PHBS WORKING PAPER SERIES

Revisiting the Forward Rate Unbiasedness Hypothesis

Cindy S.H. Wang*
Peking University

Cheng Hsiao
University of Southern California

Yi-Chi Chen
National Cheng Kung University

November 2021

Working Paper 20211104

Abstract

This research takes a forward rate unbiasedness hypothesis (FRUH) perspective to the study of event-detection and tests for the FRUH by a newly proposed correlation test for two time series with imbalanced integrated orders. This new test generalizes Hong's (1996a) statistics for two processes being imbalanced via the AR(k) approximation framework, where k is allowed to increase with sample size T. Our procedure is easy to implement, in that it obviates the need for the bootstrapping procedure and long-run variance framework. Monte Carlo studies show that Hong's generalized tests perform very well in finite samples and also avoid the spurious regression problem due to persistent time series. We illustrate the usefulness of our methodology by testing the forward rate unbiasedness hypothesis and find that the validity of the unbiasedness proposition is regime-dependent and less likely to hold during the periods of high financial instability including the global financial crisis and COVID-19's market turbulence. Our testing results from the duration of Hong's tests discover that co-movements between the spot and forward exchange rates have a significant time-lag effect. This finding agrees with the theoretical justification which attributes the deviations from the parity to infrequent revisions of investor portfolio decisions. Finally, our AR-filtered correlation tests during financial crises or economic downturns exhibit distinctive patterns that differed from those during tranquil periods, which could serve as the basis for an early warning system when systemic risks were to emerge in financial markets.

Keywords: Fractionally integrated process, autoregressive (AR) approximation, spurious regression, persistent time series, foreign exchange rate unbiasedness hypothesis (FRUH).
JEL Classification: C1, C4

Peking University HSBC Business School
University Town, Nanshan District
Shenzhen 518055, China

Revisiting the Forward Rate Unbiasedness Hypothesis

Cindy S.H. Wang † **Cheng Hsiao**†† **Yi-Chi Chen**†††

March 16, 2021

Abstract This research takes a forward rate unbiasedness hypothesis (FRUH) perspective to the study of event-detection and tests for the FRUH by a newly proposed correlation test for two time series with imbalanced integrated orders. This new test generalizes Hong's (1996a) statistics for two processes being imbalanced via the AR(k) approximation framework, where k is allowed to increase with sample size T. Our procedure is easy to implement, in that it obviates the need for the bootstrapping procedure and long-run variance framework. Monte Carlo studies show that Hong's generalized tests perform very well in finite samples and also avoid the spurious regression problem due to persistent time series. We illustrate the usefulness of our methodology by testing the forward rate unbiasedness hypothesis and find that the validity of the unbiasedness proposition is regime-dependent and less likely to hold during the periods of high financial instability including the global financial crisis and COVID-19's market turbulence. Our testing results from the duration of Hong's tests discover that co-movements between the spot and forward exchange rates have a significant time-lag effect. This finding agrees with the theoretical justification which attributes the deviations from the parity to infrequent revisions of investor portfolio decisions. Finally, our AR-filtered correlation tests during financial crises or economic downturns exhibit distinctive patterns that differed from those during tranquil periods, which could serve as the basis for an early warning system when systemic risks were to emerge in financial markets.

KEYWORDS: Fractionally integrated process, autoregressive (AR) approximation, spurious regression, persistent time series, foreign exchange rate unbiasedness hypothesis (FRUH).

† Peking University, HSBC Business School, China.
† Corresponding Author, e-Mail: cindywang9177@gmail.com;
†† University of Southern California, U.S.A. and WISE, Xiamen University, China.
††† Department of Economics, National Cheng Kung University, Taiwan, R.O.C.

1. Introduction

Economists and financial analysts are often interested in knowing if a shock in one market affects other markets. Developing a useful and robust correlation test is the first step towards investigating how information is transmitted across assets and markets and for exploring the significance of its effect. In general, the common approach to measuring the correlation between two time series is to perform a regression analysis in the pre-specified functional form, and then test for significance of the regression coefficients. However, a series of studies recognize that the most widely discussed problem undermining confidence in the reliable regression test is the uncertainty about the unobservable time-series properties of the variables, such as the degrees of the persistence. When the integrated orders of two series are imbalanced, i.e., the regressor is persistent (near unit root or long memory) and the regressand displays covariance stationary time series , empirical facts cast doubt on the finite sample accuracy of standard t-tests based on least squares regressions. For example, under the local-to-unity assumptions, the distribution of the standard t-statistics is non-standard and thus conventional statistical inference of current empirical findings could be called into question (e.g., Cavanagh *et al.*, 1995; Maynard and Phillips, 2002; Liu and Maynard, 2005). Tsay and Chung (2000) found that the usual t-statistic in a regression between two uncorrelated stationary long memory processes (I_{d_1} and $I(d_2)$, $d_1, d_2 \in (0, 0.5)$) diverges, given that both orders of integration sum up to a value greater than 0.5. This divergence of the usual t-statistic is a defining characteristic of a spurious regression. Furthermore, Tsay and Chung (Theorem 4, 2000) showed that the insignificant testing problem first considered by Robinson (1993) could also arise when the integrated order of one long memory process is negative, and that of the other one is positive, because the t-ratio converges to zero, leading to the actual size of using the t-test for the null hypothesis of no correlation between the two unrelated long memory processes being significantly below the nominal size. These results imply that the usual t-statistic could be of no use in empirical testing with long memory processes.

Few previous attempts in practice consider tests that are immune to such persistent behavior, although many studies document regression estimator, and tests have fundamentally different properties in the presence of persistent regressors. Bekaert and Hodrick (2001), and Liu and Maynard (2005) are notable exceptions, nevertheless, there exists a restriction on the testing procedure of Bekaert and Hodrick (2001) and both literature rule out long memory models (see Maynard, 2006). Maynard (2003) and Maynard (2006) suggest the covariance schemes to test for the forward exchange rate hypothesis (FRUH) that are free from the problems associated with persistent regressors. However, the spurious cross-correlation resulted from two time series with imbalanced integrated orders by Wang *et al.* (2020) could be on account of the biased empirical results using the covariance framework.

As a remedy, this paper develops a tool to test for the correlation between two integrated-order imbalanced time series, that allows for persistent behavior; i.e., one series displays a near unit root or stationary long memory process, and the other displays a covariance stationary time series (or I(0) process). Empirically, it is not surprising to test for the relationship between two time series with imbalanced orders. A notable example of this type in international finance is the forward premium puzzle, in which the return of the spot rate exhibits an I(0) process, and the forward discount is near unit root or long memory. In this regard, we generalize Hong's (1996) correlation statistics based on an $AR(k)$ pre-filtered approach to test for two time series with imbalanced integrated-orders.

This study makes the following theoretical and empirical contributions. First, we show that the limiting distribution of Hong's correlation test follows the standard normal distribution when the integrated orders of two time series are imbalanced. Hence, Hong's correlation test is easy to implement. One simply uses an AR model to filter an $I(0)$, $I(d)$, or a differenced near unit root process and then constructs the Hong's correlation test by the AR-filtered residuals. The main advantage of using an AR model to filter a $I(d)$ process is that it does not require the prior knowledge and estimation of an exactly identified $I(d)$ model. It can also substantially reduce the computational burden and inaccurate parameter estimates of the integrated order d. In fact, when d is unknown, the maximum likelihood estimation (MLE) of the fractional parameter in a parametric ARFIMA process is quite time-consuming and is not very accurate in finite samples, especially when d is close to 0.5 and the sample size is small ($T = 100$ and 200). Therefore, simulation methods based on the estimated ds could lead to severe bias in finite samples. Although Sun (2004) proposed a heteroskedasiticity and autocorrelation consistent t-test (HAC-t test) to avoid the spurious regression problem, the limiting distribution of the HAC-t test is non-standard. The critical values thus vary with the use of different kernel functions, fractional parameters (d), and bandwidth parameters (M), which make the application of the HAC-t test difficult. Thus, using Hong's correlation test could avoid the pitfalls associated with an estimation of the fractional parameter and the HAC-t testing procedure.

Second, Monte Carlo simulations demonstrate the desirability of Hong's correlation test in finite samples. Results confirm our theoretical justification. The size-control ability of the Hong's statistics is convincing for various combinations of integrated orders of two time series. In addition, the power performance is very promising relative to that of the HAC-t (Sun, 2004). More importantly, the spurious correlation due to the near unit root or long memory process in the conventional t test is less likely to occur. This situation is particularly noticeable in the forward premium puzzle.

Third, the commonly cited bias in regression analysis, such as the model mis-specifications, or omitted variables in pre-determined functional forms, could be averted in Hong's correlation testing framework. In addition, our testing procedure also allows for examining the cross-correlations over various lags.

Fourth and Finally, as an illustration, we demonstrate the applicability of the Hong correlation testing procedure by examining the time-varying pattern of foreign exchange rate unbiasedness hypothesis (FRUH) for several developed economies from 1996 to 2020 with the rolling window scheme. This long-standing anomaly is of specific interest to us, due to the vulnerability of the conventional regression-based testing procedure to the biased statistical inference, when the forward discount follows an $I(d), d \in (0, 0.5)$, or a near unit root process.

In brief, our testing framework aims to bridge the gap in the current literature on FRUH. For instances, the spurious effect and inaccurate estimate of the long memory parameter d, could both occur with high possibility in the covariance test and two-stage rebalancing estimation for FRUH as in Maynard (2003, 2005) and Maynard *et al.* (2013). Empirical evidence indicates that during or prior to the period of the general global economic events or crises, the strength of the correlation between spot rate return and forward discount for each currency increases, and thus, the unbiasedness hypothesis for several countries would be rejected. Additionally, our correlation tests suggest that the forward premium responds with a time lag to a change in the spot rate. Such an empirical observation can be reconciled by the theoretical argument that investors may react slowly to changes in financial markets on the basis of a limited set of information. Collectively, these findings could provide important information for business practitioners in the portfolio adjustment and risk control.

The rest of the paper runs organized as follows. Section 2 provides the theoretical justification of the test statistics. Section 3 uses Monte Carlo studies to investigate the finite sample properties of Hong's test statistics constructed by the AR-filtered residuals. Section 5 presents an empirical application of Hong's correlation test on the forward rate unbiasedness hypothesis. Concluding remarks are given in Section 5. All proofs are in the Appendix.

2. The Model and Test Statistics

Many researchers and business practitioners are interested in the lead-lag cross-correlations, because the duration of cross-correlation and elucidation of various causalities between two series are quite crucial to signal the pattern of spillover effects in financial markets.

Haugh (1976) proposed the following test statistics for two ARMA processes being independent :

$$S_M = T \sum_{l=-M}^{M} \widehat{\rho}_{ij}(l)^2 \qquad \text{and} \qquad S_M^* = T^2 \sum_{l=-M}^{M} (T-|l|)^{-1} \widehat{\rho}_{ij}(l)^2, \tag{1}$$

where

$$\widehat{\rho}_{ij}(l) = \frac{\sum_{t=k+1}^{T-l} \widehat{e}_{t,k,i}\widehat{e}_{t+l,k,j}}{\left(\sum_{t=k+1}^{T} \widehat{e}_{t,k,i}^2\right)^{1/2} \left(\sum_{t=k+1}^{T} \widehat{e}_{t+l,k,j}^2\right)^{1/2}}, \tag{2}$$

and $\widehat{e}_{t,k,i}$ and $\widehat{e}_{t,k,j}$ are prewhitened $y_{t,i}$ and $y_{t,j}$ processes, derived from residuals of fitting ARMA model for $y_{t,i}$ and $y_{t,j}$, respectively. Haugh (1976) proved that S_M and S_M^* are both asymptotically distributed as χ^2_{2M+1}. Because of the lower power of the Haugh test, Hong (1996) proposed a powerful test of no contemporaneous or lagged correlations between two stationary covariance time series ($I(0)$ processes) $y_{t,i}$ and $y_{t,j}$ based on Haugh's (1976) test statistic. In this study, we further generalize Hong's (1996) test statistics to the case when two time series have different orders of integration; i.e., either $I(0)$, $I(d), d \in (0, 0.5)$, or a nearly unit root process as defined in Assumptions 1-3.

Assumption 1. *The processes y_t follows an ARMA(p,q) processes of the form:*

$$\phi(L)y_t = \theta(L)e_t,$$

where (i) the autoregressive (AR) and moving average (MA) polynomials $\phi(\cdot)$ and $\theta(\cdot)$ in the lag operator L have all roots outside the unit circle; (ii) $\phi(\cdot)$ and $\theta(\cdot)$ have no common roots; (iii) e_t is an i.i.d. process with $\mathbf{E}[e_t] = 0$, $\mathbf{E}[e_t^2] = \sigma_e^2$, and $\mathbf{E}[e_t^4] < \infty$.

Assumption 2. *y_t is generated as:*

$$\phi(L)(1-L)^d y_t = \theta(L)e_t, \tag{3}$$

where (i) $d \in (0, 0.5)$ such that $(1-L)^d = \sum_{j=0}^{\infty} \beta_j L^j$ is the fractional differencing operator; (ii) AR-and MA-polynomials $\phi(L) = \sum_{j=0}^{\infty} \phi(j) L^j$ and $\theta(L) = \sum_{j=0}^{\infty} \theta_j L^j$ have all roots outside the unit circle; (iii) $\phi(L)$ and $\theta(L)$ have no common zeroes; (iv) e_t is an i.i.d. process as defined in Assumption 1.

Assumption 2 here ensures the conditions in Theorem 3 of Hosking (1996) hold and allows an $I(d)$ process, y_t, to be represented by an infinite-order moving average process;

$$y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}, \quad \text{where} \quad \psi_j = O\left(j^{d-1}\right) \quad \text{as} \quad j \to \infty, \tag{4}$$

or an infinite order autoregressive process, $AR(\infty)$:

$$y_t = \sum_{j=1}^{\infty} \beta_j y_{t-j} + e_t, \quad \text{where} \quad \beta_j = O\left(j^{-d-1}\right) \quad \text{as} \quad j \to \infty. \tag{5}$$

Assumption 3. *y_t is generated by the process of the form:*

$$[1-(1+\frac{c}{T})L]y_t = u_t, \quad u_t = e_t + \theta_T e_{t-1}, \quad \theta_T = -1 + \delta/\sqrt{T} \tag{6}$$

where $c < 0$ and e_t is an i.i.d. process as defined in Assumption 1 (e.g., Perron and Ng, 1996 and 1998).

Based on the analysis of Berk (1974), Hong (1996a) suggested two modified test statistics, Q_T and Q_{T^*}:

$$Q_T = \frac{T\sum_{l=1-T}^{T-1} g^2(l/M)\widehat{\rho}_{ij,AR}(l)^2 - S_T(g)}{\{2D_T(g)\}^{1/2}}, \tag{7}$$

and

$$Q_{T^*} = \frac{T\sum_{l=1-T}^{T-1} g^2(j/M)\widehat{\rho}_{ij,AR}(l)^2 - MS(g)}{\{2MD(g)\}^{1/2}}, \tag{8}$$

where the smoothing parameter M is an increasing function of T, $M = M(T) \to \infty$, $M/T \to 0$, as $T \to \infty$,

$$S_T(g) = \sum_{l=1-T}^{T-1} (1-|l|/T)g^2(l/M),$$
$$D_T(g) = \sum_{l=2-T}^{T-2} (1-|l|/T)(1-(|l|+1)/T)g^4(l/M),$$

$$S(g) = \int_{-\infty}^{\infty} g^2(z)dz,$$
$$D(g) = \int_{-\infty}^{\infty} g^4(z)dz,$$

$$\widehat{\rho}_{AR,ij}(l) = \frac{T^{-1}\sum_{t=k+1}^{T-l} \widehat{e}_{t,i,k}\widehat{e}_{t+l,j,k}}{\sqrt{T^{-1}\sum_{t=k+1}^{T} \widehat{e}_{t,i,k}^2}\sqrt{T^{-1}\sum_{t=k+1}^{T} \widehat{e}_{t+l,j,k}^2}},$$

where $y_{t,i}$ and $y_{t,j}$ series are fitted by the AR(k) approximation as follows,

$$\widehat{\phi}_i(L)y_{t,i} = \widehat{e}_{t,i,k} \quad \text{and} \quad \widehat{\phi}_j(L)y_{t,j} = \widehat{e}_{t,j,k}$$

$\widehat{e}_{t,i,k}$ and $\widehat{e}_{t,j,k}$ are AR-filtered residuals and asymptotically mimic $e_{t,i}$ and $e_{t,j}$, correspondingly (see Berk,1974; Poskitt, 2007). In addition, $g(.)$ is a kernel function assigning different weights to various lags and satisfying the following assumption:

Assumption 4. *(i) For all $x \in R$, $|g(x)| \leq 1$ and $g(x) = g(-x)$; $g(0) = 1$; $g(x)$ is continuous at zero and for almost all $x \in R$. (ii) $|g(x) - g(y)| < c|x-y|$ for some $c > 0$. (iii) $|g(x)| - |g(y)| \geq 0$ for $|x| \leq |y|$. (iv) $\int_{-\infty}^{\infty} g^2 dz < \infty$.*

Assumption 4 allows the utilization of some commonly-used kernels such as the Bartlett, Daniell, Parzen, quadratic-spectral (QS), and the truncated kernels (Priestley, 1981). Furthermore, Hong (1996a, 1996b) showed that under some additional conditions on g and M, one can replace $D_T(g)$ by $MD(g)$ without affecting the asymptotic distribution of Q_T. Both Q_T and Q_{T^*} are asymptotically normally distributed under the null hypothesis of the independence and have the same asymptotic power properties. Particularly, the lag order M indicates the lag length is used to characterize the co-movement between two markets, to the extent that we are able to measure the cross-correlation between the financial series with a time horizon of M. In our empirical example, we shall discuss how the forward rate correlates with the spot rate in the context of FRUH under different lag orders.

Hong (1996) suggested the following optimal kernel that maximizes the power of Q_T and Q_{T^*} over a suitable class of kernel functions :

$$g(\tau) = \{g \quad \text{satisfies Assumption 4}, g^2 = 0.5\tau^2, G(\lambda) \geq 0 \quad \text{for} \quad \lambda \in (-\infty, \infty)\},$$

where

$$G(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(z) e^{-iz\lambda} dz.$$

When the characteristic exponent of the function $g(z)$, $m = 2$, where $m > 0$ is the largest positive integer such that

$$g^m := lim_{z \to 0} \frac{1 - g(z)}{|z|^m}$$

exists finite and nonzero.

Assumption 5. *(i) g(.) satisfies Assumption 1, (ii)* $|g(x)| \leq C_1 |x|^{-b}$ *for some* $b > 1 + 1/m$ *and some* $C_1 < \infty$.

Hong (1996) showed that Q_T and Q_{T^*} deliver better power than Haugh's test for two $I(0)$ processes being correlated. We now show that the two test statistics, Q_T and Q_T^*, developed for testing the independence between two $I(0)$ time series, can also be used to test for the independence between two integrated order-imbalanced time series.

3. Asymptotic Properties of Hong Tests

3.1. The AR approximation of a time series

We note that $y_{t,i}$ satisfying either Assumption 1, 2 or 3 can be represented by an infinite-order autoregressive process (AR(∞)):

$$\phi_i(L) y_{t,i} = e_{t,i}, i = 1, 2, \cdot, N, \tag{9}$$

where $\phi_i(L) = \sum_{j=0}^{\infty} \phi_{ji} L^j$ with a lag operator L, and $|\phi_i(L)|$ is bounded away from zero, and $e_{t,i}$ is a sequence of independent identically distributed random variables with zero mean and variance $\mathbf{E}\, e_{t,i}^2 = \sigma_i^2$. Approximating $y_{t,i}$ by an AR(k) process:

$$y_{t,i} = \sum_{j=1}^{k} \beta_{ji}(k) y_{t-j,i} + e_{t,i,k}, \tag{10}$$

where k increases with T (e.g., Berk, 1974; Poskitt, 2007). If $y_{t,i}$ follows Assumption 1, then Berk (1974) shows that if $(k, T) \to \infty$ and $k^3/T \to 0$, then $\widehat{\beta}_{ji}(k) \to \phi_{ji}$ and $E(e_{t,i,k}^2) \to \sigma_i^2$. If $y_{t,i}$ follows Assumption 2, then Poskitt (2007) proves that as long as $k = o((\frac{T}{logT})^{0.5-d})$ [1], when $(k, T) \to \infty$, $\widehat{\beta}_{ji}(k) \to \phi_{ji}$ and $lim_{k\to\infty} E(\widehat{e}_{t,i,k} - e_{t,i})^2 = 0$. If $y_{t,i}$ follows Assumption 3, Perron and Ng (1998) show that the first differences of the data can be approximated by an AR(k) process;

$$\Delta y_{t,i} = \sum_{j=1}^{k} \beta_{ji} \Delta y_{t-j,i} + \epsilon_{t,i,k}, \tag{11}$$

[1] This condition satisfies the model selection criterion for the AR(k) approximation of a long memory process discussed in Section 5 of Poskitt (2007).

where $\Delta y_{t,i} = y_{t,i} - y_{t-1,i}$. In addition, the estimate of residual variance resulting from the AR-fitted model converges to the true error variance as $k, T \to \infty$ and $k^{11}/T \to 0$.

3.2 The AR-filtered version of sample correlation coefficient

As shown in Wang *et al.* (2020 a) and Wang *et al.* (2020 b), $\widehat{\rho}_{ij}$ does not necessarily have the limiting distribution for the cases when the integrated orders of two time series are imbalanced, i.e., one time series is an $I(d_i), d_i \in (0, 0.5)$ process and the other is an $I(0)$ process or an $I(d_j), d_j \in (0, 0.5)$, $d_i \le d_j$. Therefore, following Haugh (1976) and Hong (1996), instead of testing the correlation between $y_{t,i}$ and $y_{t,j}$, we propose to consider the correlation of the two individually AR-filtering residuals of $y_{t,i}$ and $y_{t,j}$, i.e., $e_{t,i}$ and $e_{t,j}$.[2] The intuition behind this procedure is that if two series, $y_{t,i}$ and $y_{t,j}$ are independent, so are $e_{t,i}$ and $e_{t,j}$. On the other hand, if there exists a common shock between $e_{t,i}$ and $e_{t,j}$, then $y_{t,i}$ and $y_{t,j}$ are correlated. Filtering two series by the AR scheme would not alter the independence of the filtered time series, but could lead to correct asymptotic distribution of the correlation coefficients.

We then demonstrate the Pearson sample correlation coefficient based on the residuals of two individual AR approximations. The properties of the AR-filtering version of Pearson sample correlation coefficient $\widehat{\rho}_{ij,AR}$ for two imbalanced-order time series, either $I(0)$, $I(d), d \in (0, 0.5)$ or nearly unit root process[3] are summarized as follows.

Lemma 1. *Suppose $y_{t,i}$ follows an ARMA process defined as Assumption 1 and $y_{t,j}$ follows either a stationary long memory process or a nearly unit root process as defined by Assumptions 2 and 3, respectively. Under the null hypothesis of the independence between $y_{t.i}$ and $y_{t,j}$, as $(k_1, k_2, k_3, T) \to \infty$,*

1. $\sqrt{T}\widehat{\rho}_{ij,AR} \xrightarrow{d} N(0,1)$, when (i) $y_{t,j}$ follows a nearly unit root process; (ii) $k_1 = o(T^{1/3})$ is the lag length for the AR(k_1) approximation of an $I(0)$ process; (iii) $k_2 = o(T^{1/11})$ is the lag length for the AR(k_2) approximation of a differenced nearly unit root process.

2. $\sqrt{T}\widehat{\rho}_{ij,AR} \xrightarrow{d} N(0,1)$, when (i) $y_{t,j}$ follows an $I(d)$, $d \in (0, 0.5)$ process; (ii) $k_1 = o(T^{1/3})$ is the lag length for the AR(k_1) approximation of an $I(0)$ process; (iii) $k_3 = o((T/logT)^{0.5-d})$ is the lag length for the AR(k_3) approximation of an $I(d), d \in (0, 0.5)$ process.

Lemma 2. *If $y_{t,i}$ and $y_{t,j}$ follow a stationary long memory process and a nearly integrated process as defined by Assumptions 2 and 3, respectively, then, under the null hypothesis of no cross-sectional dependence between $y_{t.i}$ and $y_{t,j}$, as $(k_1, k_2, T) \to \infty$,*

1. $\sqrt{T}\widehat{\rho}_{ij,AR} \xrightarrow{d} N(0,1)$, where (i) $k_1 = o((T/logT)^{0.5-d})$ is the lag length for the AR(k_1) approximation of an $I(d), d \in (0, 0.5)$ process; (ii) $k_2 = o(T^{1/11})$ is the lag length for the AR(k_2) approximation of a differenced nearly unit root process.

Lemma 3. *When $y_{t,i}$ follows an $I(d_i)$, $d_i \in (0, 0.5)$ process, $y_{t,j}$ follows an $I(d_j)$ process, $d_j \in (0, 0.5)$ then as $(k_1, k_2, T) \to \infty$ and under the null hypothesis of the independence between $y_{t,i}$ and $y_{t,j}$, with $k_1 = o((T/logT)^{0.5-d_i})$ for an $I(d_i) \in (0, 0.5)$ process and $k_2 = o((T/logT)^{0.5-d_j})$ for an $I(d_j) \in (0, 0.5)$*

[2] This framework is inspired by Haugh (1976) and Hong (1996) which consider to use $e_{t,i}$ and $e_{t,j}$ to examine the correlation between $y_{t,i}$ and $y_{t,j}$. In other words, using individually constructed AR model to filter $y_{t,i}$ and $y_{t,j}$ does not change the correlation or independence between $y_{t,i}$ and $y_{t,j}$.

[3] When one of a pair follows a nearly unit root process, one could take the first difference on this nearly unit root process.

process, where k_1 and k_2 are lag lengths for the AR(k) approximation of the $I(d_i)$ and $I(d_j)$ processes, respectively, we then have the following results:

1. $\sqrt{T}\widehat{\rho}_{AR,ij} \xrightarrow{d} N(0,1)$, *when $y_{t,j}$ follows an $I(d_j)$, $d_j \in (0, 0.5)$ process and $(d_i^2+d_i)+(d_j^2+d_j)-1<0$.*

Remark 1. Under the null hypothesis of the independence between $y_{t,i}$ and $y_{t,j}$, Wang *et al.* (2020) suggest that inference based on $\sqrt{T}\widehat{\rho}_{ij}$ without the AR-filtered version of sample correlation coefficient between $y_{t,i}$ and $y_{t,j}$ could lead to the spurious correlation. Theorems 1-3 establish the consistency of AR-filtering version of Pearson sample correlation coefficient ($\widehat{\rho}_{ij,AR}$) and its limiting distribution following a standard normal distribution provided that the order k of AR(k) approximation satisfies the given conditions. Thus, we can construct the conventional Pearson correlation coefficients from the residuals of $AR(k)$ model and derive the limiting properties of the usual correlation-related tests using the analysis of $\widehat{\rho}_{ij,AR}$ when imbalanced integrated orders of two time series occur.

3.3. Asymptotic Properties of Hong's Statistics

Theorem 1. *Let $y_{t,i}$ and $y_{t,j}$ satisfy Assumption 1 and 2, respectively. Suppose the kernel functions with $r=2$ satisfy Assumption 4 and 5. When the smooth parameter $M=M(T)\to\infty$, $M/T\to 0$, as $k_1,k_2,T\to\infty$, both Q_T and Q_T^* are asymptotically distributed as $N(0,1)$ if $e_{t,1}$ is independent of $e_{s,2}$ for all t and s, where (i) $k_1=o(T^{1/3})$ is the lag length for the AR(k_1) approximation of an $I(0)$ process; (ii) $k_2=o((T/logT)^{0.5-d_j})$ is the lag length for the AR(k_2) approximation of an $I(d_j), d_j\in(0,0.5)$ process.*

Theorem 2. *Let $y_{t,i}$ and $y_{t,j}$ satisfy Assumption 1 and 3, respectively. Suppose the kernel functions with $r=2$ satisfy Assumptions 4 and 5. When the smooth parameter $M=M(T)\to\infty$, $M/T\to 0$, as $k_1,k_2,k_3,T\to\infty$, both Q_T and Q_T^* are asymptotically distributed as $N(0,1)$ if $e_{t,1}$ is independent of $e_{s,2}$ for all t and s, where (i) $k_1=o(T^{1/3})$ is the lag length for the AR(k_1) approximation of an $I(0)$ process; (ii) $k_2=o(T^{1/11})$ is the lag length for the AR(k_2) approximation of a differenced nearly unit root process.*

Theorem 3. *Let $y_{t,i}$ and $y_{t,j}$ satisfy Assumption 2 and 3, respectively. Suppose the kernel functions with $r=2$ satisfy Assumptions 4 and 5. When the smooth parameter $M=M(T)\to\infty$, $M/T\to 0$, as $k_1,k_2,T\to\infty$, both Q_T and Q_T^* are asymptotically distributed as $N(0,1)$ if $e_{t,1}$ is independent of $e_{s,2}$ for all t and s, where (i) $k_1=o(T^{1/11})$ is the lag length for the AR(k_1) approximation of a differenced nearly unit root process; (ii) $k_2=o((T/logT)^{0.5-d_j})$ is the lag length for the AR(k_2) approximation of an $I(d_j), d_j\in(0,0.5)$ process.*

Theorem 4. *Let $y_{t,i}$ and $y_{t,j}$ satisfy $I(d_i)$ and $I(d_j)$, $d_i,d_j\in(0,0.5)$ as defined in Assumption 3. Suppose the kernel functions with $r=2$ satisfy Assumptions 4 and 5. When the smooth parameter $M=M(T)\to\infty$, $M/T\to 0$, as $k_1,k_2,T\to\infty$, both Q_T and Q_T^* are asymptotically distributed as $N(0,1)$ if $e_{t,1}$ is independent of $e_{s,2}$ for all t and s. where (i) $k_1=o((T/logT)^{0.5-d_i})$ is the lag length for the AR(k_1) approximation of an $I(d_i)\in(0,0.5)$ process; (ii) $k_2=o((T/logT)^{0.5-d_j})$ is the lag length for the AR(k_2) approximation of an $I(d_j)\in(0,0.5)$ process.*

Theorems 1-4 shows that Q_T and Q_{T^*} constructed from the residuals of ever-increasing order AR models are asymptotically normally distributed under the null hypothesis of the independence between two integrated-order imbalanced time series. Although three AR(k) models with ever-increasing orders, $k=o(T^{1/3})$, $k=o((T/logT)^{0.5-d})$ and $k=o(T^{1/11})$, are sufficient to establish the asymptotic distributions of

Hong's statistics, our interest is not in getting consistent estimate of β_j, but rather in obtaining fitted residuals that are close to white noises, often a finite order AR(k) process is sufficient, because $(y_{t-1}, y_{t-2}, \cdots, y_{t-k})$ are collinear. When regressors are correlated, dropping collinear variables may bias the estimate of β_j, but it does not affect the prediction and hence the estimated residuals.

The circumstances discussed in Theorems 1 and 2 are commonly found in empirical applications such as the forward discount anomaly proposed in Baillie and Bollerslev (2000), Maynard and Phillips (2001), and Liu and Maynard (2005). The common characteristic of these examples is the regression model including a stationary $I(d)$ or nearly unit root regressor (forward premium) and an $I(0)$ dependent variable (spot return). Theorem 4 illustrates the empirical finding of the FRUH in Maynard *at al.* (2013), where both the spot rate return and forward discount display $I(d)$ processes.

Remark 2. *Using an ever increasing AR(k) model to transform the residuals of $y_{t,i}$ into white noise residuals as $(k,T) \to \infty$, Berk (1974) showed that if $y_{t,i}$ is an $I(0)$ process, $k = o(T^{1/3})$ and Poskitt (2007) showed that if $y_{t,i}$ is an $I(d)$ process, $k = o((T/logT)^{0.5-d})$. When $y_{t,i}$ is a differenced nearly unit root process, $k = o(T^{1/11})$. In implementation of our adjusted AR(k) filtered correlation coefficients, we suggest to use a common bound, i.e, $k^{**} = o((T/logT)^{0.5})$, for an AR($k^{**}$) model to approximate the $I(0)$, $I(d)$, and differenced processes for Q_T and Q_T^*.*

4. Order Selection

Lemmas 1-3 and Theorems 1-4 establish the asymptotic properties of Q_T and Q_T^* based on the AR-filtered residuals of $I(0)$, nearly unit root and $I(d)$ processes. However, we only have finite T observations for the observed series. We thus encounter an issue of how to select the order of an AR approximation based on T observations.

The data generating process (DGP) of y_t under our Assumptions 1-2 and 3 can be represented by

$$y_t = \sum_{j=1}^{\infty} \beta_j y_{t-j} + \epsilon_t \tag{12}$$

and

$$\Delta y_t = \sum_{j=1}^{\infty} \beta_j \Delta y_{t-j} + \epsilon_t, \tag{13}$$

respectively,[4] where ϵ_t is i.i.d with mean 0 and variance σ^2. Thus, before computing the correlation between two series, we suggest first using an AR filter, $B_h(L) = (1 - \beta_{1h}L - \beta_{2h}L^2 - \cdots - \beta_{hh}L^h)$ to transform y_t into a white noise process

$$\epsilon_{th} = \beta_h(L)y_t = (y_t - \sum_{j=1}^{h} \beta_{jh} y_{t-j}), \tag{14}$$

where $h = 1, \cdots, M_T$, M_T increases with T while $\frac{M_T}{T} \to 0$ as $T \to \infty$. Let $\sigma_h^2 = \frac{1}{T}\sum_{t=1}^{T} \epsilon_{th}^2$ and k_T be the order of the AR filter such that

$$\sigma_{k_T}^2 = min(\sigma_1^2, \cdots, \sigma_{M_T}^2). \tag{15}$$

Let $\widehat{\sigma}_h^2 = \frac{1}{T}\sum_{t=1}^{T} \widehat{\epsilon}_{th}^2$, where

$$\epsilon_{th} = y_t - \widehat{\beta}_{1h} y_{t-1} - \cdots - \widehat{\beta}_{hh} y_{t-h}, \tag{16}$$

[4] Equation (13) is considered in Perron and Ng (1998) as the equation (A.1) of Perron and Ng (1998) on page 585.

and β_{jh} is the least square estimate of β_{jh}.[5] The popular order selection criterion takes the form

$$SC_T(h) = log\widehat{\sigma}_h^2 + \frac{hC_T}{T}, \tag{17}$$

where $C_T > 0$, $\frac{C_T}{T} \to 0$ as $T \to \infty$. If $C_T = 2$, we have AIC (see Akaike, 1970). If $C_T = logT$, we have BIC (see Schwartz, 1978).

Let $\widehat{\sigma}^*_{k_T}$ be the estimate of variance constructed by the residuals from the AR($\widehat{k}^*_T$) approximation of y_t based on the selection criterion (17). Shibata (1980) defined a sequence of selected orders $\widehat{k}^*_T$ as efficient if the limit of sequence $lim_{T\to\infty}\frac{SC_T(\widehat{k}^*_T)}{SC_T(\widehat{k}_T)} = 1$ as $T \to \infty$. Poskitt (2007) showed that if y_t is a stationary long memory $I(d)$ process satisfying Assumption 2, where $0 < d < 0.5$, then letting $C_T = 2$ (AIC) is the efficient condition of Shibata (1980). When y_t is a covariance stationary as defined in Assumption 1 or a differenced nearly unit root process as defined in Assumption 3, $\widehat{\beta}_h$ converges to β_h at the speed of faster rate relative to that of a stationary long memory (see Berk, 1974; Perron and Ng, 1998; Poskitt, 2007), thus, using AIC also satisfies Shibata (1980) efficiency condition, for that being $\frac{\widehat{k}^*_T}{k_T} = 1 + o(1)$. We shall therefore using AIC to select the order of AR to filter y_t process.[6] [7]

5. Monte Carlo Simulations

Monte Carlo experiments are conducted to examine finite sample properties of our analytical results. The Monte Carlo experiment for each case is based on 2,000 replications with sample size T. For each T, we generate $T + 200$ observations and discard the first 200 to reduce the effects of initial values. We simulate a pair of time series $y_{t,1}$ and $y_{t,2}$ generated as the following combinations of DGPs and assume that there is no cross correlation between them, i.e., $\rho_{e_{t,1},e_{t,2}} = 0$ (see Haugh, 1976)[8] are independent (see Hong, 1996).

$$\begin{aligned}
\text{DGP (a).} \quad & (1+0.7L)y_{t,1} = (1+0.5L)e_{t,1}, \quad (1+0.95L)y_{t,2} = (1-0.5L)e_{t,2},\\
\text{DGP (b).} \quad & (1+0.7L)y_{t,1} = (1+0.4L)e_{t,1}, \quad (1+0.8L)(1-L)^{0.45}y_{t,2} = (1+0.2L)e_{t,2},\\
& \text{where} \quad d_{y_{t,2}} = \{0.1, 0.2, 0.3, 0.4, 0.45\}.\\
\text{DGP (c).} \quad & (1-0.95L)y_{t,1} = (1-0.5L)e_{t,1}, \quad (1-L)^{d_{y_{t,2}}}y_{t,2} = e_{t,2},\\
& \text{where} \quad d_{y_{t,2}} = \{0.1, 0.2, 0.3, 0.4, 0.45\}.\\
\text{DGP (d).} \quad & (1+0.8L)(1-L)^{d_{y_{t,1}}}y_{t,1} = (1+0.2L)e_{t,1}, \quad (1-L)^{d_{y_{t,2}}}y_{t,2} = e_{t,2},\\
& \text{where} \quad d_{y_{t,1}} = \{0.33, 0.36, 0.38, 0.4, 0.44\}, d_{y_{t,2}} = \{0.33, 0.36, 0.38, 0.4, 0.44\}.
\end{aligned}$$

Here, $e_{t,1}$ and $e_{t,2}$ are identically and independently distributed $N(0, 1)$ random variables. DGP (a) represents the combination of two stationary time series, but one of them follows a nearly unit root process. DGP (b) indicates a combination of a stationary $I(d)$ process and an I(0) process. DGP (c) presents the two integrated-order imbalanced time series; i.e., one of them follows a nearly $I(1)$ process and the other exhibits a stationary long memory process. DGP (d) displays two stationary long memory processes with

[5] For ease of notation, we let $y_{-j} = 0$ for $j = 1, 2, \cdots, M_T$.

[6] In fact, Mellow's criterion also satisfies the Shibata (1980) condition.

[7] Although BIC is known to be consistent, Schmidt and Tschernig (1993) found that the BIC criterion sometimes performs poorly. When the true model is of infinite dimension, AIC yields the best finite dimensional approximating model (e.g., Shibata (1981), Ng and Perron (2005), Beran (1995) and Poskitt (2007)).

[8] It is well known that $y_{t,1}$ and $y_{t,2}$ are independent if and only if the innovations $e_{t,1}$ and $e_{t,2}$

different differencing parameters. It should be noted that DGPs (a), (b), and (c) are widely discussed in the FRUH literature, such as Maynard (2003) and Choi *et al.* (2010).

4.1. Finite sample properties of Hong's statistics for DGPs (a)-(c)

We report the finite sample performance of Hong's statistics for DGPs (a)-(c) in Tables 1-2. To save space, only the simulation results for $T = 200$ are displayed in this paper, but the results for the other T are available from the authors upon request. We follow the simulation design of Hong (1996). Two commonly-used kernels, Daniell and Parzen, having $D(k) = 1.209200/\tau$ and $1.325414/\tau$ are considered. Three rates of M are $[ln(T)], [3T^{0.2}]$ and $[3T^{0.3}]$, where $[a]$ denotes the integer part of a. For $T = 200$, these rates deliver $M = 5, 9, 15$. We use AIC to choose an order k of the AR(k) approximation of each time series. Furthermore, Hong (1996a, 2001) pointed out that both Q_T and Q_{T^*} diverge to positive infinity in probability as $T \to \infty$ under a general class of alternatives, which implies that asymptotically, negative values of Q_T and Q_{T^*} occur only under the null hypothesis. Therefore, Q_T and Q_{T^*} could be one-sided test, and the upper-tailed $N(0, 1)$ critical value should thus be used in our simulation. Tables 1-3 examine the size and power of Hong's Q_T and Q_{T^*} tests with the Parzen kernel at the 5% significant level, respectively, where $y_{t,1}$ and $y_{t,2}$ follows the DGPs (a)-(c). Results for the case where the Daniell kernel is used for constructing two tests are available on request. Results show that the empirical sizes are very close to 5% for every pair of time series with imbalanced orders where the two processes are independent with each other no matter which kernel is used. The nominal size is close to the actual size for Hong's two test statistics for DGPs (a)-(c).

We further assess the power of Hong's statistic when many financial time series share a similar pattern of short cross-correlations, as described in Hong (1996). We assume that the error term $e_{t,1}$ and $e_{t,2}$ are generated as: $\rho_{12}(j) = 0.2$ for $j = 0$ and $\rho_{12}(j) = 0$ for $j \neq 0$, where $\rho_{12}(j)$ denotes the cross correlation function of e_{1t} and e_{2t} at lag j. The results in Tables 1-2 show that the rejection percentages of 1000 replications at the 5% significance level are similar to those of Table 2 in Hong (1996 a), where the two time series are short memory processes. Both Q_T and Q_{T^*} perform similarly for the two kernels considered. For each kernel, the more slowly M grows, the better is the power of the test. On the other hand, the power performance of Hong's statistics does not depend on the long memory or nearly $I(1)$ characteristics of the data series, as long as the data series themselves are stationary. We also compare the power performance of Hong's statistic by considering two cross correlated long memory processes with imbalanced integrated orders. We assume that $y_{t,1}$ and $y_{t,2}$ are generated with $\rho_{y_{12}}(j) = 0.2$ for $j = 0$ and $\rho_{y_{12}}(j) = 0$ for $j \neq 0$, where $\rho_{y_{12}}(j)$ denotes the cross correlation function of $y_{t,1}$ and $y_{t,2}$ at lag j. The power patterns of the results yielded by the Paren kernel are similar to those obtained by the Daniell kernel, when the significant level is at 5%. The results of Theorems 7 and 8 are clearly supported in our simulation studies. Simulation evidence indicates that the generalizations of both Hong tests to two integrated-order imbalanced time series are applicable.

4.2. Finite sample Properties of Hong's Statistics for DGP (d)

We summarize the finite sample performance of Hong's statistics for DGP (d) in Tables 3-4. We as well use AIC to choose an order k of the AR(k) approximation of each time series. Tables 3 and 4 show the size and power of the Hong's tests Q_T and Q_{T^*} with the Parzen kernel at the 5% significant level, respectively, where both $y_{t,1}$ and $y_{t,2}$ follow I(d) processes. The results show that the empirical sizes are very close to 5% for every pair of d_1 and d_2 where the two processes are independent, regardless of the kernel functions. The investigation of the power of the Hong's statistic for DGP (d) is reported in Table 5. The results in Table 4 indicate that the rejection percentage of 1000 replications at the 5% significant level is also similar to

Table 2 in Hong (1996), where the two time series are short memory processes. Both Q_T and Q_{T^*} perform similarly for the two kernels considered. For each kernel, the more slowly M grows, the better is the power of the test. On the other hand, the power performance of the Hong's statistics does not depend on the long memory characteristics of the data series, as long as the processes are stationary. We also compare the power performance of the Hong's statistic by considering two cross-correlated long memory processes. We further assume that $y_{t,1}$ and $y_{t,2}$ are generated with $\rho_{y_{12}}(j) = 0.2$ for $j = 0$ and $\rho_{y_{12}}(j) = 0$ for $j \neq 0$ where $\rho_{y_{12}}(j)$ denotes the cross-correlation function of $y_{t,1}$ and $y_{t,2}$ at lag j. The power patterns in Table 4 are similar to those presented in Tables 1-2 when the significance level is set at 5%. The results of Theorem 4 are clearly supported by our simulation studies, demonstrating that the Hong's statistics derived from prewhitening stationary $I(d)$ processes by the AR(k) approximation behave well in finite samples.

In a comparison, we present the conventional procedure of examining the correlation between two stationary long memory processes, $y_{t,1}$ and $y_{t,2}$, that is, regressing $y_{t,1}$ on $y_{t,2}$,

$$y_{t,1} = \alpha + \delta y_{t,2} + u_t, \quad t = 1, \cdots, T,$$

where we first estimate the fractional parameter, d of two long memory processes by the maximum likelihood method, and then construct the usual t-ratio,

$$t = (\widehat{\delta} - \delta)/se(\widehat{\delta}),$$

to test whether $\delta = 0$ or not, where $\widehat{\delta}$ is the OLS estimate of δ and $se(\widehat{\delta})$ is the standard error of $\widehat{\delta}$. The results are reported in Table 5. Generally, when the value of the fractional parameter d increases, the bias of the maximum likelihood estimates of d becomes larger (e.g., Sowell, 1992). At the 5% significance level, t-test may exhibit significant size distortions. As a result, the usual t-ratios without proper filtering may induce spurious correlations.

Consequently, we further consider the HAC-t statistics proposed by Sun (2004). The HAC-t test is defined as HAC-$t = (\widehat{\delta} - \delta)/\widetilde{\sigma}$, where $\widetilde{\sigma}$ is the HAC estimator given by

$$\widetilde{\sigma}^2 = \left(\sum_{t=1}^{T} (y_{t,2} - \bar{y_2})^2 \right)^{-1} T\widehat{\Sigma} \left(\sum_{t=1}^{T} (y_{t,2} - \bar{y_2})^2 \right)^{-1},$$

where

$$\widehat{\Sigma} = \sum_{j=-T+1}^{T-1} k(j/M')\widetilde{\Gamma}(j),$$

$$\widetilde{\Gamma}(j) = \begin{cases} \frac{1}{T} \sum_{t=1}^{T-j} (y_{t+j,2} - \bar{y_2})\widehat{u}_{t+j}\widehat{u}_t(y_{t,2} - \bar{y_{t,2}}) & \text{for} \quad j \geq 0, \\ \frac{1}{T} \sum_{t=-j+1}^{T} (y_{t+j,2} - \bar{y_2})\widehat{u}_{t+j}\widehat{u}_t(y_{t,2} - \bar{y_{t,2}}) & \text{for} \quad j < 0, \end{cases}$$

$k(.)$ is a kernel function, and $M' = bT, b \in (0, 1]$ is the bandwidth parameter. We follow the Sun's (2004) simulation procedure to obtain the limiting distribution of the HAC-t test. The simulation results of Sun (2004) show that the limiting distribution becomes close to the standard normal when b is close to 0.1 and $k(.)$ is the Bartlett kernel with b being 0.1 (e.g. Sun, 2004, p. 955-956). We also use 5000 replications and a sample size of $T = 200$ to calculate the 95% quantile of the limiting distribution of $\widetilde{t}$ for different (d_1, d_2) combinations. The results for all combinations of d_1 and d_2 in this study are qualitatively similar. The corresponding 95% quantiles for different (d_1, d_2) combinations are given in Table 6. Table 7 illustrates the size and power performance of the HAC-t test, when $\rho_{y_{12}}(j) = 0.2$ for $j = 0$ and $\rho_{y_{12}}(j) = 0$ for $j \neq 0$. It appears that although the HAC-t test eliminates the spurious effect presented in Table 7, the power of the

HAC-t is lower than that of the Hong's tests when DGPs are long memory processes. Additionally, Hong's Q_T and Q_T^* can be used to measure the cross-correlations with various lags (not just $j = 0$), and tend to have better power than the HAC-t test, which focuses exclusively on instantaneous correlation between two long memory series. All the simulation results have confirmed that the Hong's test based on the AR(k) approximation is a very useful tool for examining the interactions between two stationary long memory processes.

6. An Application to the Forward Premium Anomaly

6.1. Forward rate unbiasedness hypothesis

The forward premium, or forward discount puzzle, has been a long standing issue in international finance, and is of great practical relevance to financial decisions (e.g., Taylor, 2003). The puzzle or the anomaly comes from the empirical failure of the forward rate unbiasedness hypothesis (hereafter "FRUH"). The consensus about its failure is based, largely, on the failure of the regression of the change in the future spot rate on the current forward premium to produce unity slope coefficients. Very often, researchers find that this slope coefficient is statistically negative, implying that forward rates are biased predictors of future spot rates. In a high-cited study, for example, Froot and Thaler (1990) reported that the average of the estimated slope coefficient is -0.88 across 75 studies. To some extent the variations in the forward discount anomaly across studies can be attributed to model specifications, data availability, sample period of interest, and econometrics analysis. To date, however, the empirical quest for the FRUH has not been fully settled.

Under this backdrop, recent studies have suggested that the anomaly is exaggerated and simply a statistical artifact because of improper treatments of the FRUH test regression. To be specific, this regression itself is not balanced, in the sense that the orders of integration of the dependent variable and the regressors are different. For example, Baillie and Bollerslev (1994) and Baillie and Bollerslev (2000) noted that the long memory process of the forward discount could explain the anomaly when the return of the spot rate displays an $I(0)$ process. Similar results are also provided in Maynard and Phillips (2001) and Maynard *et al.* (2013). Moreover, when both the forward discount and return of spot rate follow $I(d), d \in (0, 0.5)$ processes, the test result of FRUH may be spurious (e.g., Maynard *et al.*, 2013). In other words, with traditional regression-based tests, the validity of the FRUH could be called into question owing to the misleading statistical inference caused by improper statistical treatment. We therefore adopt a simple, intuitive, and straightforwardly implementable approach to the forward rate anomaly.

In contrast to regression-based tests, we consider to test for the forward rate unbiasedness directly using the generalized Hong (1996a) correlation tests as this hypothesis may be rewritten as a test of correlation between excess spot returns and the lagged forward premium. The most appealing advantage of this extension is that the AR-filtered correlation tests can accommodate simultaneously the effects by the imbalanced and spurious regressions (Wang *et al.*, 2021). Our proposed procedure is robust to the persistent behavior of the forward premium, and the literature on test for the FRUH using a correlation test remains unexplored. In particular, long memory in the forward premium has been widely documented and is critical to the validity of standard tests of the FRUH. In this study, we particularly emphasize the use of the generalized Hong (1996a) correlation testing procedure as it does not require us to model explicitly the extent, or even the nature of the persistence in the forward premium. Accordingly, we can avoid the debate about characterizations of the persistence in the forward premium, even though our method remains valid in the presence of a long memory forward premium. The details of the correlation testing procedure for the FRUH are as follows.

Let the current forward exchange rate be denoted as f_t and the next period spot exchange rate be denoted as s_{t+1}. FRUH (Fama, 1984) implies:

$$s_{t+1} = f_t + \epsilon_t,$$

where $E(\epsilon_t|f_t) = 0$. Hence, a test of FRUH is often put in the regression framework (e.g., Maynard *et al.*, 2013),

$$s_{t+1} - f_t = \alpha + \beta(f_t - s_t) + \epsilon_{t+1}.$$

The null hypothesis is:[9]

$$H_0 : \alpha = \beta = E_t(\epsilon_{t+1}) = 0.$$

However, generally $(s_{t+1} - f_t)$ is a short memory process $(I(0))$ while $f_t - s_t$ displays a long memory process (e.g., Maynard and Phillips, 2001; Choi and Zivot, 2007; Maynard *et al.*, 2013).

Instead of regressing current $(s_{t+1} - f_t)$ on $(f_t - s_t)$, we use the generalized Hong's test statistics to determine whether $(s_{t+1} - f_t)$ are correlated with the current and past $(f_t - s_t)$. Moreover, our proposed test is useful not only to test contemporaneous correlations, but also to identify past or future correlations which can reveal a more interesting pattern of market interactions across various lags.

6.2. Empirical analysis

All exchange rates considered in this study are taken from Bloomberg. We use the log spot and 1-month log forward exchange rates at monthly frequency from June 1996 to October 2020. These exchange rates are end-of-month national currency units per US dollar quoted by the arithmetic average of the bid and ask rates for 5 advanced countries: Australia (AUD), Canada (CAD), Japan (JPY), Swiss (CHF), and the United Kingdom (GBP). The inclusion of these countries provides a diverse, yet comparable landscape for examining the validity of FRUH. First of all, most of these countries experienced the currency or financial crises in the 1990s and 2000s when the role of default or counterpart risks became increasingly important in affecting the parity. In particular, since 2008 it has been prominent that short rates have effectively hit the zero interest rate bound, which prompted a sequence of research on the impact of narrowing interest rate differentials: see Chinn and Quayyum (2013), Du *et al.* (2018), Burnside (2019), and Bussiere *et al.* (2019), among others. Second, the data analyzed in this paper span through the period of the COVID-19 pandemic. Uncertainty and anxiety over the economic fallout induced by COVID-19 have roiled global financial markets and raised the prospect of far-reaching, unintended effects on the world economy and investors (see Baker *et al.* and references therein). It is worthwhile to examine whether COVID-19 has any effects on the behavior of exchange rates and thus the validity of FRUH.

We summarize the descriptive statistics of the exchange rate data in Table 8. There are several key features. First, the correlation coefficient between excess spot return $(s_{t+1} - f_t)$ and the forward premium $(f_t - s_t)$ shows that $(s_{t+1} - f_t)$ negatively correlates with $(f_t - s_t)$ for all major currencies except for JPY with a negligibly small positive correlation.[10] Second, the first-order autocorrelation coefficient (ρ) of the forward premium is high (ranging from 0.971 for AUD to 0.912 for GBP). Hence, the forward premium for the industrial countries exhibits persistence. Finally, results (the last column) also show $s_{t+1} - f_t$ and $f_t - s_t$ are integrated of different orders as the former is less persistent while the latter exhibits non-stationary long memory with $d > 0.5$. We thus examine the correlation between the excess spot return and forward premium

[9] This is equivalent to testing for $\beta = 1$ in the "Fama" regression that regresses the expected change in the exchange rate on the forward premium.

[10] It is interesting to note that the correlation for JPY turns negative if we excluded the period of 2020.

for each exchange rate market by the Hong's tests, in order to validate the predictive ability of the forward premium on the excess spot return.

Because Hong (1996a) demonstrated that two kernels, the Parzen and Daniell kernels, deliver similar performance in finite samples, we only report the results of the Parzen kernel-based tests Q_T for brevity. In addition, FRUH is tested in five major currencies based on the Parzen kernel-based test statistic Q_T with $M = 3, 6, 9, 12$. The analysis is performed in the rolling-window framework using a window size $W = 48$ [11]. The order k of the AR(k) approximation for each exchange rate is selected by the AIC criterion.

Figures 1-5 illustrate the rolling correlation patterns between $(s_{t+1} - f_t)$ and $(f_t - s_t)$ for all currencies. There are several striking features by plotting the test statistics over time. First, it exhibits time-varying co-movements between excess return and the forward premium, though with varying degrees of statistical significance (cutoff value for the one-tailed t-test is 1.645). This is consistent with empirical evidence that the predictability coefficient is unstable over time (e.g., Bansal, 1997; Baillie and Chang, 2011; Chinn and Meredith, 2005). Second, our present results complement previous findings by showing that the parity condition is less likely to hold in the 2000s, particularly during periods of high financial instability as with the early 2000s or the during the global financial crisis. Third, our filtered correlation test can detect the location of every turning point where the spot-forward parity changes from an upward trend to a downsize trend when there is growing uncertainty about the financial markets, in particular, in the midst of the 2020 COVID-19 pandemic.

Figures 1-2 show the correlation patterns of the most commonly cited funding currencies, JPY and CHF. It is clear that the correlation for both hedging currencies is less likely to be of the same order of magnitude prior to and in the aftermath of each global event. Regardless of the choice of M, the value of the correlation test statistic of Q_T for CHF is greater than 1.645 from the end of 2008 through 2015, covering the 2008 subprime crisis and European debt crisis. On the other hand, our AR-filtered test statistics for JPY indicate the unbiasedness hypothesis is rejected more frequently in the 2000s, even though the strength of the rejection is tapering off sporadically. In fact, our AR-filtered test statistics indicate the unbiasedness hypothesis is more likely to be rejected when the yen carry trade gained its popularity throughout the 2000s, although the yen carry trade with the US dollar took a brief hiatus in the late 2005 leading up to the global financial crisis. Starting in the late 2010s, our proposed test indicates that there is strong evidence to refute the validity of FRUH for both hedging currencies. The rejection was particularly evident for both CHF and JPY when the financial markets were hit hard by the recent COVID-19 pandemic as increased demand to hedge exchange rate risks in countries' net foreign asset positions with forward contracts, leading to misalignment in spot and forward exchange rates (Liao and Zhang, 2020). We note that our results from the AR-filtered tests are consistent with Chinn and Quayyum (2013) in which they found the evidence for FRUH became weaker for both the franc and the yen from the mid-1990s onward (up to 2012) in a time of extraordinarily low interest rates when the major central banks were actively pursuing a zero-bound-interest-rate policy. In addition, it is observed that the correlations tend to increase or suddenly jump prior to the aforementioned crisis. That is, by testing the FRUH proposition for both of which are widely considered "safe haven" currencies, JPY and CHF, our proposed method provides a reasonably good means to detect the recent financial crises, which is highly correlated with sudden shift of test statistics.

As shown in Figure 3, the unbiasedness hypothesis for GBP is strongly rejected by our AR-filtered test statistics in the wake of the historical economic downfalls including the 2001-02 dot-com bubble bursts, the

[11] Rossi and Inoue (2012) indicate the use of different window sizes may lead to different empirical results in practice. Thus, we also consider the case of $W = 60$. The resulting results are similar to those based on $W = 48$. Results are available on request.

2009 global financial crisis, and the ensuing European debt crisis. In particular, several abrupt shifts in patterns of correlation took place in the early 2000s and during the periods of 2006-2007 and 2012-2013. We also observe the significant spikes in 2020 which reflect the devastating impact of the COVID-19 pandemic as the UK has struggled to contain the spread of coronavirus.

Two commodity currencies, AUD and CAD, are closely connected to the price of gold and of oil. Compared to AUD, CAD is more sensitive to the Global financial crisis. Our AR-filtered statistics in Figure 4 indicate that FRUH for CAD is rejected in the wake of the global financial crisis and oil prices fell amid the global slowdown in 2013-2014. For AUD in Figure 5, the unbiasedness hypothesis is significantly rejected over a sustained period of time in 2008 through 2016, which is heavily influenced by commodity prices. The strength of the rejection began to taper off in the early 2017, and remained a steady decline until 2020.

Our empirical results from the Hong's correlation tests are consistent with previous observations. First, FURH may not be a general phenomenon, but in fact is regime-dependent. Given the reduced-form nature of the Fama regression, the spot-forward parity could break down in the face of changes in policy regimes or expectation formations due to the heightened counterparty or default risks (Baba and Packer, 2009; Coffey *et al.*, 2009; Griffolli and Ranaldo, 2011; Bussiere et al., 2019). Second, the failure of FRUH proposition is frequently observed during periods of financial crisis or a significant shift in the global economic environment, whereas the support for the parity regains its strength after the financial crisis faded. In particular, the favorable evidence for FRUH in recent years could be attributable to the decline and compression of interest rates near the zero lower bound during this period (Chinn and Quayyum, 2013; Burnside, 2019). Finally, the pattern of the correlation tests is consistent with the recent revival of exchange rate reconnect (Lilley *et al.*, 2019). Countries with large positive external imbalances (e.g., Japan and Swiss) tend to experience a deviation from the covered interest rate parity during the periods of increased market volatility since the financial distress increases hedging demand of global investors which in turn tightens balance sheet constraints for financial intermediaries (Liao and Zhang, 2020). Such an exchange rate hedging channel can explain the observed time varying pattern of the spot-forward parity and the rejection of the unbiasedness proposition occurs in times of financial crises including the COVID-induced financial turmoil.

6.3. Correlation tests under different lags

One of the major advantages of our proposed correlation tests is to allow us to examine the pattern of co-movements between the spot and the forward exchange markets across various lags. Such an evaluation of patterns is important to better understand whether the investors would respond to market shocks with a time-lag. In this subsection, we investigate how the strength of correlation changes as the lag order M increases. We calculate the test statistics for cross-correlations based on our proposed AR-filtered methods during the period 1996–2020 with varying lags from 1 to 48. Figure 6 depicts the correlation tests as functions of lag. Tests for AUD, JPY and CHF sharply increase at lags 2-6, reach peaks at lags 7-10 before gradually decreasing and finally achieve stabilization at protracted lags. By contrast, both CAD and GBP drop until hitting the bottom at lag 4, and then reach stabilization when lags $M \geq 12$. This trend confirms that the market takes time (almost a year as shown in Figure 6) to adapt to sudden shocks, and cross-correlations between the spot and forward rates have time-lags.

The time-lag effect observed above is consistent with the so-called 'delayed overshooting' put forth by Bacchetta and van Wincoop (2010), who provided a theoretical framework of information heterogeneity as a potential solution to the forward premium puzzle. In the same vein of Froot and Thaler (1990) and Eichenbaum and Evans (1995), and Bacchetta and van Wincoop (2010) formally modelled the notion of infrequent revisions of investor portfolio decisions where many investors do not actively manage their foreign exchange portfolios which can be rationalized by the costs of portfolio adjustments, or the costs of evaluations with new information. The argument can be presented as follows. Investors' international portfolios consist of both domestic and foreign bonds in order to diversify foreign exchange risks. In the event of rising interest rates, foreign currency appreciation is driven by a rising demand for the foreign bonds. However, only some investors changes their portfolio positions with the purchases of foreign bonds to exploit potential arbitrage gains, while the rest remain inattentive to the spike in the interest rate. As time goes on, most investors will reassess their portfolios and gradually buy the foreign bonds, and therefore the foreign currency appreciates further for a sustained period of time. Simply put, "after the initial increase in the foreign interest rate, the currency appreciates initially but continues to appreciate for some time as the whole market gradually adjusts its portfolio. The early movers might earn a high excess return, but the expected return differential will die out over time" (Engel, 2014, p.510). Thus, an increase in the interest rate can lead to a continued exchange rate appreciation if investors make infrequent portfolio decisions, and thus deviations from the parity will be left uncorrected.

7. Concluding Remarks

This research considers testing correlations between two long memory processes, where a stationary long memory process in a finite sample can be approximated very well by an AR(k) model if the lag length k is selected appropriately. Moreover, we have demonstrated the applicability of Hong's (1996a) statistics to test for two long memory processes being uncorrelated based on the sample cross-correlation function of the AR-filtered stationary long memory processes. The newly developed test procedure does not require the pre-specified functional form as in traditional regression-based approaches, and thus avoids the difficulties arising from an inaccurate estimation of the fractional parameter d, or the spurious regression induced by the long memory processes in the sense of Tsay and Chung (2000). The desirability of using the Akaike criterion seems to be a useful tool to select the order of an autoregression in approximating long memory processes. The Monte Carlo experiments conducted in this paper confirm our theoretical results. We find in finite samples that Hong's statistics based on the AR pre-whitening appear more useful for testing correlation between two stationary long memory processes than some currently popular methods.

For the empirical application of our approach, this study re-visits the forward rate unbiasedness hypothesis for currencies of five developed countries, with special focus on the pattern of cross-correlations when there is an imminent systematic risk in the financial system. While our empirical results are consistent with the previous findings that the validity of FRUH is regime-dependent, the shift in the patterns of the correlation between the spot and forward rates coincides with the timing of the global financial crisis or the economic environment changes, implying the proposed test statistics can be used as an early warning signals of imminent market crashes or potential adverse events.

correlation is generally a symmetric measure of dependence that does not provide information about the direction of the association. Consequently, an interesting extension of this work would be to consider a Granger causality test based on directional predictability between time series. More specifically, our AR-filtered metrics for measuring contributions to systemic risk could be extracted using Granger-causality tests in tail events or extreme risk. An example of such a test can be found in Hong (2001), Hong et al. (2009) and Wang et al. (2020) based on the occurrence of tail events and to detect the possible causality among such events. In this context, the main challenge to resolve is the extension of this test to a conditional setup. We leave this as an issue for future research.

REFERENCES

Akaike, H., 1969. Power spectrum estimation through autoregressive model fitting. Annals of Institute of Statistics Mathematics 21, 407-419.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001a. The Distribution of realized stock return volatility. Journal of Financial Economics 61, 43-76.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001b. The distribution of exchange rate volatility. Journal of the American Statistics Association 96, 42-55.

Andersen, T. G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579-625.

Baba, N., Packer, F., 2009. Interpreting deviations from covered interest parity during the financial market turmoil of 2007-08. Journal of Banking and Finance 33, 1953-1962.

Baillie, R. T., Bollerslev, T., 1990. Intra-day and inter-market volatility in foreign exchange rates. Review of Economic Studies 58, 565-585.

Baillie, R. T., 1996. Long Memory Processes and Fractional Integration in Econometrics. Journal of Econometrics 73, 5-59.

Barndorff-Nielsen, O. E., Shephard, N., 2001. Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics. Journal of Royal Statistical Society Series B 63, 167-241.

Beran, J., Bhansali, R.J., Ocker, D., 1995. On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. Biometrika 85, 921-934.

Berk, K.N. 1974. Consistent Autoregressive Spectral Estimates. The Annals of Statistics 2, 489-502.

Brockwell, P.J. and R.A. Davis (1991) *Time Series: Theory and Methods*, 2nd Edn. New York: Springer-Verlag.

Bollerslev, T., Wright, J.H., 2000. Semiparametrics estimation of long-memory volatility dependencies: the role of high-frequency Data. Journal of Econometrics 98, 81-106.

Burnside, C., 2019. Exchange Rates, Interest Parity, and the Carry Trade. Oxford Research Encyclopedia of Economics and Finance.

Bussiere, M., Chinn, M.D., Ferrara, L., Heipertz, J., 2019. The new Fama puzzle, NBER Working Paper No. 24342.

Chinn, M.D. and Quayyum, S., 2013. Long horizon uncovered interest parity re-assessed, NBER Working Paper No. 18482.

Christensen, B.J., Nielsen, M.∅., 2006. Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. Journal of Econometrics 133, 343-371.

Choi, K., Yu, W.C., Zivot, E., 2010. Long memory versus structural breaks in modeling and forecasting realized volatility. Journal of International Money and Finance 29, 857-875. 857-875.

Coffey, N., Hrung, W.B., Sarkar, A., 2009. Capital constraints, counterparty risk, and deviations from covered interest rate parity. Federal Reserve Bank of New York Staff Report 393.

Diebold, F.X., Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. The Economic Journal 119, 158-171.

Du, W., Tepper, A., Verdelhan, A., 2018. Deviations from covered interest rate parity, Journal of Finance 73, 915-957.

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.

Forbes, K., Rigobon, R., 2001. No contagion, only interpendence: measuring stock market co-movements. The Journal of Finance 57, 2223-2261.

Griffolli, M.T., Ranaldo, A., 2011. Limits to arbitrage during the crisis: Funding liquidity constraints and covered interest parity. Working paper, International Monetary Fund and University of St. Gallen.

Haugh, L.D., 1976. Checking the independence of two covariance stationary time series: a univariate residual cross-correlation approach. Journal of American Statistical Association 71, 378-385.

Hong, Y., 1996a. Testing for independence between two covariance stationary time series. Biometrika 83, 615-625.

Hong, Y., 1996b. Consistent Testing for Serial Correlation of Unknown Form. Econometrica 64, 837-864.

Hong, Y., 2001. A Test for Volatility Spillover with application to exchange rate. Journal of Economterics 103, 183-224.

Hong, Y., Liu, Y., Wang, S., 2009. Granger causality in risk and detection of extreme risk spillover between financial markets. Journal of Econometrics 150, 271-287.

Hosking, J. R. M., 1996. Asymptotic Distributions of the Sample Mean, Autocovariances, and Autocorrelations of Long-Memory Time Series. Journal of Econometrics 73, 261-284.

King, M., Sentana, E., Wadhwani, S., 1994. Volatility and links between national stock markets. Econometrica 62, 901-933.

Lin, W.L., Engle, R.F., Ito, T., 1994. Do bulls and bears move across borders? International transmission of stock returns and volatility. Review of financial Studies 7, 507-538.

Mallows, C.L., 1973. Come Comments on C_p. Technometrics 15, 661-675.

Martin, V. L., Wilkins, N.P., 1999. Indirect estimation of ARFIMA and VARFIMA models. Journal of Econometrics 93, 149-175.

McLeod, A.I., Hipel, K.W., 1978. Preservation of the Rescaled Adjusted Range 1: A Reassessment of the Hurst Phenomenon. Water Resources Research 14, 491-508.

Ng, S. Perron, P., 2005. Practitioners' Corner A Note on the Selection of Time Series Models. Oxford Bulletin of Economics and Statistics 67, 1, 115-134.

Priestley, M.B., 1981. *Spectral Analysis and Time Series*. London: Academic Press.

Poskitt, D.S., 2007. Autoregressive approximation in nonstandard situations; the Fractional Integrated and Non-Invertible Cases. Annals of Institute of Statistical Mathematics 59, 697-725.

Robinson, P.M., 1993. Highly Insignificant F-Ratios. Econometrica 61, 687-696.

Robinson, P.M., 2001. Higher-order Kernel Semiparametric M-estimation of Long memory. Journal of Econometrics 114, 1-27.

Rossi, B., Inoue, A., 2012. Out-of-sample forecast tests robust to the choice of window size. Journal of Business and Economic Statistics 30, 432-453.

Sun, Y., 2004. A convergent t-statistic in spurious regressions. Econometric Theory 20, 943-962.

Tsay, W.J., Chung, C.F., 2000. The spurious regression of fractionally integrated processes. Journal of Econometrics 96, 155-182.

Wang, S.H., Luc, B., Hsiao, C., 2013. Forecasting long memory processes subject to structural breaks. Journal of Econometrics 177, 171-184.

Wang, C.S.H., Hsiao, C., Wen, G.C.C. 2020 a. Autoregression filtered correlation tests for panel data exhibiting different order of integration in its components. *working paper.*

Wang, S.H., Hsiao, C., Yang, H., 2021 b. Market integration, systemic risk and diagnostic tests in large mixed panels. Econometric Reviews, *forthcoming.* https://doi.org/10.1080/07474938.2021.1889209.

APPENDIX

To prove Theorems 1-4, the following arguments are similar to those in Wang *et al.*(2021 a) and Wang *et al.* (2021 b). All the following analytical results are based on (B.1) and (B.2) of Wang *et al,* (2021 b).

Thus, by Wang *et al.* (2021 a) and Wang *et al.* (2021 b), we note that the AR(k)-filtered residual cross-correlation function, $\widehat{\rho}_{AR,ij}, i,j = 1,2,,\cdots,N, i \neq j$, asymptotically converges to the population cross-correlation function $\gamma_{ij}, i,j = 1,2,,\cdots,N, i \neq j$, i.e., $\widehat{\rho}_{AR,ij} \to \gamma_{ij}$, under combinations of any two of three of the following time series, an $I(0)$, a nearly unit root, and an $I(d), d \in (0,d)$ processes, in order to obtain the proofs of Theorems 1-4.

$\widehat{\rho}_{AR,ij}$ and ρ_{ij} are denoted as follows.

$$\widehat{\rho}_{AR,ij} = \frac{\widehat{R}_{ij}}{\{\widehat{R}_{ii}\widehat{R}_{jj}\}^{1/2}},$$

$$\widehat{R}_{ij} = T^{-1}\widehat{\varepsilon}'_{k,i}\widehat{\varepsilon}_{k,j},$$

$$\widehat{R}_{ii} = T^{-1}\widehat{\varepsilon}'_{k,i}\widehat{\varepsilon}_{k,i} \quad \text{and} \quad \widehat{R}_{jj} = T^{-1}\sum_{t=k+1}^{T}\widehat{e}^2_{t,k,j} = T^{-1}\widehat{\varepsilon}'_{k,j}\widehat{\varepsilon}_{k,j}.$$

where (i)$\widehat{R}_{ij}$ is the cross-covariance function constructed by estimates of AR-filtered residuals; (ii) R_{ii} and $\widehat{R}_{jj}$ are the variance functions based on estimates of AR-filtered residuals. (iii) $\widehat{\varepsilon}_{k,i}$ and $\widehat{\varepsilon}_{k,j}$ are AR-filtered residuals of $y_{t,i}$ and $y_{t,j}$.

$$\gamma_{ij} = \frac{R_{ij}}{\{R_{ii}R_{jj}\}^{1/2}},$$

$R_{ij} = T^{-1}\varepsilon'_{T-k,i}\varepsilon_{T-k,j}$, $R_{ii} = T^{-1}\varepsilon'_{T-k,i}\varepsilon_{T-k,i}$ and $R_{jj} = T^{-1}\varepsilon'_{T-k,j}\varepsilon_{T-k,j}$. where (i) R_{ii} and R_{jj} are the population variance functions; (ii) R_{ij} is the cross-covariance function; (iii) $\varepsilon_{T-k,i}$ and $\varepsilon_{T-k,j}$ are white noise error terms.

A. Proof of Theorem 1

On the analysis of Hong (1996) and Hong (2001), using Lemma A.4 of Wang *et al.* (2021 b), we note that and

$$\sum_{s=1-T}^{T-1} g^2(l/M)\widehat{\rho}_{AR,ij}(l)^2 = \sum_{s=1-T}^{T-1} g^2(l/M)(\gamma_{ij}(l)^2 + O_p(T^{-2d_j-1}(logT)^{-2d_j-1})) \qquad (A.1)$$

Following results of Hong (2001) and use Markove's inequality, we obtain

$$\sum_{s=1-T}^{T-1} g^2(l/M)\gamma_{ij}(l)^2 = O_p(M/T) \qquad (A.2)$$

Because (A.1), (A.2) and $M^{-1}D_T(k) \to D(k)$ (see, Hong (1996 b)) hold and when $y_{t,i}$ and $y_{t,j}$ are the stationary I(0) process and stationary I(d) processes respectively, the AR-filtered residuals $\widehat{\varepsilon}_{k,i}$ and $\widehat{\varepsilon}_{k,j}$ behave as those constructed by two white noise series, $\varepsilon_{T-k,i}$ and $\varepsilon_{T-k,j}$, asymptotically, thus once we consider a class of kernels with $r = 2$, by assumptions 4, 5 and using the same argument on the page 620 of Hong (1996 a), when $\frac{M}{T} \to 0$,

$$\begin{aligned}
Q_T &= \frac{T\sum_{l=1-T}^{T-1} g^2(l/M)\widehat{\rho}_{ij}(l)^2 - S_T(k)}{2D_T(k)^{1/2}} \\
&= \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)(\gamma_{ij}(l)^2 + O_p(T^{-2d_j-1}(logT)^{-2d_j-1}) - S_T(k)}{\{2D_T(k)\}^{1/2}} \\
&\leq \frac{T\sum_{l=1-T}^{T-1} g^2(l/M)\gamma_{ij}(l)^2 - S_T(k)}{\{2MD(k)\}^{1/2}} \\
&+ \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)O_p(T^{-2d_j}(logT)^{-2d_j-1})}{\{2MD(k)\}^{1/2}} \\
&= \frac{T\sum_{l=1-T}^{T-1} g^2(j/M)\gamma_{ij}(l)^2 - S_T(k)}{\{2D_T(k)\}^{1/2}} + o_p(1) = \bar{Q_T} + o_p(1).
\end{aligned} \qquad (A.3)$$

By Theorem 1 of Hong (1996 a), we know that Q_T converges to $N(0,1)$ in distribution even when two series follow the DGPs of Theorem 7. Likewise, by the same reasoning of (B.47), as $\frac{M}{T} \to 0$,

$$\begin{aligned}
Q_{T^*} &= \frac{T\sum_{l=1-T}^{T-1} g^2(l/M)\widehat{\rho}_{AR,ij}(l)^2 - MS_T(k)}{\{2MD_T(k)\}^{1/2}} \\
&= \frac{T\sum_{l=1-T}^{T-1} g^2(l/M)\gamma_{ij}(l)^2 - MS_T(k)}{\{2MD_T(k)\}^{1/2}} + o_p(1)
\end{aligned}.$$

and Q_T^* converges to $N(0,1)$ in distribution when two series follow the DGPs of Theorem 1. Hence, we can conclude that as the DGP processes satisfy Theorem 1, both Q_T and Q_T^* are asymptotically distributed as $N(0,1)$ when $\varepsilon_{T-k,t}$ is uncorrelated of $\varepsilon_{T-k,s}$ for all t and s. Thus the cross-correlation estimates constructed with the AR-filtered residuals behave as those constructed by two white noise series, $\varepsilon_{T-k,i}$ and $\varepsilon_{T-k,j}$, asymptotically.

A. Proof of Theorem 2

Similarly, using (A.2) and the Lemma A.4 of Wang *et al.* (2021 a), when $y_{t,i}$ follows an $I(0)$ process and $y_{t,j}$ follows an nearly I(1) process and $\frac{M}{T} \to 0$

$$\begin{aligned}
Q_T &= \frac{T\sum_{l=1-T}^{T-1} g^2(l/M)\widehat{\rho}_{ij}(l)^2 - S_T(k)}{2D_T(k)^{1/2}} \\
&= \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)(\gamma_{AR,ij}(l)^2 + O_p(T^{-1})) - S_T(k)}{\{2D_T(k)\}^{1/2}} \\
&= \frac{T\sum_{l=1-T}^{T-1} g^2(l/M)\gamma_{ij}(l)^2 - S_T(k)}{\{2MD(k)\}^{1/2}} \\
&+ \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)O_p(T^{-1})}{\{2MD(k)\}^{1/2}} = \bar{Q_T} + o_p(1).
\end{aligned} \qquad (A.3)$$

$$
\begin{aligned}
Q_{T^*} &= \frac{T\sum_{l=1-T}^{T-1} g^2(j/M)\widehat{\rho}_{ij}(l)^2 - MS_T(k)}{\{2MD_T(k)\}^{1/2}} \\
&= \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)\gamma_{ij}(l)^2 - MS_T(k)}{\{2MD_T(k)\}^{1/2}} + o_p(1)
\end{aligned}.
$$

when we consider a class of kernels with $r = 2$, by using assumptions 4, 5 and the same argument on the page 620 of Hong (1996 a).

Proofs of Theorem 3

By (A.1) and (A.2) and Lemma 2 (or Lemma A.4 of Wang *et al.*, 2021 a), we know that when $y_{t,i}$ follows an $I(d), d \in (0, 0.5)$ process and $y_{t,j}$ follows an nearly I(1) process, and as we consider a class of kernels with $r = 2$, by using assumptions 4, 5 and the same argument on the page 620 of Hong (1996 a), as $\frac{M}{T} \to 0$,

$$
\begin{aligned}
Q_T &= \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)\widehat{\rho}_{AR,ij}(l)^2 - S_T(k)}{2D_T(k)^{1/2}} \\
&= \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)(\gamma_{ij}(l)^2 + O_p(k^{-2d_i-1}) - S_T(k)}{\{2D_T(k)\}^{1/2}} \\
&= \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)\rho_{AR,ij}(l)^2 - S_T(k)}{\{2MD(k)\}^{1/2}} \\
&+ \frac{T\sum_{s=1-T}^{T-1} g^2(s/M)O_p(k^{-2d_i-1}T^{-1})}{\{2MD(k)\}^{1/2}} = \bar{Q_T} + o_p(1).
\end{aligned} \tag{A.4}
$$

By Theorem 1 of Hong (1996 a), we know that Q_T converges to $N(0, 1)$ in distribution even when two series follow the DGPs of Theorem 3. Likewise,

$$
\begin{aligned}
Q_{T^*} &= \frac{T\sum_{s=1-T}^{T-1} g^2(j/M)\widehat{\rho}_{ij}(s)^2 - MS_T(k)}{\{2MD_T(k)\}^{1/2}} \\
&= \frac{T\sum_{s=1-T}^{T-1} g^2(s/M)\rho_{ij}(s)^2 - MS_T(k)}{\{2MD_T(k)\}^{1/2}} + o_p(1)
\end{aligned}.
$$

and Q_T^* converges to $N(0, 1)$ in distribution when two series follow the DGPs of Theorem 3. Hence, we can conclude that as the DGP processes satisfy Theorem 3 both Q_T and Q_T^* are asymptotically distributed as $N(0, 1)$.

Proofs of Theorem 4

Likewise, By (A.1), (A.2) and Lemma 4 (or Lemma A.4 of Wang *et al.*, 2021 b), we know that when $y_{t,i}$ follows an $I(d_i), d_i \in (0, 0.5)$ process and $y_{t,j}$ follows an $I(d_j), d_j \in (0, 0.5)$ process, as $\frac{M}{T} \to 0$,

$$
\begin{aligned}
Q_T &= \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)\widehat{\rho}_{AR,ij}(l)^2 - S_T(k)}{2D_T(k)^{1/2}} \\
&= \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)(\gamma_{ij}(l)^2 + O_p(k_1^{-2d_i-1}k_2^{-2d_j-1}T^{2d_i+2d_j-2})) - S_T(k)}{\{2D_T(k)\}^{1/2}} \\
&= \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)\rho_{AR,ij}(l)^2 - S_T(k)}{\{2MD(k)\}^{1/2}} \\
&+ \frac{T\sum_{s=1-T}^{T-1} g^2(s/M)O_p(k_1^{-2d_i-1}k_2^{-2d_j-1}T^{2d_i+2d_j-2}))}{\{2MD(k)\}^{1/2}} = \bar{Q_T} + o_p(1).
\end{aligned} \tag{A.5}
$$

By Theorem 1 of Hong (1996 a), we know that Q_T converges to $N(0,1)$ in distribution even when two series follow the DGPs of Theorem 4. Likewise,

$$Q_{T^*} = \frac{T\sum_{l=1-T}^{T-1} g^2(j/M)\widehat{\rho}_{ij}(l)^2 - MS_T(k)}{\{2MD_T(k)\}^{1/2}}$$
$$= \frac{T\sum_{l=1-T}^{T-1} g^2(s/M)\rho_{ij}(l)^2 - MS_T(k)}{\{2MD_T(k)\}^{1/2}} + o_p(1)$$

and Q_T^* converges to $N(0,1)$ in distribution when two series follow the DGPs of Theorem 3. Hence, we can conclude that as the DGP processes satisfy Theorem 3 both Q_T and Q_T^* are asymptotically distributed as $N(0,1)$, when we consider a class of kernels with $r=2$, by using assumptions 4, 5 and the same argument on the page 620 of Hong (1996 a).

Table 1. The Size Performance of the Q_T and Q_{T^*} Test at 5% Level of Significance When $y_{t,1}$ and $y_{t,2}$ follow DGPs (a), (b) and (c).

	DGP	$y_{t,2}$	k	M	0.1	0.2	0.3	0.40	0.45
	$DGP(c)$	$y_{t,1}$							
Q_T			k_{AIC}	5	5.1	5.2	5.3	5.5	5.8
				9	5.0	5.1	5.1	5.3	5.4
				15	4.8	4.9	5.0	5.2	5.3
Q_{T^*}			k_{AIC}	5	5.0	5.0	5.2	5.3	5.5
				9	4.7	4.9	5.0	5.1	5.3
				15	4.5	4.7	4.8	4.9	5.1
	$DGP(b)$	$y_{t,1}$							
Q_T			k_{AIC}	5	5.2	5.4	5.5	5.7	5.7
				9	5.1	5.2	5.2	5.5	5.6
				15	4.9	4.9	5.1	5.1	5.4
Q_{T^*}			k_{AIC}	5	5.2	5.3	5.3	5.4	5.6
				9	4.8	4.8	5.0	5.0	5.4
				15	4.6	4.9	4.9	5.0	5.0
	DGP(a)	$y_{t,2}$	k	M	nearly I(1)				
Q_T			k_{AIC}	5	5.8				
				9	5.6				
				15	5.2				
Q_{T^*}			k_{AIC}	5	5.2				
				9	5.0				
				15	5.5				

Notes: $T = 200$. The chosen kernel is the Parzen kernel.

Table 2. The Power Performance of the Q_T and Q_{T^*} Test at 5% Level of Significance When $y_{t,1}$ and $y_{t,2}$ follow DGPs (a), (b)and (c).

	DGP	$y_{t,2}$	k	M	0.1	0.2	0.3	0.40	0.45
	$DGP(c)$	$y_{t,1}$							
Q_T			k_{AIC}	5	68.1	68.6	68.3	68.9	69.8
				9	57.2	57.5	57.9	58.3	59.2
				15	46.2	46.2	46.9	47.3	47.6
Q_{T^*}			k_{AIC}	5	69.0	69.2	69.3	69.8	70.3
				9	57.6	58.1	58.3	58.9	60.1
				15	47.0	47.4	47.6	48.2	48.2
	DGP(b)	$y_{t,2}$	k	M					
Q_T			k_{AIC}	5	68.9	69.0	69.1	69.2	69.2
				9	59.1	60.2	60.0	60.0	60.3
				15	48.2	49.2	50.0	50.2	50.4
Q_{T^*}			k_{AIC}	5	69.1	69.1	69.3	69.3	69.4
				9	57.1	57.3	57.9	58.2	58.3
				15	47.3	47.3	47.9	48.3	48.5
	DGP(a)	$y_{t,2}$	k	M	nearly I(1)				
Q_T			k_{AIC}	5	69.0				
				9	67.1				
				15	51.2				
Q_{T^*}			k_{AIC}	5	68.8				
				9	56.8				
				15	48.1				

Notes: $T = 200$. The chosen kernel is the Parzen kernel.

Table 3. The Size Performance of the Q_T and Q_T^* Tests When $y_{t,1}$ = DGP(b), $y_{t,2}$ = DGP(b)

	Y_{1t}	d_1	k	M	0.33	0.36	0.38	0.40	0.44
Size	Y_{2t}	d_2							
Q_T		0.36	k_{AIC}	5	5.6	5.6	5.7	5.8	5.8
				9	5.4	5.3	5.4	5.4	5.6
				15	5.2	5.2	5.2	5.2	5.7
		0.40	k_{AIC}	5	5.7	5.6	5.6	5.5	5.9
				9	5.3	5.7	5.4	5.5	5.8
				15	5.2	5.2	5.5	5.6	5.6
		0.44	k_{AIC}	5	5.2	5.3	5.4	5.6	6.1
				9	5.0	5.0	5.4	5.3	5.7
				15	5.0	5.0	5.3	5.2	5.5
Q_T^*		0.36	k_{AIC}	5	5.5	5.6	5.6	5.6	5.9
				9	4.5	4.4	4.6	4.7	5.2
				15	4.4	4.4	4.4	4.6	4.6
		0.40	k_{AIC}	5	5.3	5.2	5.2	5.3	5.6
				9	5.0	5.1	5.1	5.2	5.2
				15	4.6	4.7	4.6	4.8	5.0
		0.44	k_{AIC}	5	5.4	5.3	5.6	6.0	6.2
				9	5.9	5.6	5.7	5.5	5.8
				15	5.0	5.1	5.2	5.5	5.7

Notes: $T = 200$. The chosen kernel is the Parzen kernel.

Table 4. The Power Performance of the Q_T and Q_{T^*} Tests at 5% Level of Significance When $y_{t,1}$ = DGP (b), $y_{t,2}$ = DGP(b)

	Y_{1t}	d_1	k	M	0.33	0.36	0.38	0.40	0.45
Size	Y_{2t}	d_2							
Q_T		0.36	k_{AIC}	5	67.2	67.1	67.6	67.8	68.2
				9	57.2	58.1	58.7	58.5	58.7
				15	46.1	46.2	46.9	46.7	47.1
		0.40	k_{AIC}	5	67.2	67.0	66.9	67.4	67.4
				9	57.4	58.6	58.6	58.6	59.3
				15	45.8	46.0	45.9	45.9	46.7
		0.44	k_{AIC}	5	67.6	67.2	67.5	67.8	67.9
				9	58.4	59.2	59.7	59.3	59.8
				15	45.6	45.5	45.6	45.6	45.9
Q_{T^*}		0.36	k_{AIC}	5	64.3	64.6	64.7	64.8	65.0
				9	54.9	54.9	54.9	55.2	55.1
				15	44.2	44.0	43.9	44.7	45.1
		0.40	k_{AIC}	5	65.6	65.6	65.5	65.6	66.2
				9	55.4	55.7	55.4	55.2	55.8
				15	44.0	44.3	44.6	44.5	44.5
		0.44	k_{AIC}	5	65.7	65.7	65.8	65.6	66.0
				9	57.1	57.0	57.3	57.3	57.6
				15	43.3	43.3	43.6	43.6	43.9

Notes: $T = 200$. The chosen kernel is the Parzen kernel.

Table 5. The Rejection Percentages of the t test at 5% Level of Significance When $y_{t,1}$ = DGP (b), $y_{t,2}$ = DGP (b) and $\rho_{12}(j) = 0$ for all j

Y_{1t}	d_1	0.33	0.36	0.38	0.40	0.44
Y_{2t}	d_2					
	0.33	17.9	19.9	22.3	24.8	29.1
	0.36	19.1	23.1	24.8	26.7	30.1
	0.38	22.3	23.9	25.4	27.7	31.8
	0.40	24.1	26.1	27.3	28.9	32.1
	0.44	29.3	30.3	31.9	32.1	33.2

Table 6. 95 % Quantiles for $\widetilde{t}$ Test

Y_{1t}	d_1	0.33	0.36	0.38	0.40	0.44
Y_{2t}	d_2					
	0.36	2.0522	2.0531	2.0539	2.0584	2.0616
	0.40	2.0580	2.0595	2.0601	2.0625	2.0632
	0.44	2.0611	2.0618	2.0625	2.0631	2.0644

Table 7. The Size and Power Performance of the $\widetilde{t}$ Test When $y_{t,1}$ = DGP (b)

,

$y_{t,2}$ = DGP (b)

	Y_{1t}	d_1	0.33	0.36	0.38	0.40	0.44
Size	Y_{2t}	d_2					
		0.36	5.6	5.5	5.5	5.6	5.8
		0.40	5.5	5.5	5.7	5.9	6.0
		0.44	5.7	5.9	5.9	6.1	6.3
Power		0.36	28.2	29.1	29.3	29.9	30.6
		0.40	29.6	30.2	30.6	31.3	31.9
		0.44	30.3	30.7	31.0	31.6	32.5

Notes: $T = 200$.