0.102
SJC	250	0.172	0.295	0.023	0.070	0.011	0.037	0.054	0.102	0.114	0.158	0.061	0.116	0.055	0.101
with	500	0.148	0.290	0.035	0.081	0.017	0.062	0.053	0.104	0.066	0.122	0.062	0.111	0.057	0.108
$\tau_l = 0.1$	750	0.123	0.272	0.035	0.073	0.026	0.059	0.061	0.109	0.066	0.121	0.055	0.108	0.055	0.106
$\tau_u = 0.05$	1000	0.121	0.248	0.050	0.098	0.036	0.081	0.051	0.096	0.060	0.116	0.049	0.108	0.061	0.104
SJC	250	0.162	0.304	0.030	0.072	0.013	0.046	0.062	0.109	0.120	0.171	0.073	0.129	0.066	0.118
with	500	0.141	0.283	0.036	0.084	0.018	0.055	0.056	0.109	0.076	0.125	0.062	0.114	0.054	0.107
$\tau_l = 0.1$	750	0.136	0.279	0.041	0.088	0.027	0.065	0.054	0.107	0.064	0.114	0.062	0.110	0.056	0.106
$\tau_u = 0.1$	1000	0.129	0.265	0.049	0.098	0.030	0.081	0.055	0.106	0.059	0.105	0.049	0.098	0.053	0.106

Notes: The table shows the size. Q5, Q10 and Q15 mean the quantile of the returns c^A for the cumulative probability 5%, 10% and 15%, respectively. The data are generated from linear model (18) and symmetric Joe-Clayton (SJC) copula model (19) for the nonlinear and asymmetric phenomenon. All results are based on 10,000 Monte Carlo simulation and asymptotic critical values.

Table 2
Empirical proportion of rejections (power) for the statistics.

DGP	T	Granger stat		HJ stat		DP stat		J stat		J stat(Q5)		J stat(Q10)		J stat(Q15)	
		5% level	10% level	5% level	10% level	5% level	10% level	5% level	10% level	5% level	10% level	5% level	10% level	5% level	10% level
	250	0.498	0.625	0.063	0.128	0.054	0.129	0.189	0.284	0.184	0.267	0.203	0.293	0.226	0.321
Linear	500	0.705	0.802	0.091	0.169	0.096	0.187	0.305	0.422	0.248	0.351	0.318	0.434	0.346	0.464
model	750	0.843	0.912	0.105	0.193	0.123	0.225	0.425	0.546	0.329	0.442	0.433	0.549	0.481	0.599
	1000	0.924	0.961	0.119	0.212	0.151	0.262	0.537	0.661	0.415	0.534	0.535	0.653	0.604	0.718
SJC	250	0.879	0.941	0.275	0.433	0.209	0.388	0.462	0.588	0.421	0.522	0.519	0.630	0.571	0.684
with	500	0.993	0.998	0.556	0.704	0.531	0.719	0.728	0.825	0.644	0.742	0.776	0.860	0.830	0.885
$\tau_l = 0.05$	750	1.000	1.000	0.720	0.830	0.729	0.859	0.895	0.940	0.775	0.856	0.900	0.944	0.940	0.970
$\tau_u = 0.05$	1000	1.000	1.000	0.846	0.929	0.881	0.948	0.949	0.973	0.886	0.928	0.963	0.980	0.980	0.990
SJC	250	0.948	0.974	0.383	0.550	0.320	0.525	0.586	0.699	0.538	0.639	0.638	0.743	0.687	0.784
with	500	0.997	0.999	0.746	0.866	0.724	0.866	0.871	0.925	0.756	0.825	0.888	0.936	0.929	0.964
$\tau_l = 0.1$	750	1.000	1.000	0.899	0.959	0.908	0.959	0.966	0.980	0.902	0.946	0.967	0.982	0.983	0.990
$\tau_l = 0.05$	1000	1.000	1.000	0.962	0.987	0.975	0.991	0.990	0.995	0.961	0.983	0.991	0.996	0.995	0.998
SJC	250	0.983	0.993	0.533	0.704	0.477	0.674	0.712	0.807	0.624	0.718	0.740	0.825	0.800	0.871
with	500	1.000	1.000	0.869	0.941	0.862	0.943	0.940	0.972	0.861	0.909	0.949	0.974	0.972	0.987
$\tau_l = 0.1$	750	1.000	1.000	0.973	0.990	0.975	0.993	0.990	0.996	0.955	0.975	0.994	0.997	0.998	0.999
$\tau_u = 0.1$	1000	1.000	1.000	0.996	0.999	0.998	0.999	0.997	0.997	0.986	0.996	0.998	0.999	1.000	1.000

Notes: The table shows the power. Q5, Q10 and Q15 mean the quantile of the returns c^A for the cumulative probability 5%, 10% and 15%, respectively. The data are collected by reversing the roles of $\{R_t^A\}$ and $\{R_t^B\}$ in Table 1. All results are based on 10,000 Monte Carlo simulation and asymptotic critical values.

$$
\begin{aligned}
&+\frac{1}{T-k}\sum_{t=1}^{T-k}\sum_{i=1}^{m}\lambda_i\{I[R_{t+k}^B<-c_i^B]\\
&-pred^-(c_i,k)\}I[R_t^A<-c_i^A]\Big/\left\{\frac{1}{T-k}\sum_{t=1}^{T-k}I[R_t^A<-c_i^A]\right\}\\
&-\frac{1}{T-k}\sum_{t=1}^{T-k}\sum_{i=1}^{m}\lambda_i\{I[R_{t+k}^B<c_i^B]\\
&-npred^+(c_i,k)\}I[R_t^A>c_i^A]\Big/\left\{\frac{1}{T-k}\sum_{t=1}^{T-k}I[R_t^A>c_i^A]\right\}\\
&-\frac{1}{T-k}\sum_{t=1}^{T-k}\sum_{i=1}^{m}\lambda_i\{I[R_{t+k}^B>-c_i^B]\\
&-npred^-(c_i,k)\}I[R_t^A<-c_i^A]\Big/\left\{\frac{1}{T-k}\sum_{t=1}^{T-k}I[R_t^A<-c_i^A]\right\}\\
&+\sum_{i=1}^{m}\lambda_i(pred(c_i,k)-npred(c_i,k))\\
=&\frac{1}{T-k}\sum_{t=1}^{T-k}\sum_{i=1}^{m}\lambda_i(\eta_t^{p+}(c_i,k)+\eta_t^{p-}(c_i,k)\\
&-\eta_t^{np+}(c_i,k)-\eta_t^{np-}(c_i,k))\\
&+\frac{1}{T-k}\sum_{t=1}^{T-k}\sum_{i=1}^{m}\lambda_i\eta_t^{p+}(c_i,k)\left(\frac{Pr(R_t^A>c_i^A)}{\frac{1}{T-k}\sum_{t=1}^{T-k}I[R_t^A>c_i^A]}-1\right)\\
&+\frac{1}{T-k}\sum_{t=1}^{T-k}\sum_{i=1}^{m}\lambda_i\eta_t^{p-}(c_i,k)\left(\frac{Pr(R_t^A<-c_i^A)}{\frac{1}{T-k}\sum_{t=1}^{T-k}I[R_t^A<-c_i^A]}-1\right)\\
&-\frac{1}{T-k}\sum_{t=1}^{T-k}\sum_{i=1}^{m}\lambda_i\eta_t^{np+}(c_i,k)\left(\frac{Pr(R_t^A>c_i^A)}{\frac{1}{T-k}\sum_{t=1}^{T-k}I[R_t^A>c_i^A]}-1\right)\\
&-\frac{1}{T-k}\sum_{t=1}^{T-k}\sum_{i=1}^{m}\lambda_i\eta_t^{np-}(c_i,k)\left(\frac{Pr(R_t^A<-c_i^A)}{\frac{1}{T-k}\sum_{t=1}^{T-k}I[R_t^A<-c_i^A]}-1\right)\\
&+\sum_{i=1}^{m}\lambda_i(pred(c_i)-npred(c_i))\\
=&\frac{1}{T-k}\sum_{t=1}^{T-k}\sum_{i=1}^{m}\lambda_i(\eta_t^{p+}(c_i,k)+\eta_t^{p-}(c_i,k)\\
&-\eta_t^{np+}(c_i,k)-\eta_t^{np-}(c_i,k))\\
&+o_p\Bigg(\frac{1}{T-k}\sum_{t=1}^{T-k}\sum_{i=1}^{m}\lambda_i(\eta_t^{p+}(c_i,k)+\eta_t^{p-}(c_i,k)\\
&-\eta_t^{np+}(c_i,k)-\eta_t^{np-}(c_i,k))\Bigg)\\
=&\frac{1}{T-k}\sum_{t=1}^{T-k}\eta_t(k)+o_p\left(\frac{1}{T-k}\sum_{t=1}^{T-k}\eta_t(k)\right),
\end{aligned}
\tag{A.1}
$$

where $\lambda=(\lambda_1,\ldots,\lambda_m)'$ denotes an vector $(m\times 1)$ satisfying $\lambda'\lambda=1$, $\eta_t(k)=\sum_{i=1}^{m}\lambda_i(\eta_t^{p+}(c_i,k)+\eta_t^{p-}(c_i,k)-\eta_t^{np+}(c_i,k)-\eta_t^{np-}(c_i,k))$ and

$$
\begin{aligned}
\eta_t^{p+}(c_i,k)&=\frac{\left(I[R_{t+k}^B>c_i^B]-pred^+(c_i)\right)I[R_t^A>c_i^A]}{Pr(R_t^A>c_i^A)}\\
\eta_t^{p-}(c_i,k)&=\frac{\left(I[R_{t+k}^B<-c_i^B]-pred^-(c_i)\right)I[R_t^A<-c_i^A]}{Pr(R_t^A<-c_i^A)}\\
\eta_t^{np+}(c_i,k)&=\frac{\left(I[R_{t+k}^B<-c_i^B]-pred^+(c_i)\right)I[R_t^A>c_i^A]}{Pr(R_t^A>c_i^A)}\\
\eta_t^{np-}(c_i,k)&=\frac{\left(I[R_{t+k}^B>c_i^B]-pred^-(c_i)\right)I[R_t^A<c_i^A]}{Pr(R_t^A<c_i^A)}.
\end{aligned}
$$

In addition, in (A.1) we use the fact that under null hypothesis, $pred(c,k)=npred(c,k)$, we know that to prove $\sqrt{T}(\widehat{pred}(c,k)-\widehat{npred}(c,k))\xrightarrow{d}\mathcal{N}(0,\Omega(k))$, we consider

$$
\begin{aligned}
&\sqrt{T}(\widehat{pred}(c,k)-\widehat{npred}(c,k))\\
&\quad=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_t(k)+o_p\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_t(k)\right).
\end{aligned}
\tag{A.2}
$$

We have

$$
\Omega(k)=\lim_{T\to\infty}Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_t(k)\right]=\sum_{l=-\infty}^{\infty}Cov(\eta_t(k),\eta_{t-l}(k)).\tag{A.3}
$$

By the central limit theorem for mixing processes (White, 2014), we have

$$
\frac{\sqrt{T}\lambda'(\widehat{pred}(c,k)-\widehat{npred}(c,k))}{\sqrt{\lambda'\Omega(k)\lambda}}\xrightarrow{d}\mathcal{N}(0,1).\tag{A.4}
$$

And hence

$$
\begin{aligned}
&T(\widehat{pred}(c,k)-\widehat{npred}(c,k))'\Omega^{-1}(k)(\widehat{pred}(c,k)\\
&\quad-\widehat{npred}(c,k))\xrightarrow{d}\chi^2(m).
\end{aligned}
\tag{A.5}
$$

Next, we show that $\hat{\Omega}(k)\xrightarrow{p}\Omega(k)$. Write $\hat{\Omega}(k)-\Omega(k)=\left(\hat{\Omega}(k)-E[\hat{\Omega}(k)]\right)+\left(E[\hat{\Omega}(k)]-\Omega(k)\right)$. By the Andrews (1991) Lemma 1, Assumption A.2 implies that Assumption A of Andrews (1991) holds. It follows the Andrews (1991) Proposition 1(a) that $Var[\hat{\Omega}(k)]=E[(\hat{\Omega}(k)-E[\hat{\Omega}(k)])^2]=O(p/T)$. Therefore we have $\hat{\Omega}(k)-E[\Omega(k)]=O_p(\sqrt{p/T})$ by Chebyshev's inequality. In addition, because Assumption A.2(ii) implies $\sum_{j=-\infty}^{\infty}\Omega(j,k)\le C$, and because of Assumption A.4 and dominated convergence, we have

$$
\begin{aligned}
E[\hat{\Omega}(k)]-\Omega(k)=&\sum_{j=1-T}^{T-1}[(1-|j|/T)k(j/p)-1]\Omega(j,k)\\
&+\sum_{|j|>T}\Omega(j,k)\to 0,
\end{aligned}
\tag{A.6}
$$

as $T\to\infty$. Consequently, $\hat{\Omega}(k)\xrightarrow{p}\Omega(k)$. By Slutsky's theorem, we then obtain

$$
\begin{aligned}
J(k)=&\,T\left(\widehat{pred}(c,k)-\widehat{npred}(c,k)\right)'\hat{\Omega}^{-1}(k)\left(\widehat{pred}(c,k)\right.\\
&\left.-\widehat{npred}(c,k)\right)\xrightarrow{d}\chi^2(m).
\end{aligned}
\tag{A.7}
$$

Suppose that the bandwidth parameter p is a function of the data, which is expressed as $\hat{p}$, and $\hat{\Omega}^*(k)$ is the kernel estimator of $\Omega(k)$.

$$
\begin{aligned}
\hat{\Omega}_{ij}^*(k)-\hat{\Omega}_{ij}(k)&=\sum_{l=1-T}^{T-1}[k(l/\hat{p})-k(l/p)]\hat{\delta}_l(c_i,c_j,k)\\
&=\sum_{|l|\le q}[k(l/\hat{p})-k(l/p)]\hat{\delta}_l(c_i,c_j,k)\\
&\quad+\sum_{q<|l|<T}[k(l/\hat{p})-k(l/p)]\hat{\delta}_l(c_i,c_j,k)\\
&=I_1(c_i,c_j,k)+I_2(c_i,c_j,k),
\end{aligned}
\tag{A.8}
$$

where $q=T^{\kappa}$ for κ.

We now consider the first term $I_1(c_i,c_j,k)$. Using the Assumptions and the triangle inequality, we have

$$
\begin{aligned}
|I_1(c_i, c_j, k)| &\le \sum_{|l|\le q} C|l/\hat{p} - l/p| \cdot |\hat{\delta}_l(c_i, c_j, k)| \\
&\le C|\hat{p}^{-1} - p^{-1}|q \sum_{|l|\le q} |\hat{\delta}_l(c_i, c_j, k) - \delta_l(c_i, c_j, k)| \\
&\quad + C|\hat{p}^{-1} - p^{-1}|q \sum_{|l|\le q} |\delta_l(c_i, c_j, k)| \\
&= |\hat{p}^{-1} - p^{-1}| O_p(q/T^{1/2} + q) \\
&= O_p(q \cdot |\hat{p}^{-1} - p^{-1}|), \qquad \text{(A.9)}
\end{aligned}
$$

where $\sum_{l=-\infty}^{\infty} |\delta_l(c_i, c_j, k)| \le C$ and $\sup_{0<l<T} E[\hat{\delta}_l(c_i, c_j, k) - \delta_l(c_i, c_j, k)]^2 = O(T^{-1})$ from Hannan (2009).

For the second term $I_2(c_i, c_j, k)$, using the Assumptions, we have

$$
\begin{aligned}
|I_2(c_i, c_j, k)| &\le \sum_{q<|l|<T} C(|l/\hat{p}|^{-b} + |l/p|^{-b}) \hat{\delta}(c_i, c_j, k) \\
&\le C(\hat{p}^b + b^b) q^{-b} \sum_{q<|l|<T} |\hat{\delta}_l(c_i, c_j, k) - \delta_l(c_i, c_j, k)| \\
&\quad + C(\hat{p}^b + b^b) q^{-b} \sum_{q<|l|<T} |\delta_l(c_i, c_j, k)| \\
&= C(\hat{p}^b + b^b) q^{-b} [O_p(q/T^{1/2}) + o_p(1)], \qquad \text{(A.10)}
\end{aligned}
$$

where again $\sum_{l=-\infty}^{\infty} |\delta_l(c_i, c_j, k)| \le C$ and $\sup_{0<l<T} E[\hat{\delta}_l(c_i, c_j, k) - \delta_l(c_i, c_j, k)]^2 = O(T^{-1})$.

Therefore, we have $\hat{\Omega}^*(k) - \hat{\Omega}(k) = o_p(1)$.

References

Andrews, D.W., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 817–858.

Ang, A., Chen, J., 2002. Asymmetric correlations of equity portfolios. J. Financ. Econ. 63 (3), 443–494.

Diks, C., Panchenko, V., 2006. A new statistic and practical guidelines for nonparametric granger causality testing. J. Econom. Dynam. Control 30 (9), 1647–1669.

Granger, C.W., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 424–438.

Hannan, E.J., 2009. Multiple Time Series. Vol. 38. John Wiley & Sons.

Hiemstra, C., Jones, J.D., 1994. Testing for linear and nonlinear granger causality in the stock price-volume relation. J. Finance 49 (5), 1639–1664.

Hong, Y., Tu, J., Zhou, G., 2007. Asymmetries in stock returns: statistical tests and economic evaluation. Rev. Financ. Stud. 20 (5), 1547–1581.

Li, F., 2014. Identifying asymmetric comovements of international stock market returns. J. Financ. Econ. 12 (3), 507–543.

Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Rev. Econ. Stud. 61 (4), 631–653.

Patton, A.J., 2006. Modelling asymmetric exchange rate dependence*. Int. Econom. Rev. 47 (2), 527–556.

White, H., 2014. Asymptotic Theory for Econometricians. Academic Press.