Journal of Statistical Planning and Inference 148 (2014) 82–94

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Factor double autoregressive models with application to simultaneous causality testing

Shaojun Guo [a], Shiqing Ling [b], Ke Zhu [a,*]

[a] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, HaiDian District, Beijing, China
[b] Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

ARTICLE INFO

Article history:
Received 19 October 2012
Received in revised form
4 September 2013
Accepted 9 December 2013
Available online 19 December 2013

Keywords:
Asymptotic normality
Causality-in-mean
Causality-in-variance
Factor DAR model
Instantaneous causality
Score test
Strong consistency

ABSTRACT

Testing causality-in-mean and causality-in-variance has been largely studied. However, none of the tests can detect causality-in-mean and causality-in-variance simultaneously. In this paper, we introduce a factor double autoregressive (FDAR) model. Based on this model, a score test is proposed to detect causality-in-mean and causality-in-variance simultaneously. Furthermore, strong consistency and asymptotic normality of the quasi-maximum likelihood estimator (QMLE) for the FDAR model are established. A small simulation study shows good performances of the QMLE and the score test in finite samples. A real data example on the causal relationship between Hong Kong stock market and US stock market is given.

1. Introduction

Since the seminal work of Granger (1969), the Granger causality test has been broadly used in finance and economics. Principally, it tells us whether the past information of some specified series can improve the prediction of the current and future values of the other series. The study of causality is of theoretical interest; see, e.g., Geweke (1984a) and Gouriéroux and Monfort (1997) for earlier works and Nishiyama et al. (2011) and the references therein for more recent ones. In practice, the causality-in-mean has been widely identified between many macroeconomic variables, e.g., Sims (1972, 1980), Geske and Roll (1983), Ram and Spencer (1983), Stock and Watson (1989), and Lee (1992) to name a few. Recently, the nonlinear causality has received more attention. As a special case of the nonlinear causality, the causality-in-variance becomes particularly essential, because it manifests the volatility spillover across different assets or markets; see, e.g., Baillie and Bollerslev (1990), Engle et al. (1990), Hamao et al. (1990), Ng (2000), and Hong (2001). For more discussions on the explanation of causality-in-variance, we refer to Ross (1989) and Hong (2001).

Testing causality-in-mean and causality-in-variance has been largely but separately studied. For the causality-in-mean, Granger (1969) constructed a F-test based on the regression; Geweke (1982, 1984b) measured the linear dependence including causality-in-mean for the multiple time series; Boudjellaba et al. (1992) gave a testing procedure for the vector ARMA model; and many others. For the causality-in-variance, Cheung and Ng (1996) proposed a residual cross-correlation function test (CCF test); Hong (2001) modified the CCF test by adding the weight function; Hafner and Herwartz (2006) gave

* Corresponding author.
E-mail addresses: guoshaoj@amss.ac.cn (S. Guo), maling@ust.hk (S. Ling), kzhu@amss.ac.cn (K. Zhu).

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.12.007

a Wald test for the multivariate GARCH model; see also Hiemstra and Jones (1994) and Nishiyama et al. (2011) for other nonlinear tests. However, none of the tests aforementioned can detect causality-in-mean and causality-in-variance simultaneously. The empirical studies have demonstrated that these two causality patterns may co-exist; see, e.g., Hamao et al. (1990), Cheung and Ng (1996), and Ng (2000). Pantelidis and Pittis (2004) showed that without filtering out causality-in-mean, the test for causality-in-variance could suffer severe size distortions in the present of causality-in-mean. Therefore, it urges us to develop a tool to detect them simultaneously.

In this paper, we introduce a factor double autoregressive (hereafter FDAR) model. This causal model not only includes Granger's linear causality model as a special case, but also characterizes the causality-in-variance. An extended FDAR model is also presented to capture the instantaneous causality-in-mean and causality-in-variance together. We next propose a score test to detect causality-in-mean and causality-in-variance simultaneously. In the presence of both causalities, we propose a quasi-maximum likelihood approach to estimate the parameters in the FDAR model. Under regularity conditions, strong consistency and asymptotic normality of the quasi-maximum likelihood estimator (QMLE) for the FDAR model are obtained. On the basis of this FDAR model, we analyze the causal relationship between Hong Kong stock market and US stock market. The results find evidence that US stock market affects HK stock market largely in both mean and variance of returns, while the impact of HK stock market to US stock market is relatively weak. This is consistent with our sense, since US market is the largest capital market in the world.

The remainder of the paper is organized as follows. In Section 2, we introduce the FDAR model and give a sufficient and necessary condition for testing causality-in-mean and causality-in-variance. In Section 3, we propose a score test to detect causality-in-mean and causality-in-variance, simultaneously. The asymptotic properties of the QMLE for the FDAR model are studied in Section 4. A simulation study is carried out in Section 5 to examine the performances of the score test and the QMLE in finite samples. A real example is offered in Section 6. All of the proofs are provided in the Appendix.

2. The causal model

Suppose that we observe two series x_t and y_t and consider how y_t causes x_t. Let $\mathcal{I}_{1,t}$ and $\mathcal{I}_{2,t}$ be σ-fields of $\{x_t\}$ and $\{y_t\}$ available at period t, respectively. Denote $\mathcal{I}_t=\sigma(\mathcal{I}_{1,t},\mathcal{I}_{2,t})$. Following Granger (1969), y_t is said to cause x_t in mean if

$$P\{E(x_t|\mathcal{I}_{1,t-1})\neq E(x_t|\mathcal{I}_{t-1})\}>0. \tag{2.1}$$

Next, following Granger et al. (1986), y_t is said to cause x_t in variance if

$$P\{E\{[x_t-E(x_t|\mathcal{I}_{t-1})]^2|\mathcal{I}_{1,t-1}\}\neq E\{[x_t-E(x_t|\mathcal{I}_{t-1})]^2|\mathcal{I}_{t-1}\}\}>0. \tag{2.2}$$

The causality-in-mean, as a special case of linear causality, is often called the first order causality; see Nishiyama et al. (2011). The causality-in-variance is a kind of the nonlinear causality defined by Hiemstra and Jones (1994). Both of them are also two special cases of general causalities defined by Granger (1980). It is easy to see that any of (2.1) and (2.2) holds if and only if

$$P\{E\{[x_t-E(x_t|\mathcal{I}_{1t-1})]^2|\mathcal{I}_{1t-1}\}\neq E\{[x_t-E(x_t|\mathcal{I}_{t-1})]^2|\mathcal{I}_{t-1}\}\}>0. \tag{2.3}$$

Thus, testing (2.1)–(2.2) altogether is equivalent to testing (2.3). See also Comte and Lieberman (2000). However, without any other information, (2.3) can hardly be testable. For instance, it may cause the curse of dimensionality if the conditional expectation $E(x_t|\mathcal{I}_{t-1})$ is estimated nonparametrically.

To make (2.3) easily testable, a natural approach is to specify a meaningful causal relationship between x_t and y_t. In this paper, we assume that given $\{(x_s,y_s),s<t\}$, x_t's are generated from the following model:

$$x_t=\phi_0+\sum_{i=1}^{p}\phi_i x_{t-i}+\sum_{i=1}^{q}\psi_i y_{t-i}+\eta_t\sqrt{\alpha_0+\sum_{i=1}^{p}\alpha_i x_{t-i}^2+\sum_{i=1}^{q}\beta_i y_{t-i}^2}, \tag{2.4}$$

where all α_i and β_i are non-negative constant parameters, $\{\eta_t\}$ is a sequence of *i.i.d.* random variables with zero mean and unit variance and η_t is independent of $\mathcal{I}_{t-1}$ for each $t\geq 1$. We call model (2.4) as the factor double autoregressive (FDAR) model. When all α_i and β_i are zeros, it reduces to Granger's linear causal model. When the factor y_t is absent, it reduces to the DAR model in Weiss (1986) and Ling (2004, 2007), and furthermore, it reduces to the ARCH model in Engle (1982) if all ϕ_i's are zeros. Throughout the paper, we assume that (x_t,y_t) are stationary and ergodic.

Since our main goal here is to detect how y_t causes x_t, we do not specify the generation mechanism of y_t, whether or not dependent of x_t, only assuming that y_t is stationary and ergodic. Of course, the series y_t can be modeled in practice. In the end of this section, we give a remark of how to model y_t. In simulation studies, we choose three generation mechanisms of y_t, showing that all the procedures proposed in Sections 3 and 4 work well.

Based on model (2.4) and Assumption 2.1 below, an equivalent but testable condition for (2.3) is derived.

Assumption 2.1. (i) $y_{t-i}\notin\sigma(\mathcal{I}_{1,t-1},\mathcal{I}_{2,t-i-1})$ for any $i\geq 1$; (ii) $E|\eta_t|^2<\infty$, $E|x_t|^2<\infty$ and $E|y_t|^2<\infty$.

We now give our first proposition, which presents a sufficient and necessary condition for testing (2.3) under model (2.4).

Proposition 2.1. *Suppose that Assumption 2.1 holds. Then, the inequality (2.3) holds if and only if some ψ_i or β_i is not zero. Particularly, (2.1) holds if and only if some ψ_i is not zero; and (2.2) holds if and only if some β_i is not zero.*

Proof. See Appendix A.

Although model (2.4) captures the causality-in-mean and causality-in-variance simultaneously from y_t to x_t, it is often meaningful to describe the instantaneous causality-in-mean and causality-in-variance between x_t and y_t. Motivated by this, we proceed to consider the following extended FDAR model:

$$x_t=\phi_0+\sum_{i=1}^{p}\phi_i x_{t-i}+\sum_{i=0}^{q}\psi_i y_{t-i}+\eta_t\sqrt{\alpha_0+\sum_{i=1}^{p}\alpha_i x_{t-i}^2+\sum_{i=0}^{q}\beta_i y_{t-i}^2}, \tag{2.5}$$

where all α_i, β_i, and $\{\eta_t\}$ are defined as in model (2.4) except that η_t is independent of $\sigma(\mathcal{I}_{1,t-1},\mathcal{I}_{2,t})$. Clearly, the extended FDAR model reduces to FDAR model when $\psi_0=\beta_0=0$. As in Hong (2001), we say that there is an instantaneous causality-in-mean between x_t and y_t if

$$P\{E(x_t|\mathcal{I}_{t-1})\neq E(x_t|\mathcal{I}_{1,t-1},\mathcal{I}_{2,t})\}>0 \tag{2.6}$$

and an instantaneous causality-in-variance between x_t and y_t if

$$P\{E\{[x_t-E(x_t|\mathcal{I}_{1,t-1},\mathcal{I}_{2,t})]^2|\mathcal{I}_{t-1}\}\neq E\{[x_t-E(x_t|\mathcal{I}_{1,t-1},\mathcal{I}_{2,t})]^2|\mathcal{I}_{1,t-1},\mathcal{I}_{2,t}\}\}>0. \tag{2.7}$$

Analogous to Proposition 2.1, our second proposition below gives a sufficient and necessary condition for testing (2.6) and (2.7) under model (2.5) and Assumption 2.2.

Assumption 2.2. (i) $y_t\notin\mathcal{I}_{t-1}$; (ii) $E|\eta_t|^2<\infty$, $E|x_t|^2<\infty$ and $E|y_t|^2<\infty$.

Proposition 2.2. *Suppose that Assumption 2.2 holds. Then, relation (2.6) holds if and only if $\psi_0\neq 0$; and relation (2.7) holds if and only if $\beta_0\neq 0$.*

Proof. The proof is directly from Assumption 2.2 and hence omitted. □

Till now, we have not restricted the specification of y_t. Although not being necessary, it is also worthwhile to model y_t by an extended FDAR model in practice, especially when y_t exhibits the conditional heteroskedasticity. That is, we consider another extended FDAR model for y_t:

$$y_t=\pi_0+\sum_{i=0}^{r}\pi_i x_{t-i}+\sum_{i=1}^{s}\omega_i y_{t-i}+\zeta_t\sqrt{\tau_0+\sum_{i=0}^{r}\tau_i x_{t-i}^2+\sum_{i=1}^{s}\nu_i y_{t-i}^2}. \tag{2.8}$$

Likewise, model (2.8) shares the same property as model (2.5). In what follows, we call models (2.5) and (2.8) as the bivariate extended FDAR model. Based on this bivariate extended FDAR model, Assumptions 2.1–2.2 hold if η_t and ζ_t are independent. Intuitively, if η_t and ζ_t are dependent, there may exist either a common factor z_t affecting both x_t and y_t or some other nonlinear causal relation besides the causality-in-variance between x_t and y_t. In this case, we suggest to use a multivariate extended FDAR model to deal with the problem of common factors. If η_t and ζ_t remain dependent after filtering out the impact of common factors, a further nonlinear test in Hiemstra and Jones (1994) can be implemented to detect whether there are some other nonlinear causal relations besides the causality-in-variance.

3. Simultaneous causality test

In this section, we propose a score test to simultaneously detect causality-in-mean and causality-in-variance from y_t to x_t under model (2.4). We first assume that both p and q are known. In the end of this section, the case that p and q are unknown is discussed. Let $\theta=(\phi',\psi',\alpha',\beta')'$ be the unknown parameters of model (2.4), where $\phi=(\phi_0,\ldots,\phi_p)'$, $\alpha=(\alpha_0,\ldots,\alpha_p)'$, $\psi=(\psi_1,\ldots,\psi_q)'$, and $\beta=(\beta_1,\ldots,\beta_q)'$. According to Proposition 2.1, we would like to test the hypotheses:

$$H_0:\psi\equiv\beta\equiv 0. \tag{3.1}$$

Given the observations $\{(x_t,y_t)\}_{t=1}^n$, we denote $X_t=(1,x_{t-1},\ldots,x_{t-p})'$, $X_t^*=(1,x_{t-1}^2,\ldots,x_{t-p}^2)'$, $Y_t=(y_{t-1},\ldots,y_{t-q})'$ and $Y_t^*=(y_{t-1}^2,\ldots,y_{t-q}^2)'$. By assuming that η_t follows standard normal distribution, the quasi-log-likelihood function (ignoring a constant) of model (2.4) is

$$L_n(\theta)=-\frac{1}{n}\sum_{t=m}^{n}l_t(\theta)\quad\text{and}\quad l_t(\theta)=\log\sqrt{h_t(\theta)}+\frac{\varepsilon_t^2(\theta)}{2h_t(\theta)}, \tag{3.2}$$

where $m=1+\max(p,q)$, $\varepsilon_t(\theta)=x_t-\phi'X_t-\psi'Y_t$ and $h_t(\theta)=\alpha'X_t^*+\beta'Y_t^*$. Here, $h_t(\theta)$ is the conditional variance of x_t, given $\mathcal{I}_{t-1}$.

Under H_0, model (2.4) becomes a DAR(p) model with parameters $(\phi',\alpha')'$. Denote $\Theta_1\triangleq\Theta_\phi\times\Theta_\alpha$ be the parameter space of this DAR(p) model. Let $\overline{\theta}_{10}\triangleq(\overline{\phi}_0',\overline{\alpha}_0')'$ be the true value of $(\phi',\alpha')'\in\Theta_1$. As in Ling (2007), the quasi-maximum likelihood estimator (QMLE) $\hat{\theta}_{1n}\triangleq(\hat{\phi}_n',\hat{\alpha}_n')'$ for $\overline{\theta}_{10}$ is obtained by maximizing $L_n(\theta)$ with respect to $(\phi',\alpha')'\in\Theta_1$ under the constraint that $(\psi',\beta')\equiv 0$. Moreover, let $\hat{\theta}_n=(\hat{\phi}_n',0_{1\times q},\hat{\alpha}_n',0_{1\times q})'$ and

$$T_n(\theta)=\left(\frac{\partial L_n(\theta)}{\partial\psi'},\frac{\partial L_n(\theta)}{\partial\beta'}\right)' \tag{3.3}$$

be the score function for ψ and β. To construct the score statistics, we desire to prove that $T_n(\hat{\theta}_n)$ is asymptotically normal with mean zero under H_0 and regularity conditions. To accomplish it, we need the following two assumptions.

Assumption 3.1. $\overline{\theta}_{10}$ is an interior point in Θ_1, and Θ_1 is compact with $\alpha_i^L \le \alpha_i \le \alpha_i^U$ for all i, where α_i^L and α_i^U are some positive constants.

Assumption 3.2. $E|\eta_t|^4 < \infty$, $E|x_t|^\iota < \infty$ for some $\iota > 0$, and $E|y_t|^4 < \infty$.

To be convenient, we make some notations before the theorem:

$$J = \begin{pmatrix} 1 & -\frac{E\eta_t^3}{\sqrt{2}} \\ -\frac{E\eta_t^3}{\sqrt{2}} & \frac{E\eta_t^4 - 1}{2} \end{pmatrix} \quad \text{and} \quad A_t(\overline{\theta}_{10}) = \operatorname{diag}\left\{ \frac{\Gamma_1 X_t - Y_t}{\sqrt{h_t(\overline{\theta}_0)}}, \frac{Y_t^* - \Gamma_2 X_t^*}{\sqrt{2} h_t(\overline{\theta}_0)} \right\}, \tag{3.4}$$

where $\overline{\theta}_0 = (\overline{\phi}_0', 0_{1\times q}, \overline{\alpha}_0', 0_{1\times q})'$,

$$\Gamma_1 = E\left(\frac{Y_t X_t'}{h_t(\overline{\theta}_0)} \right) \left[E\left(\frac{X_t X_t'}{h_t(\overline{\theta}_0)} \right) \right]^{-1} \quad \text{and} \quad \Gamma_2 = E\left(\frac{Y_t^* X_t^{*\prime}}{h_t^2(\overline{\theta}_0)} \right) \left[E\left(\frac{X_t^* X_t^{*\prime}}{h_t^2(\overline{\theta}_0)} \right) \right]^{-1}.$$

Then, we can give our first main result as follows:

Theorem 3.1. *Suppose that Assumptions 2.1(i) and 3.1–3.2 hold and J is positive definite. Then, under H_0, as $n \to \infty$,*

$$\sqrt{n} T_n(\hat{\theta}_n) \to_d N(0, \Xi),$$

where $\to_d$ denotes the convergence in distribution, $\Xi = E[A_t(\overline{\theta}_{10}) J A_t'(\overline{\theta}_{10})]$, and J and $A_t(\overline{\theta}_{10})$ are defined in (3.4).

Proof. See Appendix B.

It is important to point out that Γ_1 and Γ_2 are both well defined, since the matrixes $E[X_t X_t'/h_t(\overline{\theta}_0)]$ and $E[X_t^* X_t^{*\prime}/h_t^2(\overline{\theta}_0)]$ are positive definite by Lemma B.5 in Ling (2007). Also, it is readily shown that $J > 0$ if and only if $P(\eta_t^2 - c\eta_t - 1 = 0) < 1$ for any $c \in R$. A simple condition for this is that η_t has a positive density on some interval. In particular, when $\eta_t \sim N(0,1)$, J becomes the identity matrix.

In practice, given the observations $\{(x_t, y_t)\}_{t=1}^n$, the matrix Ξ can be consistently estimated by its sample mean $\hat{\Xi}_n$. Under H_0, if the conditions in Theorem 3.1 hold, it is not hard to show that $\hat{\Xi}_n = \Xi + o_p(1)$. Therefore, we construct a score test statistic

$$S_n = n T_n'(\hat{\theta}_n) \hat{\Xi}_n^{-1} T_n(\hat{\theta}_n)$$

to test (3.1). The following corollary gives its asymptotic distribution, as expected.

Corollary 3.1. *Suppose that the conditions in Theorem 3.1 hold. Then, under H_0, as $n \to \infty$,*

$$S_n \to_d \chi_{2q}^2,$$

where χ_k^2 is a chi-square distribution with degree of freedom k.

Proof. The proof is directly from Theorem 3.1, and hence it is omitted. □

Remark 3.1. Based on model (2.5), a score test $S_n^\diamond$, which is similar to S_n, can be used to detect the hypothesis

$$H_0^\diamond : \theta_2^\diamond \equiv 0,$$

where $\theta_2^\diamond = (\psi', \beta', \psi_0, \beta_0)'$. If Assumption 2.2(i) and the conditions in Theorem 3.1 hold, by using the same method as in Corollary 3.1, we can easily show that under $H_0^\diamond$, $S_n^\diamond$ converges to $\chi_{2(1+q)}^2$ as $n \to \infty$.

Indeed, the test statistic S_n always depends on the orders p and q. Without confusion, we shorten the notation $S_n(p,q)$ to S_n for brevity. In practice, both p and q are often unknown, and should be determined before using S_n. This can be done by Akaike's information criterion (AIC). In this case, we propose our testing procedure as follows:

1. Determine the values of p and q by AIC under FDAR model (2.4).
2. Calculate the test statistic S_n and compare it to the upper-tailed critical value of χ_{2q}^2 at an appropriate level.
3. If S_n is larger than the critical value, then the null hypothesis H_0 is rejected. Otherwise, H_0 is not rejected.

Clearly, the above procedure is also applicable to detect $H_0^\diamond$ via replacing S_n and χ_{2q}^2 by $S_n^\diamond$ and $\chi_{2(1+q)}^2$, respectively. In order to accomplish Step 1 aforementioned, it is necessary for us to consider the estimation for the FDAR model. The full study on this topic is given in the next section.

4. The QMLE

In this section, we study the QMLE for model (2.4). Denote $\Theta \triangleq \Theta_\phi \times \Theta_\psi \times \Theta_\alpha \times \Theta_\beta$ be the parameter space of model (2.4), where $\Theta_\phi \subset R^{1+p}$, $\Theta_\psi \subset R^q$, $\Theta_\alpha \subset R_+^{1+p}$ and $\Theta_\beta \subset R_+^q$ with $R_+ = [0,\infty)$. Let $\theta_0 \triangleq (\overline{\phi}_0', \overline{\psi}_0', \overline{\alpha}_0', \overline{\beta}_0')'$ be the true value of $\theta \in \Theta$, and $\tilde{\theta}_n \triangleq (\tilde{\gamma}_n', \tilde{\delta}_n')'$ be the maximizer of $L_n(\theta)$ in Θ, i.e.,

$$\tilde{\theta}_n = \arg\max_{\theta \in \Theta} L_n(\theta). \tag{4.1}$$

where $L_n(\theta)$ is defined in (3.2), $\tilde{\gamma}_n = (\tilde{\phi}_n', \tilde{\psi}_n')'$, and $\tilde{\delta}_n = (\tilde{\alpha}_n', \tilde{\beta}_n')'$. We call $\tilde{\theta}_n$ be the QMLE of θ_0. To derive the asymptotic property of $\tilde{\theta}_n$, we make the following assumptions:

Assumption 4.1. The true value θ_0 is an interior point in Θ, and Θ is compact with $\alpha_i^L \le \alpha_i \le \alpha_i^U$ and $\beta_j^L \le \beta_j \le \beta_j^U$ for all i and j, where α_i^L, α_i^U, β_j^L and β_j^U are some positive constants.

Assumption 4.2. $E|x_t|^\iota < \infty$ and $E|y_t|^\iota < \infty$ for some $\iota > 0$.

Assumptions 4.1–4.2 are analogous to Assumptions 3.1–3.2 except that only the fractional moment of y_t is required. This is because the conditional variance $h_t(\theta)$ itself as one sort of weight can control the log-likelihood function (3.2). When y_t is absent and $p=1$ (i.e., DAR(1) model), Borkovec and Klüppelberg (2001) showed that the condition $E(\ln|\phi+\eta_t\sqrt{\alpha}|) < 0$ is sufficient for the stationarity of x_t. Note that this condition does not rule out the case that $|\phi| \ge 1$. Hence, it implies that the stationary region of DAR(1) model is larger than that of AR(1) model; see Ling (2004, 2007) for more discussions on it.

We are now ready to give our second main result as follows:

Theorem 4.1. *Suppose that Assumptions 2.1(i) and 4.1–4.2 hold, $E\eta_t^4 < \infty$ and J is positive definite. Then, as $n \to \infty$,*

(i) $\tilde{\theta}_n \to \theta_0$ *a.s.*;

(ii) $\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_d \mathrm{N}(0, \Omega_0^{-1}\Sigma_0\Omega_0^{-1})$,

where $\Omega_0 = E[B_t(\theta_0)B_t'(\theta_0)]$, $\Sigma_0 = E[B_t(\theta_0)JB_t'(\theta_0)]$, and

$$B_t(\theta) = \left(\frac{1}{\sqrt{h_t(\theta)}} \frac{\partial \varepsilon_t(\theta)}{\partial \theta'}, \frac{1}{\sqrt{2}h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \theta'} \right)'.$$

Proof. See Appendix B.

Remark 4.1. Similar to (4.1), we can define the QMLE $\theta_n^\diamond$ of $\theta_0^\diamond$ for model (2.5), where $\theta_0^\diamond = (\theta_0', \psi_{00}, \beta_{00})'$ is the true value of model (2.5). If Assumption 2.2(i) and the conditions in Theorem 4.1 hold, by using the similar method as for Theorem 4.1, the strong consistency and asymptotic normality of $\theta_n^\diamond$ can be obtained as well.

By a direct calculation, we can see that

$$\frac{\partial \varepsilon_t(\theta)}{\partial \theta} = (-K_t', 0_{1\times(p+q)})' \quad \text{and} \quad \frac{\partial h_t(\theta)}{\partial \theta} = (0_{1\times(1+p+q)}, K_t^{*\prime})',$$

where $K_t = (X_t', Y_t')'$ and $K_t^* = (X_t^{*\prime}, Y_t^{*\prime})'$. Thus we can show that $\Omega_0 > 0$ and $\Sigma_0 > 0$ if $J > 0$ and Assumption 2.1 holds. When y_t is absent, the asymptotic variance in Theorem 4.1 is the same as the one for the DAR(p) models in Ling (2007). Furthermore, if $E\eta_t^3 = 0$, then $\Omega_0^{-1}\Sigma_0\Omega_0^{-1}$ reduces to a block diagonal matrix

$$\mathrm{diag}\left\{ \left[E\left(\frac{1}{h_t(\theta_0)} K_t K_t' \right) \right]^{-1}, \kappa \cdot \left[E\left(\frac{1}{2h_t^2(\theta_0)} K_t^* K_t^{*\prime} \right) \right]^{-1} \right\},$$

with $\kappa = (E\eta_t^4 - 1)/2$.

In the end, we proceed to discuss the diagnostic checking of model (2.4). Denote $\hat{\eta}_t$ be the residual of model (2.4). A portmanteau test $Q^2(M)$ defined in the same way as the Li–Mak test can be used to test the independence of $\{\eta_t\}$. If $\{\eta_t\}$ is independent, by a similar method as in Li and Mak (1994), we can show that $Q^2(M) \to_d \chi_M^2$ as $n \to \infty$. Therefore, model (2.4) is not adequate if $Q^2(M)$ is larger than the upper-tailed critical value of χ_M^2 at an appropriate level. Moreover, if we further consider a bivariate extended FDAR model, the test statistic $C(M) \triangleq n\sum_{i=-M}^M \hat{r}_i^2$ defined in the same way as the CCF test can be used to detect the independence of $\{\eta_t\}$ and $\{\zeta_t\}$, where $\hat{r}_i$ is the sample cross-correlation of the squared residuals $\{\hat{\eta}_t^2\}$ and $\{\hat{\zeta}_t^2\}$ at lag i. If $\{\eta_t\}$ and $\{\zeta_t\}$ are independent, by a similar method as in Cheung and Ng (1996), it is not hard to show that $C(M) \to_d \chi_{2M+1}^2$ as $n \to \infty$. Hence, we reject the hypothesis that $\{\eta_t\}$ and $\{\zeta_t\}$ are independent, if $C(M)$ is larger than the upper-tailed critical value of χ_{2M+1}^2 at an appropriate level.

5. Simulation

In this section, we first give a simulation study to assess the performance of $\tilde{\theta}_n$ in finite samples. The model used to generate data samples is

$$x_t = \phi_0 + \phi_1 x_{t-1} + \psi_1 y_{t-1} + \eta_t \sqrt{\alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 y_{t-1}^2}, \tag{5.1}$$

Table 1
Estimators for model (5.1) when $\{y_t\}$ is generated from model (a).

ϕ_0	ϕ_1	ψ_1	α_0	α_1	β_1		$\tilde{\phi}_{0n}$	$\tilde{\phi}_{1n}$	$\tilde{\psi}_{1n}$	$\tilde{\alpha}_{0n}$	$\tilde{\alpha}_{1n}$	$\tilde{\beta}_{1n}$
0.0	0.5	0.5	1.0	0.5	0.5	Bias	−0.0012	−0.0027	−0.0009	0.0091	−0.0063	−0.0026
						SD	0.0484	0.0351	0.0549	0.1144	0.0505	0.1023
						AD	0.0487	0.0356	0.0527	0.1119	0.0489	0.0984
0.0	0.0	−0.3	1.0	0.5	0.5	Bias	−0.0019	−0.0017	−0.0010	0.0028	−0.0049	0.0012
						SD	0.0466	0.0398	0.0511	0.1072	0.0579	0.0999
						AD	0.0462	0.0390	0.0501	0.1059	0.0585	0.0926
0.0	0.6	0.0	1.0	0.6	0.3	Bias	−0.0017	−0.0044	−0.0013	−0.0008	−0.0038	−0.0020
						SD	0.0461	0.0375	0.0501	0.1069	0.0541	0.0783
						AD	0.0476	0.0373	0.0483	0.1077	0.0547	0.0786
0.0	−0.2	0.7	1.0	0.3	0.6	Bias	0.0013	−0.0008	0.0001	−0.0029	−0.0029	−0.0016
						SD	0.0449	0.0341	0.0504	0.1033	0.0437	0.0938
						AD	0.0446	0.0342	0.0504	0.1007	0.0427	0.0959

Table 2
Estimators for model (5.1) when $\{y_t\}$ is generated from model (b).

ϕ_0	ϕ_1	ψ_1	α_0	α_1	β_1		$\tilde{\phi}_{0n}$	$\tilde{\phi}_{1n}$	$\tilde{\psi}_{1n}$	$\tilde{\alpha}_{0n}$	$\tilde{\alpha}_{1n}$	$\tilde{\beta}_{1n}$
0.0	0.5	0.5	1.0	0.5	0.5	Bias	−0.0021	−0.0032	−0.0047	−0.0038	−0.0053	0.0006
						SD	0.0419	0.0383	0.1040	0.0862	0.0542	0.2921
						AD	0.0411	0.0369	0.1024	0.0850	0.0533	0.2851
0.0	0.0	−0.3	1.0	0.5	0.5	Bias	0.0005	−0.0030	−0.0011	−0.0011	−0.0044	0.0022
						SD	0.0392	0.0402	0.1001	0.0834	0.0606	0.2795
						AD	0.0397	0.0397	0.0989	0.0827	0.0616	0.2730
0.0	0.6	0.0	1.0	0.6	0.3	Bias	0.0012	−0.0005	0.0022	−0.0061	−0.0032	0.0277
						SD	0.0433	0.0382	0.1033	0.0879	0.0548	0.2592
						AD	0.0429	0.0374	0.1037	0.0877	0.0554	0.2648
0.0	−0.2	0.7	1.0	0.3	0.6	Bias	0.0006	−0.0010	0.0029	0.0019	−0.0048	−0.0053
						SD	0.0393	0.0336	0.0997	0.0791	0.0493	0.2813
						AD	0.0376	0.0360	0.0955	0.0775	0.0489	0.2702

Table 3
Estimators for model (5.1) when $\{y_t\}$ is generated from model (c).

ϕ_0	ϕ_1	ψ_1	α_0	α_1	β_1		$\tilde{\phi}_{0n}$	$\tilde{\phi}_{1n}$	$\tilde{\psi}_{1n}$	$\tilde{\alpha}_{0n}$	$\tilde{\alpha}_{1n}$	$\tilde{\beta}_{1n}$
0.0	0.5	0.5	1.0	0.5	0.5	Bias	0.0002	−0.0018	−0.0005	−0.0032	−0.0030	−0.0007
						SD	0.0714	0.0354	0.0398	0.1406	0.0469	0.0520
						AD	0.0697	0.0354	0.0399	0.1362	0.0445	0.0514
0.0	0.0	−0.3	1.0	0.5	0.5	Bias	−0.0023	−0.0013	0.0012	−0.0028	−0.0011	−0.0032
						SD	0.0498	0.0384	0.0462	0.0989	0.0601	0.0699
						AD	0.0483	0.0389	0.0456	0.0987	0.0581	0.0694
0.0	0.6	0.0	1.0	0.6	0.3	Bias	0.0004	−0.0028	0.0007	−0.0048	−0.0041	−0.0013
						SD	0.0540	0.0383	0.0370	0.1021	0.0560	0.0426
						AD	0.0533	0.0381	0.0373	0.1060	0.0559	0.0426
0.0	−0.2	0.7	1.0	0.3	0.6	Bias	0.0017	−0.0012	−0.0035	−0.0032	−0.0016	−0.0037
						SD	0.0516	0.0312	0.0422	0.1026	0.0396	0.0648
						AD	0.0514	0.0327	0.0424	0.1018	0.0379	0.0634

where η_t follows the standard normal distribution. The factor sample $\{y_t\}_{t=1}^n$ are generated from three different models:

(a) $y_t = 0.5y_{t-1} + \zeta_t$ (AR(1) model);

(b) $y_t = \zeta_t\sqrt{0.1+0.5y_{t-1}^2}$ (ARCH(1) model);

(c) $y_t = 0.5y_{t-1} + 0.5x_{t-1} + \zeta_t\sqrt{0.1+0.2y_{t-1}^2+0.3x_{t-1}^2}$ (FDAR model),

where ζ_t follows the standard normal distribution and is independent of η_t. We take the sample size $n=1000$ and use 1000 replications. The true parameters are $\theta_0 = (0.0, 0.5, 0.5, 1.0, 0.5, 0.5)$, $(0.0, 0.0, -0.3, 1.0, 0.5, 0.5)$, $(0.0, 0.6, 0.0, 1.0, 0.6, 0.3)$,

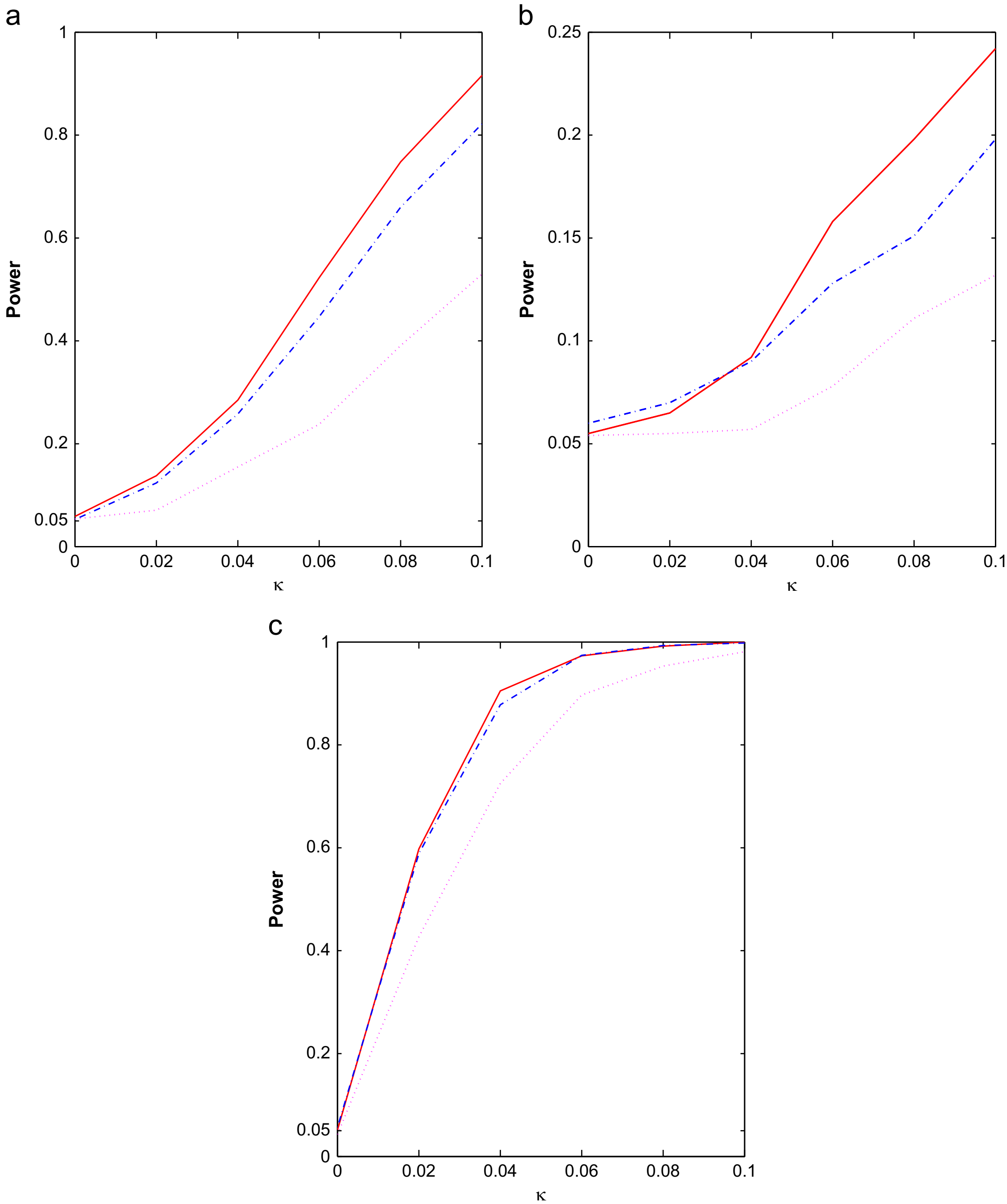

Fig. 1. (a) Power curves for $\rho=0$ (solid line), $\rho=0.4$ (dashed line), and $\rho=0.8$ (dotted line), based on model (a); (b) power curves for $\rho=0$ (solid line), $\rho=0.4$ (dashed line), and $\rho=0.8$ (dotted line), based on model (b); (c) power curves for $\rho=0$ (solid line), $\rho=0.4$ (dashed line), and $\rho=0.8$ (dotted line), based on model (c).

and (0.0, −0.2, 0.7, 1.0, 0.3, 0.6), respectively. Based on models (a)–(c), Tables 1–3 list the sample biases, the sample standard deviations (SD) and the average estimated asymptotic standard deviations (AD) of $\tilde{\theta}_n$, respectively. Each estimated asymptotic standard deviation is obtained from Theorem 4.1 with Ω_0 and Σ_0 being estimated by their sample averages. From Tables 1–3, we can see that $\tilde{\theta}_n$ has very small bias and its SD and AD are very close to each other. Interestingly, the way in which $\{y_t\}$ is generated does not affect the performance of $\tilde{\theta}_n$, hence it gives us enough freedom to choose factor in practice.

Next, we assess the performance of our score test (S_n) in finite samples. The model used to generate data samples is

$$x_t = 0.5x_{t-1} + \psi_1 y_{t-1} + \eta_t\sqrt{1.0 + 0.5x_{t-1}^2 + \beta_1 y_{t-1}^2}, \tag{5.2}$$

where $(\psi_1, \beta_1) = \kappa(1.0, 1.0)$ with $\kappa = \{0.0, 0.02, 0.04, \ldots, 0.1\}$, and the factor samples $\{y_t\}_{t=1}^n$ are generated from models (a)–(c). Here, $\{\eta_t\}_{t=1}^n$ and $\{\zeta_t\}_{t=1}^n$ are random samples generated from a bivariate normal distribution with mean zero, variance one, and covariance ρ. Again, we set the sample size $n = 1000$ and use 1000 replications, and choose the significance level $\alpha = 0.05$. For $\rho = 0.0$, 0.4, and 0.8, the power curves are plotted in Fig. 1(a)–(c), based on models (a)–(c), respectively. The sizes correspond to the cases when $\kappa = 0.0$. From Fig. 1, it is clear that the sizes of S_n are close to their nominal ones. Although the power becomes weaker as the value of ρ increases, S_n performs well no matter how the factor samples are generated. Overall, the numerical study shows that both $\tilde{\theta}_n$ and S_n have good performances in finite samples.

6. An example

In this section, we study the causal relationship between Hong Kong (HK) stock market and US stock market. We choose the Hang Seng index (HSI) and SP500 Composite index (SPCI) as the proxies for the HK stock market and the US stock market, respectively. The data sets used are the daily closing HSI data and SPCI data from June 16, 2008 to June 10, 2010, and each of them has in total 501 observations; see Fig. 2(a). Furthermore, we denote the log-return of HSI and SPCI by x_t and y_t, respectively, and plot them in Fig. 2(b).

We first consider the causal relation from y_t to x_t. Unless stated otherwise, we set the significance level $\alpha = 0.05$. According to AIC, we choose $p=2$ and $q=3$ in model (2.4). Then, we obtain $S_n = 73.6$, which is greater than 12.59 (the 95% upper percentile of χ_6^2). So there exists the simultaneous causality-in-mean and causality-in-variance from y_t to x_t. Therefore, we use the following FDAR model:

$$x_t = \phi_0 + \sum_{i=1}^{2}\phi_i x_{t-i} + \sum_{i=1}^{3}\psi_i y_{t-i} + \eta_t\sqrt{\alpha_0 + \sum_{i=1}^{2}\alpha_i x_{t-i}^2 + \sum_{i=1}^{3}\beta_i y_{t-i}^2}, \tag{6.1}$$

to fit the data set $\{x_t\}$. All parameters are estimated through the QMLE method and these results are reported in Table 4 with the standard errors in parentheses. Based on the residuals $\{\hat{\eta}_t\}$, the Li–Mak tests $Q^2(6)$ and $Q^2(12)$ reported in Table 4

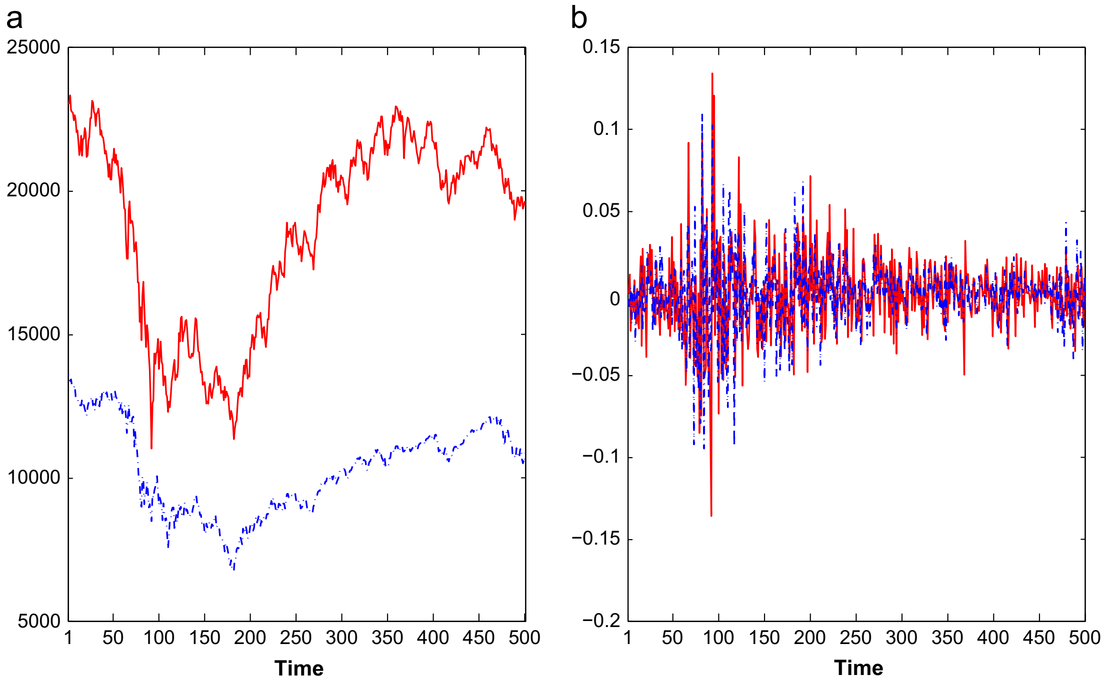

Fig. 2. (a) The daily closing HSI (—) and SP500 (× 10) (-.-.) and (b) the log-return of HSI (—) and SP500 (-.-.).

Table 4
Results for models (6.1)–(6.4).

Parameters	Causal models from y_t to x_t		Parameters	Causal models from x_t to y_t	
	Model (6.1)	Model (6.2)		Model (6.3)	Model (6.4)
ϕ_0	−0.0011 (0.0007)		π_0	0.0004 (0.0006)	
ϕ_1	−0.1986 (0.0493)	−0.1977 (0.0493)	π_1	−0.2717 (0.0489)	−0.2665 (0.0488)
ϕ_2	−0.1211 (0.0500)	−0.1177 (0.0499)	π_2	−0.2057 (0.0552)	−0.2095 (0.0548)
ψ_1	0.6395 (0.0471)	0.6401 (0.0472)	π_3	−0.0181 (0.0511)	
ψ_2	0.1780 (0.0681)	0.1735 (0.0679)	π_4	0.0989 (0.0517)	
ψ_3	0.1877 (0.0589)	0.1819 (0.0587)	ω_0	0.2726 (0.0440)	0.2700 (0.0442)
α_0	0.0001 (0.00002)	0.0001 (0.00002)	ω_1	0.1422 (0.0459)	0.1391 (0.0459)
α_1	0.0883 (0.0522)	0.0842 (0.0518)	τ_0	0.000016 (0.000011)	0.000020 (0.000012)
α_2	0.1064 (0.0560)	0.0986 (0.0552)	τ_1	0.000001 (0.0389)	0.000001 (0.0394)
β_1	0.0148 (0.0350)	0.0150 (0.0353)	τ_2	0.1648 (0.0714)	0.1642 (0.0719)
β_2	0.3625 (0.0984)	0.3563 (0.0976)	τ_3	0.2678 (0.0796)	0.2651 (0.0803)
β_3	0.1603 (0.0696)	0.1577 (0.0692)	τ_4	0.2476 (0.0794)	0.2428 (0.0797)
			ν_0	0.1375 (0.0520)	0.1408 (0.0534)
			ν_1	0.1342 (0.0537)	0.1247 (0.0537)
	$Q^2(6)=10.15$ $Q^2(12)=16.19$	$Q^2(6)=10.63$ $Q^2(12)=16.44$		$Q^2(6)=9.46$ $Q^2(12)=13.33$	$Q^2(6)=11.01$ $Q^2(12)=14.14$
$C(6)=17.67$ $C(12)=28.64$					

indicate that model (6.1) is adequate. However, the parameter ϕ_0 in model (6.1) is not significantly different from zero. Hence, by using the QMLE method, we re-fit the data set $\{x_t\}$ as

$$x_t=\sum_{i=1}^{2}\phi_i x_{t-i}+\sum_{i=1}^{3}\psi_i y_{t-i}+\eta_t\sqrt{\alpha_0+\sum_{i=1}^{2}\alpha_i x_{t-i}^2+\sum_{i=1}^{3}\beta_i y_{t-i}^2}, \tag{6.2}$$

where all results for model (6.2) are reported in Table 4, and indicate that model (6.2) is adequate. From this model, we observe that US market affects HK market in both the mean and variance of return. Specifically, the influence for the mean of return lasts for three days, and it becomes weak as time goes by; while the influence for the variance of return has one-day delay since β_1 closes to zero, and then starts to mitigate two days later.

Next, we consider the causal relation from x_t to y_t. Since HK stock market is one day earlier than US stock market in calendar, we use $S_n^\diamond$ instead of S_n in this case. According to AIC, we choose $p=4$ and $q=1$ in model (2.8). Then, we obtain $S_n^\diamond=127.8$, which is greater than 9.5 (the 95% upper percentile of χ_4^2). Similar to model (6.1), we obtain the following fitted model for the data set $\{y_t\}$:

$$y_t=\pi_0+\sum_{i=1}^{4}\pi_i y_{t-i}+\sum_{i=0}^{1}\omega_i x_{t-i}+\zeta_t\sqrt{\tau_0+\sum_{i=1}^{4}\tau_i y_{t-i}^2+\sum_{i=0}^{1}\nu_i x_{t-i}^2}, \tag{6.3}$$

where all results for model (6.3) are reported in Table 4, and indicate that model (6.3) is adequate. Furthermore, we find that the parameters π_0, π_3, and π_4 in model (6.3) are not significantly different from zero. Thus, similar to model (6.2), we re-fit the data set $\{y_t\}$ using the model

$$y_t = \sum_{i=1}^{2} \pi_i y_{t-i} + \sum_{i=0}^{1} \omega_i x_{t-i} + \zeta_t \sqrt{\tau_0 + \sum_{i=1}^{4} \tau_i y_{t-i}^2 + \sum_{i=0}^{1} \nu_i x_{t-i}^2}. \tag{6.4}$$

Again, all results for this adequate model are reported in Table 4. Since the parameters ω_0, ω_1, ν_0, and ν_1 in model (6.4) are significantly different from zero, we claim that HK market causes US market in both the mean and variance. However, compared with model (6.2), the impact period from HK market to US market only lasts for two days, and is shorter than the one from US market to HK market. This is consistent with the fact that the US market is the largest capital market in the world. Moreover, based on the residuals from models (6.2) and (6.4), the CCF tests $C(6)$ and $C(12)$ reported in Table 4 indicate that $\{\eta_t\}$ and $\{\zeta_t\}$ are independent, and hence the bivariate FDAR models (6.2) and (6.4) are enough for us to characterize the causal relations between HK market and US market.

Acknowledgments

The authors gratefully acknowledge the constructive suggestions and comments from the Editor, the Associate Editor and three referees that greatly improve this paper. The research of S. Guo was supported by the Chinese NSF Grants (11101408) and Key Lab of Random Complex Structures and Data Science of Chinese Academy of Sciences. The research of S. Ling was supported by the Hong Kong Grants (641912 and 603413, HKUST). The research of K. Zhu was supported by the Chinese NSF Grants (11201459) and the National Center for Mathematics and Interdisciplinary Sciences of Chinese Academy of Sciences.

Appendix A. Proof of Proposition 2.1

Proof. For brevity, we only prove that

the inequality (2.3) fails if and only if all ψ_i and β_i are zeros. (A.1)

It suffices to show the necessity of (A.1). Suppose that relation (2.3) does not hold. By Assumption 2.1 and a direct calculation, it follows that

$$E\left[\left(\sum_{i=1}^{q} \psi_i y_{t-i}\right)^2 \middle| \mathcal{I}_{1t-1}\right] - \left[E\left(\sum_{i=1}^{q} \psi_i y_{t-i} \middle| \mathcal{I}_{1t-1}\right)\right]^2 + E\left(\sum_{i=1}^{q} \beta_i y_{t-i}^2 \middle| \mathcal{I}_{1t-1}\right) = \sum_{i=1}^{q} \beta_i y_{t-i}^2 \tag{A.2}$$

a.s. Then, if $\beta_1 \neq 0$, we have $y_{t-1} \in \sigma(\mathcal{I}_{1,t-1}, \mathcal{I}_{2,t-2})$, and this is a contradiction with Assumption 2.1(i). Hence, $\beta_1 = 0$. Similarly, $\beta_2 = \cdots = \beta_q = 0$. Next, when all β_i are zeros, by (A.2) and Hölder's inequality, we know that $\sum_{i=1}^{q} \psi_i y_{t-i} \equiv constant$ a.s. Then, if $\psi_1 \neq 0$, we have $y_{t-1} \in \mathcal{I}_{2,t-2}$, and this is against Assumption 2.1(i). Hence, $\psi_1 = 0$. Similarly, $\psi_2 = \cdots = \psi_q = 0$. This completes the proof. □

Appendix B. Proofs of Theorems 3.1 and 4.1

To facilitate presentation in the proof of Theorem 3.1, we denote $\tilde{\varepsilon}_t(\phi) = x_t - \phi' X_t$ and $\tilde{h}_t(\alpha) = \alpha' X_t^*$ and let

$$\tilde{L}_n(\theta_1) = -\frac{1}{n}\sum_{t=m}^{n} \tilde{l}_t(\theta_1) \quad \text{with } \tilde{l}_t(\theta_1) = \log\sqrt{\tilde{h}_t(\alpha)} + \frac{\tilde{\varepsilon}_t^2(\phi)}{2\tilde{h}_t(\alpha)},$$

where $\theta_1 = (\phi', \alpha')'$, and $\tilde{L}_n(\theta_1) \triangleq L_n(\theta)|_{(\psi', \beta')' = 0}$ is the quasi-log-likelihood function under H_0.

Proof of Theorem 3.1. First, by (3.2), (3.3) and a direct calculation, we can show that

$$T_n\left(\hat{\theta}_n\right) = \left(-\frac{1}{n}\sum_{t=m}^{n} \frac{\tilde{\varepsilon}_t(\hat{\phi}_n)}{\tilde{h}_t(\hat{\alpha}_n)} Y_t', \ \frac{1}{2n}\sum_{t=m}^{n}\left[\frac{1}{\tilde{h}_t(\hat{\alpha}_n)} - \frac{\tilde{\varepsilon}_t^2(\hat{\phi}_n)}{\tilde{h}_t^2(\hat{\alpha}_n)}\right] Y_t^{*\prime}\right)'. \tag{B.1}$$

Recall that $\overline{\theta}_{10} = (\overline{\phi}_0', \overline{\alpha}_0')'$ and $\hat{\theta}_{1n} = (\hat{\phi}_n', \hat{\alpha}_n')'$. By Taylor's expansion, we have

$$\frac{\tilde{\varepsilon}_t(\hat{\phi}_n)}{\tilde{h}_t(\hat{\alpha}_n)} = \frac{\tilde{\varepsilon}_t(\overline{\phi}_0)}{\tilde{h}_t(\overline{\alpha}_0)} - \left(\frac{X_t'}{\tilde{h}_t(\xi_{2n})}, \ \frac{\tilde{\varepsilon}_t(\xi_{1n}) X_t^{*\prime}}{\tilde{h}_t^2(\xi_{2n})}\right)\left(\hat{\theta}_{1n} - \overline{\theta}_{10}\right),$$

$$\frac{1}{\tilde{h}_t(\hat{\alpha}_n)} = \frac{1}{\tilde{h}_t(\overline{\alpha}_0)} - \left(0, \ \frac{X_t^{*\prime}}{\tilde{h}_t^2(\xi_{2n})}\right)\left(\hat{\theta}_{1n} - \overline{\theta}_{10}\right),$$

$$\frac{\tilde{\varepsilon}_t^2(\hat{\phi}_n)}{\tilde{h}_t^2(\hat{\alpha}_n)} = \frac{\tilde{\varepsilon}_t^2(\overline{\phi}_0)}{\tilde{h}_t^2(\overline{\alpha}_0)} - 2\left(\frac{\tilde{\varepsilon}_t(\xi_{1n}) X_t'}{\tilde{h}_t^2(\xi_{2n})}, \ \frac{\tilde{\varepsilon}_t^2(\xi_{1n}) X_t^{*\prime}}{\tilde{h}_t^3(\xi_{2n})}\right)\left(\hat{\theta}_{1n} - \overline{\theta}_{10}\right),$$

where (ξ_{1n}, ξ_{2n}) lies between $\hat{\theta}_{1n}$ and $\overline{\theta}_{10}$. Note that $\tilde{\varepsilon}_t(\overline{\phi}_0)/\sqrt{\tilde{h}_t(\overline{\alpha}_0)} = \eta_t$ under H_0. Therefore, by (B.1), it follows that, under H_0,

$$T_n\left(\hat{\theta}_n\right) = \left(-\frac{1}{n}\sum_{t=m}^{n}\frac{\eta_t Y_t'}{\sqrt{\tilde{h}_t(\overline{\alpha}_0)}},\ \frac{1}{2n}\sum_{t=m}^{n}\frac{(1-\eta_t^2)Y_t^{*\prime}}{\tilde{h}_t(\overline{\alpha}_0)}\right)' + \begin{pmatrix} S_{1n} \\ S_{2n}\end{pmatrix}\left(\hat{\theta}_{1n} - \overline{\theta}_{10}\right), \tag{B.2}$$

where

$$S_{1n} = \frac{1}{n}\sum_{t=m}^{n}\left(\frac{Y_t X_t'}{\tilde{h}_t(\xi_{2n})}, \frac{\tilde{\varepsilon}_t(\xi_{1n})Y_t X_t^{*\prime}}{\tilde{h}_t^2(\xi_{2n})}\right),$$

$$S_{2n} = \frac{1}{n}\sum_{t=m}^{n}\left(\frac{\tilde{\varepsilon}_t(\xi_{1n})Y_t^* X_t'}{\tilde{h}_t^2(\xi_{2n})}, -\frac{Y_t^* X_t^{*\prime}}{2\tilde{h}_t^2(\xi_{2n})} + \frac{\tilde{\varepsilon}_t^2(\xi_{1n})Y_t^* X_t^{*\prime}}{\tilde{h}_t^3(\xi_{2n})}\right).$$

Note that for any $(i,j) \in \{1,\ldots,q\} \times \{1,\ldots,1+p\}$, the (i,j)-th entry of $Y_t X_t'$ is $x_{t-j+1}y_{t-i}$, where we set $x_t \equiv 1$ for convenience. Since $\tilde{h}_t(\alpha) \ge \alpha_0^L > 0$ holds uniformly in Θ_1 by Assumption 3.1, it is straightforward to see that

$$\begin{aligned} E\left[\sup_{\theta_1\in\Theta_1}\frac{|x_{t-j+1}y_{t-i}|}{\tilde{h}_t(\alpha)}\right] &\le O(1)E\left[\sup_{\theta_1\in\Theta_1}\frac{|x_{t-j+1}y_{t-i}|}{\sqrt{\tilde{h}_t(\alpha)}}\right] \le O(1)E\left[\frac{|x_{t-j+1}y_{t-i}|}{\sqrt{\tilde{h}_t^L}}\right] \le O(1)E\left[\frac{|x_{t-j+1}y_{t-i}|}{\sqrt{\alpha_{j-1}^L}|x_{t-j+1}|}\right] \\ &= O(1)E|y_{t-i}| < \infty, \end{aligned} \tag{B.3}$$

where $\tilde{h}_t^L = \alpha_0^L + \alpha_1^L x_{t-1}^2 + \cdots + \alpha_p^L x_{t-p}^2$, and the last inequality holds by Assumption 3.2. Thus, it follows that

$$E\left[\sup_{\theta_1\in\Theta_1}\frac{\|Y_t X_t'\|}{\tilde{h}_t(\alpha)}\right] < \infty.$$

Similarly, since $\tilde{\varepsilon}_t(\phi) = \eta_t\sqrt{\tilde{h}_t(\overline{\alpha}_0)} + (\overline{\phi}_0 - \phi)'X_t$ under H_0, as for (B.3), we can show that

$$E\left[\sup_{\theta_1\in\Theta_1}\frac{\|\tilde{\varepsilon}_t(\phi)Y_t X_t^{*\prime}\|}{\tilde{h}_t^2(\alpha)}\right] < \infty.$$

Then, by Theorem 1 of Ling and McAleer (2003) and the dominated convergence theorem, it follows that

$$\begin{aligned} S_{1n} &= \left(E\left[\frac{Y_t X_t'}{\tilde{h}_t(\xi_{2n})}\right], E\left[\frac{\tilde{\varepsilon}_t(\xi_{1n})Y_t X_t^{*\prime}}{\tilde{h}_t^2(\xi_{2n})}\right]\right) + o_p(1) \\ &= \left(E\left[\frac{Y_t X_t'}{\tilde{h}_t(\overline{\alpha}_0)}\right], E\left[\frac{\eta_t Y_t X_t^{*\prime}}{\tilde{h}_t^{3/2}(\overline{\alpha}_0)}\right]\right) + o_p(1) \\ &= \left(E\left[\frac{Y_t X_t'}{\tilde{h}_t(\overline{\alpha}_0)}\right], 0\right) + o_p(1), \end{aligned} \tag{B.4}$$

where the last equation holds due to the double expectation. Similarly, we can show that

$$S_{2n} = \left(0,\ E\left[\frac{Y_t^* X_t^{*\prime}}{2\tilde{h}_t^2(\overline{\alpha}_0)}\right]\right) + o_p(1). \tag{B.5}$$

Note that $E|x_t|^\iota < \infty$ for some $\iota > 0$ by Assumption 3.2. Thus, by Assumptions 3.1–3.2, Theorem 1 in Ling (2007) showed that $\sqrt{n}(\hat{\theta}_{1n} - \overline{\theta}_{10}) = O_p(1)$ under H_0. Therefore, by (B.2), (B.4) and (B.5), we have under H_0,

$$\sqrt{n}T_n\left(\hat{\theta}_n\right) = \left(-\frac{1}{\sqrt{n}}\sum_{t=m}^{n}\frac{\eta_t Y_t'}{\sqrt{\tilde{h}_t(\overline{\alpha}_0)}},\ \frac{1}{2\sqrt{n}}\sum_{t=m}^{n}\frac{(1-\eta_t^2)Y_t^{*\prime}}{\tilde{h}_t(\overline{\alpha}_0)}\right)' + \mathrm{diag}\left\{E\left[\frac{Y_t X_t'}{\tilde{h}_t(\overline{\alpha}_0)}\right], E\left[\frac{Y_t^* X_t^{*\prime}}{2\tilde{h}_t^2(\overline{\alpha}_0)}\right]\right\}\sqrt{n}\left(\hat{\theta}_{1n} - \overline{\theta}_{10}\right) + o_p(1). \tag{B.6}$$

Since $\hat{\theta}_{1n}$ is the QMLE of $\tilde{L}_n(\theta_1)$, by Taylor's expansion, we have

$$0 = \frac{\partial \tilde{L}_n(\hat{\theta}_{1n})}{\partial\theta_1} = \frac{\partial \tilde{L}_n(\overline{\theta}_{10})}{\partial\theta_1} + \left(\hat{\theta}_{1n} - \overline{\theta}_{10}\right)\frac{\partial^2\tilde{L}_n(\zeta_n)}{\partial\theta_1\partial\theta_1'},$$

where ζ_n lies between $\hat{\theta}_{1n}$ and $\overline{\theta}_{10}$. Then it follows that

$$\sqrt{n}\left(\hat{\theta}_{1n} - \overline{\theta}_{10}\right) = -\left(\frac{1}{n}\sum_{t=m}^{n}\frac{\partial^2\tilde{l}_t(\zeta_n)}{\partial\theta_1\partial\theta_1'}\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=m}^{n}\frac{\partial\tilde{l}_t(\overline{\theta}_{10})}{\partial\theta_1}\right).$$

By a similar argument as for (B.4), we can show that

$$\frac{1}{n}\sum_{t=m}^{n}\frac{\partial^2 \tilde{l}_t(\zeta_n)}{\partial\theta_1\partial\theta_1'} = \mathrm{diag}\left\{E\left[\frac{X_tX_t'}{\tilde{h}_t(\overline{\alpha}_0)}\right], E\left[\frac{X_t^*X_t^{*\prime}}{2\tilde{h}_t^2(\overline{\alpha}_0)}\right]\right\} + o_p(1).$$

Thus, it follows that

$$\sqrt{n}\left(\hat{\theta}_{1n}-\overline{\theta}_{10}\right) = -\mathrm{diag}\left\{\left[E\left(\frac{X_tX_t'}{\tilde{h}_t(\overline{\alpha}_0)}\right)\right]^{-1}, \left[E\left(\frac{X_t^*X_t^{*\prime}}{2\tilde{h}_t^2(\overline{\alpha}_0)}\right)\right]^{-1}\right\}\left(-\frac{1}{\sqrt{n}}\sum_{t=m}^{n}\frac{\eta_tX_t'}{\sqrt{\tilde{h}_t(\overline{\alpha}_0)}}, \frac{1}{2\sqrt{n}}\sum_{t=m}^{n}\frac{(1-\eta_t^2)X_t^{*\prime}}{\tilde{h}_t(\overline{\alpha}_0)}\right)' + o_p(1). \tag{B.7}$$

As a result, by (B.6)–(B.7) we have

$$\sqrt{n}T_n\left(\hat{\theta}_n\right) = \frac{1}{\sqrt{n}}\sum_{t=m}^{n}A_t(\overline{\theta}_{10})\left(\eta_t, \frac{1-\eta_t^2}{\sqrt{2}}\right)' + o_p(1),$$

where $A_t(\overline{\theta}_{10})$ is defined as in (3.4). Note that $\Xi > 0$, because $J > 0$ and Assumption 2.1(i) holds. Then, the conclusion follows from the martingale central limit theorem. This completes the proof. □

Next, we give the proof of Theorem 4.1. The following lemma below is needed to prove the strong consistency of $\tilde{\theta}_n$.

Lemma B.1. *For any $\theta^* \in \Theta$, let $B_\delta(\theta^*) = \{\theta \in \Theta : \|\theta-\theta^*\| < \delta\}$ be an open neighborhood of θ^* with radius $\delta > 0$. Suppose that the conditions in Theorem 4.1 hold. Then,*

(i) $E\left[\sup_{\theta\in\Theta}|l_t(\theta)|\right] < \infty$;
(ii) $E[l_t(\theta)]$ *has a unique minimum at* θ_0;
(iii) $E\left[\sup_{\theta\in B_\delta(\theta^*)}|l_t(\theta)-l_t(\theta^*)|\right] \to 0$ *as* $\delta \to 0$.

Proof. First, by Assumptions 4.1–4.2, the proof of (i) is similar to that of (B.3) (see also Lemma B.2 in Ling, 2007). Second, a direct calculation shows that

$$\begin{aligned} E\left[l_t(\theta)\right] &= E\left\{\log\sqrt{h_t(\theta)} + \frac{h_t(\theta_0)}{2h_t(\theta)}E\left[\frac{\varepsilon_t^2(\theta)}{h_t(\theta_0)}\bigg|\mathcal{I}_{t-1}\right]\right\} \\ &= E\left\{\log\sqrt{h_t(\theta)} + \frac{h_t(\theta_0)}{2h_t(\theta)}E[|\eta_t-\gamma_t|^2|\mathcal{I}_{t-1}]\right\}\left(\gamma_t = \left[(\phi-\overline{\phi}_0)'X_t + (\psi-\overline{\psi}_0)'Y_t\right]\Big/\sqrt{h_t(\theta_0)} \in \mathcal{I}_{t-1}\right) \\ &\geq E\left\{\log\sqrt{h_t(\theta)} + \frac{h_t(\theta_0)}{2h_t(\theta)}E[\eta_t^2|\mathcal{I}_{t-1}]\right\} \\ &= E\left\{\log\sqrt{\frac{h_t(\theta)}{h_t(\theta_0)}} + \frac{h_t(\theta_0)}{2h_t(\theta)} + \log\sqrt{h_t(\theta_0)}\right\} \geq E\left\{\frac{1}{2} + \log\sqrt{h_t(\theta_0)}\right\} = E\left[l_t(\theta_0)\right], \end{aligned}$$

where the last second inequality holds due to the fact that $E[\eta_t-a]^2 \geq E[\eta_t - E(\eta_t|\mathcal{I}_{t-1})]^2$ for any $a \in \mathcal{I}_{t-1}$ and the last inequality holds since the function $f(x) = \log x + 1/x$ reaches the minimum at $x=1$. Moreover, if $E[l_t(\theta)] = E[l_t(\theta_0)]$, *i.e.*, $E[l_t(\theta)]$ reaches the minimum, then we have

$$(\phi-\overline{\phi}_0)'X_t + (\psi-\overline{\psi}_0)'Y_t = 0 \text{ a.s.} \quad \text{and} \quad (\alpha-\overline{\alpha}_0)'X_t^* + (\beta-\overline{\beta}_0)'Y_t^* = 0 \text{ a.s.},$$

which implies that $\theta = \theta_0$ by Assumption 2.1(i). Thus, we claim that $E[l_t(\theta)]$ has a unique minimum at θ_0, i.e., (ii) follows.

Last, by Taylor's expansion, we have

$$l_t(\theta) - l_t(\theta^*) = (\theta-\theta^*)'\frac{\partial l_t(\xi^*)}{\partial\theta}, \tag{B.8}$$

where ξ^* lies between θ and θ^*. Similar to the proof of (B.3), by Assumptions 4.1–4.2, we can show that

$$E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial l_t(\theta)}{\partial\theta}\right\|\right] < \infty.$$

Thus, it follows from (B.8) that (iii) holds. This completes the proof. □

Proof of Theorem 4.1. By Lemma B.1, a similar proof as for Theorem 2.1 in Zhu and Ling (2011) shows that (i) holds. Next, we use Theorem 4.1.3 in Amemiya (1985) to prove (ii). So, we only need to check that

(a) $(1/n)\sum_{t=m}^{n}\partial^2 l_t(\theta)/\partial\theta\partial\theta'$ exists and is continuous in Θ.
(b) For any sequence θ_n such that $\theta_n \to \theta_0$ in probability, we have $(1/n)\sum_{t=m}^{n}\partial^2 l_t(\theta_n)/\partial\theta\partial\theta' = \Omega_0 + o_p(1)$, where Ω_0 is a finite positive definite matrix;

(c) $(1/\sqrt{n})\sum_{t=m}^{n}\partial l_t(\theta_0)/\partial\theta \to_d N(0,\Sigma_0)$ as $n\to\infty$, where Σ_0 is a finite positive definite matrix.

First, because J is positive definite and Assumption 2.1(i) holds, it is not hard to show that both Ω_0 and Σ_0 are positive definite. Second, by Assumptions 4.1–4.2 and a similar proof as for (B.3), we can show that

$$E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right\|\right]<\infty.$$

Then, part (a) follows from the ergodic theorem and part (b) is implied by Theorem 1 in Ling and McAleer (2003) and the dominated convergence theorem. Third, part (c) is directly from the martingale central limit theorem and the Crámer–Wold device. Therefore, we know that (ii) holds. This completes the proof. □

References

Amemiya, T., 1985. Advanced Econometrics. Cambridge, Harvard University Press, Cambridge, MA.
Baillie, R., Bollerslev, T., 1990. Intra-day and inter-market volatility in foreign exchange rates. Rev. Econom. Stud. 58, 565–585.
Borkovec, M., Klüppelberg, C., 2001. The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. Ann. Appl. Probab. 11, 1220–1241.
Boudjellaba, H., Dufour, J.-M., Roy, R., 1992. Testing causality between two vectors in multivariate autoregressive moving average models. J. Amer. Statist. Assoc. 87, 1082–1090.
Cheung, Y.-W., Ng, L.K., 1996. A causality-in-variance test and its application to financial market prices. J. Econometrics 72, 33–48.
Comte, F., Lieberman, O., 2000. Second-order noncausality in multivariate GARCH processes. J. Time Ser. Anal. 21, 535–557.
Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. Econometrica 50, 987–1008.
Engle, R.F., Ito, T., Lin, W.L., 1990. Meteor shower or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. Econometrica 59, 524–542.
Geske, R., Roll, R., 1983. The monetary and fiscal linkage between stock returns and inflation. J. Finance 38, 1–33.
Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. J. Amer. Statist. Assoc. 77, 304–313.
Geweke, J., 1984a. Inference and causality in economic time series. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, vol. 2, North-Holland, Amsterdam.
Geweke, J., 1984b. Measures of conditional linear dependence and feedback between time series. J. Amer. Statist. Assoc. 79, 907–915.
Gouriéroux, C., Monfort, A., 1997. Time Series and Dynamic Models. Cambridge University Press, Cambridge, UK.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–459.
Granger, C.W.J., 1980. Testing for causality: a personal view. J. Econom. Dynam. Control 2, 329–352.
Granger, C.W.J., Robins, R.P., Engle, R.F., 1986. Wholesale and retail prices: bivariate time-series modeling with forecastable error variances. In: Belsley, D.A., Kuh, E. (Eds.), Model Reliability. MIT Press, Cambridge, MA, pp. 1–17.
Hafner, C.M., Herwartz, H., 2006. A lagrange multiplier test for causality in variance. Econom. Lett. 93, 137–141.
Hamao, Y., Masulis, R.W., Ng, V., 1990. Correlations in price changes and volatility across international stock markets. Rev. Financial Stud. 3, 281–307.
Hiemstra, C., Jones, J.D., 1994. Testing for linear and nonlinear Granger causality in the stock price-volume relation. J. Finance 49, 1639–1664.
Hong, Y., 2001. A test for volatility spillover with application to exchange rates. J. Econometrics 103, 183–224.
Lee, B.-S., 1992. Causal relations among stock returns, interest rates, real activity, and inflation. J. Finance 47, 1591–1603.
Li, W.K., Mak, T.K., 1994. On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. J. Time Ser. Anal. 15, 627–636.
Ling, S., 2004. Estimation and testing stationarity for double autoregressive models. J. Roy. Statist. Soc. B 66, 63–78.
Ling, S., 2007. A double AR(p) model: structure and estimation. Statist. Sinica 17, 161–175.
Ling, S., McAleer, M., 2003. Asymptotic theory for a new vector ARMA-GARCH model. Econometric Theory 19, 280–310.
Ng, A., 2000. Volatility spillover effects from Japan and the US to the Pacific-Basin. J. Internat. Money Finance 19, 207–233.
Nishiyama, Y., Hitomi, K., Kawasaki, Y., Jeong, K., 2011. A consistent nonparametric test for nonlinear causality—specification in time series regression. J. Econometrics 165, 112–127.
Pantelidis, T., Pittis, N., 2004. Testing for Granger causality in variance in the presence of causality in mean. Econom. Lett. 85, 201–207.
Ram, R., Spencer, D.E., 1983. Stock returns, real activity, inflation and money: comment. Amer. Econom. Rev. 73, 463–470.
Ross, S.A., 1989. Information and volatility: the no-arbitrage martingale approach to timing and resolution irrelevancy. J. Finance 44, 1–17.
Sims, C.A., 1972. Money, income, and causality. Amer. Econom. Rev. 62, 540–552.
Sims, C.A., 1980. Macroeconomics and reality. Econometrica 48, 1–48.
Stock, J.H., Watson, M.W., 1989. Interpreting the evidence on money-income causation. J. Econometrics 40, 161–182.
Weiss, A.A., 1986. Asymptotic theory for ARCH models: estimation and testing. Econometric Theory 2, 107–131.
Zhu, K., Ling, S., 2011. Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. Ann. Statist. 39, 2131–2163.