Using Difference-Based Methods for Inference in Regression with Fractionally-Integrated Processes

Wen-Jen Tsay*

The Institute of Economics
Academia Sinica, Taiwan

* We would like to thank Professor C.W.J. Granger for useful discussions over an early version of this paper presented at the Western Economic Association International 2003 Pacific Rim Conference. We also thank an anonymous referee and an Associate Editor for their valuable comments and suggestions. Financial support from the NSC-94-2415-H-001-018 of Taiwan is gratefully acknowledged. Wen-Jen Tsay. The Institute of Economics, Academia Sinica, Taipei, Taiwan, R.O.C. Tel. (886-2) 27822791 ext. 296. Fax: (886-2) 27853946. E-mail: wtsay@ieas.econ.sinica.edu.tw

Abstract. This paper suggests a difference-based method for inference in the regression model involving fractionally-integrated processes. Under suitable regularity conditions, our method can effectively deal with the inference problems associated with the regression model consisting of nonstationary, stationary, and intermediate memory regressors, simultaneously. Although the difference-based method provides a very flexible modeling framework for empirical studies, the implementation of this method is extremely easy, because it completely avoids the difficult problems of choosing a kernel function, a bandwidth parameter, or an autoregressive lag length for the long-run variance estimation. The asymptotic local power of our method is investigated with a sequence of local data-generating processes (DGP) in what Davidson and MacKinnon (1985) call "regression direction". The simulation results indicate that the size control of our method is excellent even when the sample size is only 100, and the pattern of power performance is highly consistent with the theoretical finding from the asymptotic local power analysis conducted in this paper.

Some key words: Fractionally-integrated process; Long memory.

1. Introduction

It is well known that a great deal of estimation and inference techniques have been developed for weakly dependent, $I(0)$, and unit root, $I(1)$, processes, including Phillips (1987), Newey and West (1987), Phillips and Hansen (1990), Andrews (1991), and Hansen (1992). Nevertheless, inference methods for the fractionally-integrated processes are relatively scarce. Among them, Robinson and Hidalgo (1997) consider the generalized least squares (GLS) estimator for a time series regression in the presence of long-range dependence in both errors and stochastic regressors. Robinson (1998) provides a framework to effectively test a time series relationship involving stationary long memory processes under suitable regularity conditions. In addition, Hidalgo (2000) tackles the problem of causality among stationary time series allowing long memory. These studies share a common feature, i.e., all the time series considered are stationary.

This paper considers the multiple linear regression frameworks of Robinson and Hidalgo (1997), but allows for the joint presence of stationary and nonstationary fractionally-integrated processes. Since there are at least two different definitions of nonstationary fractionally-integrated processes, before presenting the regression model considered in this paper, let us first define the underlining process clearly.

Adapting the notation of Robinson (2005), we define a short memory process $\{\eta_t, t = 0, \pm 1, \ldots\}$ as a zero-mean covariance stationary process with spectral density $f(\lambda)$, which is bounded and bounded away from zero. With η_t and $d \in [-1/2, 1/2)$, we define the corresponding stationary fractionally-integrated, or $I(d)$, process ξ_t as

$$\xi_t = \triangle^{-d} \eta_t, \quad t = 0, \pm 1, \ldots, \tag{1}$$

where $\triangle = 1 - L$, and L is the usual lag operator. We denote the truncated ξ_t process as:

$$\xi_t^{\#} = \xi_t \mathbf{1}(t \geq 1), \quad t = 0, \pm 1, \ldots, \tag{2}$$

where $\mathbf{1}$ is the indicator function. Based on the process $\xi_t^{\#}$ in (2), for $q \geq 0$, we define its corresponding Type I $I(q + d)$ process[1] as:

$$W_t = \triangle^{-q} \xi_t^{\#}, \quad t = 0, \pm 1, \ldots. \tag{3}$$

[1] Please see Robinson (2005) about the comparison between Type I and Type II $I(q + d)$ processes.

This Type I $I(q+d)$ process has been used by Sowell (1990), Marinucci (2000), Tsay (2000), and Velasco and Robinson (2000), to name a few.

Given the preceding $I(d)$ and its associated Type I nonstationary $I(q+d)$ $(q \geq 1)$ process, then denoting γ as an unknown intercept, U_t as a scalar error term, and the $(k \times 1)$ column vector Z_t as the stochastic regressors whose j-th element is $Z_{t,j}$, we present our regression model as:

$$\begin{cases} C_t = \gamma + Z_t'\beta + U_t, \quad Z_{t,j} = I(d_{Z_j}), \ U_t = I(d_U), \ \max\limits_{1\leq j\leq k} d_{Z_j} = d_Z, \quad j = 1, 2, \ldots, k, \ t = 1, 2 \ldots, T, \\ d_Z > d_U, \quad \text{if } d_{Z_j} \geq 1/2, \text{ for some } j = 1, 2, \ldots, k. \end{cases} \tag{4}$$

Specifically, if each component of Z_t and U_t in (4) is stationary, then we impose no restriction on the relative magnitude of d_Z and d_U as Robinson and Hidalgo (1997) do. When the maximum differencing parameter of the regressors Z_t, i.e., d_Z defined in (4), is greater than or equal to $\frac{1}{2}$, then we require that $d_Z > d_U$ so as to ensure that the probability order of magnitude of the stochastic regressors Z_t dominates that of the error term. This condition is not stringent at all and is always imposed in the regression models involving nonstationary processes. The model in (4) also allows the simultaneous appearance of a nonstationary error term U_t and stationary regressors Z_{1t}, where Z_{1t} is a proper subset of Z_t, as long as the conditions in (4) are satisfied.[2] The possibility that some regressors in Z_t are intermediate memory or even noninvertible processes is not ruled out with the model in (4), either.

To deal with the inference problems for the regression model in (4), we suggest a difference-based method. Before discussing the details of this proposed method, let us address here that better estimation procedures sometimes exist if we know the exact data-generating process (DGP) in (4). For example, as all of the elements in Z_t are $I(1)$ processes and U_t is an $I(0)$ process, the fully-modified estimators of Phillips and Hansen (1990) are efficient for such a cointegration model. Moreover, when both Z_t and U_t are stationary long memory processes, the frequency-domain weighted least squares (WLS) estimator of Robinson and Hidalgo (1997) achieves the Gauss-Markov bound under suitable regularity conditions, including that Z_t and U_t are mutually independent. Under the same independence assumption between Z_t and U_t, Tsay (2000)

[2] However, we cannot find an empirical study where some regressors have less persistence than the error term in the regression model with other nonstationary regressors.

shows that the convergence rate of the first-differenced (FD) estimator can be faster than that of the ordinary least squares (OLS) counterpart when Z_t is a nonstationary process, indicating that the implementation of the FD estimator for the model in (4) may not incur any efficiency loss, as instead it may result in an infinite efficiency gain in the estimation of β relative to its OLS counterpart.

As clearly suggested by an anonymous referee and the Associate Editor, the benefits of differencing found in Tsay (2000) depend on the independence between the regressors and the error term. When this independence assumption is relaxed, the findings in Tsay (2000) may not hold. In fact, if Z_t and U_t are correlated, then the FD estimator is even inconsistent in this set-up. Thus, the objective of this paper is not to present a method which can 'efficiently' estimate the model in (4), but instead we want to propose a unified testing methodology which can conveniently deal with the inference problems of the regression model in (4). The advantages of using such a unified inference method is enhanced greatly once we realize that, in the regression model where the integration orders of Z_t and that of U_t can be fractional but unknown, the differencing parameters of the regressors Z_t and that of the error term U_t cannot be estimated without bias when the sample size T is relatively moderate. In addition, as clearly pointed out by an Associate Editor, when some regressors in Z_t are nonstationary, the OLS estimator generally has a non-standard distribution which is difficult to use in inference. However, our method does not require us to simulate the critical values for various combinations of d_Z, d_U, and sample size T under this circumstance.

As will be shown in our Theorem 1, the $\widehat{\beta}_{MD}$ estimator generated from our difference-based method is as simple as the usual OLS estimator and is asymptotically normally distributed under suitable regularity conditions. Accordingly, the usual t-ratio statistic can be used to test each element of β, once the variance matrix of $\widehat{\beta}_{MD}$ is consistently estimated. In fact, the computation of this variance matrix is almost as simple as the $\widehat{\beta}_{MD}$ estimator in that we adopt Robinson's (1998) long-run variance estimator which does not involve any choice of an autoregressive (AR) lag length, a kernel function, or a bandwidth parameter. The remaining parts of this paper are arranged as follows: Section 2 presents the test statistic and the main results. In Section 3 we consider the choice of the number of differencing and asymptotic local power of our method. The theoretical finding generated from the power analysis is verified through a Monte Carlo experiment. Section 4 provides a conclusion.

2. The statistics and the main results

Our difference-based method contains two stages. The first one is simply to calculate the OLS estimator for β, denoted as $\widehat{\beta}_{MD}$, in the following multiple-differenced (MD) transformation of the regression model in (4):

$$\triangle^M C_t = \triangle^M Z_t'\beta + \triangle^M U_t, \qquad M = \begin{cases} g, & \text{if } d_U = d_{Z_j} = 0, \quad \text{for all } j = 1, 2, \ldots, k; \\ g, & \text{if } d_U < 0,\, d_{Z_j} < 0, \quad \text{for all } j = 1, 2, \ldots, k; \\ g + \max\left([d_U], [d_Z]\right), & \text{otherwise,} \end{cases} \tag{5}$$

where g is a nonnegative integer, $[G]$ represents the integer part of G, and $t = M+1, M+2, \ldots, T$.[3] Please note that the choice of M in (5) actually is quite flexible, but it must ensure the resulting $\widehat{\beta}_{MD}$ is asymptotically normally distributed under suitable regularity conditions. For example, if we know the regressors and the error term are weakly stationary, then $M = 0$ can be used, because this case has already been discussed by Robinson (1998). In addition, if there is only one regressor and $0 < d_U, d_Z < \frac{1}{2}$ and $d_U + d_Z < \frac{1}{2}$, then we still can apply the results in Robinson (1998) - that is, we do not need to difference the data in this set-up and $M = g = 0$ can be chosen. By contrast, if we know that $0 < d_U, d_Z < \frac{1}{2}$ holds, but $d_U + d_Z \geq \frac{1}{2}$, then $M = g = 1$ should be used instead to ensure that the resulting $\widehat{\beta}_{MD}$ is asymptotically normally distributed. Another example where $M = 0$ can be adopted is when the differencing parameters of the regressors and that of the error term are all negative. However, as will be shown later, we still can employ $M = 1$ (first-difference) in this set-up and not necessarily increase a higher variance of our proposed difference-based method.

The preceding examples point out that if we know the exact DGP in (4), then sometimes we can find some methods existing in the literature to solve the corresponding inference problems. Nevertheless, as we have discussed in Section 1, the integration orders of Z_t and that of U_t are simply unknown in reality, and the differencing parameters of the regressors Z_t and that of the error term U_t just cannot be estimated without bias when the sample size is relatively moderate. That is why we propose a unifying testing procedure when confronting the model in (4). In other words, the message carried in (5) is that, if we are not sure about the

[3] Methods based on differencing have a long history in time series literature. Please also see Anderson (1971, p.66) for the overviews.

range and combinations of the differencing parameters in the regressors and the error term, then it is better to pick a larger value of M to ensure the differencing parameters of the transformed regressors and that of the transformed error term are all negative. In particular, if we suspect some regressors are $I(1)$ processes, but the unit root or stationarity tests cannot tell us whether $d_Z = 0$ or 1, then we simply adopt $M = 2$ to implement our difference-based method if we are sure that $\max([d_U], [d_Z])$ in (5) is less than 2. The great flexibility enjoyed with our method in choosing M is valuable in empirical applications. We will discuss this point after we present Theorem 1.

Provided that we can build a stationary $I(d)$ process ξ_t in (1) and its associated Type I nonstationary $I(q+d)$ process in (3) with the short memory process η_t in (1), we construct the processes Z_t and U_t in (4) with the corresponding short memory processes Z_t^* and U_t^*, no matter if Z_t and U_t are stationary or nonstationary. By the Wold decomposition, both short memory processes, Z_t^* and U_t^*, can be expressed as the form of infinite-order moving average (MA) processes which include stationary autoregressive (AR) and stationary and invertible ARMA processes as the special cases. Thus, after applying the MD transformation defined in (5), the resulting transformed regressors and transformed error term are stationary and can be represented with the following infinite-order MA processes:

$$\triangle^M U_t = \sum_{i=0}^{\infty} \psi_i a_{t-i}, \qquad \triangle^M Z_t = \sum_{i=0}^{\infty} \varphi_i b_{t-i}, \qquad t = M+1, \ldots, T, \tag{6}$$

where a_t and b_t are assumed to satisfy the conditions in Assumption A of Robinson (1998). As a result, ψ_i and each element of φ_i in (6) are bounded with the following rate:

$$\psi_i = O\left(i^{(d_U - M)-1}\right) = O\left(i^{d_U - M - 1}\right), \qquad \varphi_{i,j} = O\left(i^{(d_{Z_j} - M)-1}\right) = O\left(i^{d_{Z_j} - M - 1}\right), \quad j = 1, \ldots, k. \tag{7}$$

The conditions in (2.2) and (2.3) of Robinson (1998) allow each element of b_t to be correlated with each other, which is weaker than the independence assumption used in Tsay (2000). For ease of exposition, from now on, the following discussions are all based on the results in (6) and that a_t and b_t satisfy the conditions in Assumption A of Robinson (1998).

We now consider the computational procedure of the $\widehat{\beta}_{MD}$ estimator. Although we do not have an intercept in (5), before we run the OLS regression for the MD transformed model in (5), which is the first stage of our difference-based method, we still demean $\triangle^M C_t$ and $\triangle^M Z_t$ in (5) to obtain:

$$Y_t = \triangle^M C_t - \overline{\triangle^M C}, \qquad X_t = \triangle^M Z_t - \overline{\triangle^M Z}, \tag{8}$$

where $\overline{S}$ denotes the sample mean of S_t. The purpose is to make our testing procedure be more in line with the usual OLS estimator, because the estimate results from these $T_M \equiv T - M$ observations of Y_t and X_t is equivalent to that obtained from adding a constant term into the regressors $\triangle^M Z_t$ in (5).

The OLS estimation results are presented as follows:

$$Y_t = X_t' \widehat{\beta}_{MD} + e_t, \qquad \widehat{\beta}_{MD} = \left(X'X\right)^{-1} X'Y, \qquad t = M+1, M+2, \ldots, T, \tag{9}$$

where e_t represents the OLS residuals, and the t-th row of X and that of Y are X_t' and Y_t, respectively. As shown previously, the notation $\widehat{\beta}_{MD}$ points out that we multiple-difference the data series before we run the OLS regression. On the other hand, it also signifies that we first difference C_t and Z_t by M times before we conduct the subsequent OLS estimation. In the following analysis about the asymptotic properties of $\widehat{\beta}_{MD}$, we name $\widehat{\beta}_{MD}$ as the MD estimator as compared to the FD estimator employed in Tsay (2000). The asymptotic properties of $\widehat{\beta}_{MD}$ are displayed in the following Theorem 1.

THEOREM 1. *Given that $\triangle^M U_t$ and $\triangle^M Z_t$ are defined in (6), M is defined in (5), a_t and b_t satisfy the conditions in Assumption A of Robinson (1998), and the spectral density function of $\triangle^M Z_t \, \triangle^M U_t$ is finite and positive definite, then as $T \to \infty$, the MD estimator for the model in (4) is asymptotically normally distributed and $\widehat{D}^{-1/2}\sqrt{T}\left(\widehat{\beta}_{MD} - \beta\right) \xrightarrow{d} N(0, I_k)$, where $\widehat{D} = \left(T_M^{-1} X'X\right)^{-1} \widehat{V}_{T_M} \left(T_M^{-1} X'X\right)^{-1}$,*

$$\widehat{V}_{T_M} = \sum_{i=-T_M+1}^{T_M-1} \widehat{c}_i \otimes \widehat{d}_i, \qquad \widehat{c}_i = T_M^{-1} \sum_{M+1 \le t, t+i \le T} e_t \, e_{t+i}, \qquad \widehat{d}_i = T_M^{-1} \sum_{M+1 \le t, t+i \le T} X_t X_{t+i}'. \tag{10}$$

$\xrightarrow{d}$ stands for convergence in distribution, I_k is a $(k \times k)$ matrix, and $\otimes$ denotes the Kronecker product.

The proof of Theorem 1 is omitted, because the results in Theorem 1 are the direct applications of Theorem 5 of Robinson (1998). The point made here is that the important findings in Robinson (1998) also apply to the cases where the differencing parameters of the regressors and that of the error term are all negative. The intuition behind the results in Theorem 1 is as follows: When $k = 1$, both U_t and $Z_{t,1}$ are independent ARFIMA$(0, d_U, 0)$ and ARFIMA$(0, d_{Z_1}, 0)$ processes, respectively, $d_U \in [0, 1)$, $d_{Z_1} \in [0, 1)$, and then $M = \max([d_U], [d_Z]) + g = 0 + g = 1$ is chosen to ensure that the autocorrelations of $\triangle U_t$ and those of $\triangle Z_{t,1}$ at lag i are all negative as $i \neq 0$. It follows here that the autocorrelations of the product $\triangle Z_{t,1} \triangle U_t$ are positive for all lag i, as U_t is assumed to be independent of $Z_{t,1}$. Therefore, the spectral density function of $\triangle Z_{t,1} \triangle U_t$ is finite and positive definite under this set-up. Combining the fast decay rate of the autocorrelations of $\triangle Z_{t,1} \triangle U_t$, we can easily prove the asymptotic normality of the MD estimator. On the other hand, when the differencing parameters of $\triangle^M Z_{t,1}$ and that of $\triangle^M U_t$ are not all negative under some other specifications for $Z_{t,1}$ and U_t, the spectral density function of $\triangle^M Z_{t,1} \triangle^M U_t$ may degenerate to be zero based on Lemma A.1 of Tsay (2000), implying that the $\sqrt{T}$ convergence rate of the MD estimator may not hold. That explains why we have to choose a large enough M to ensure the differencing parameters of the differenced regressors and differenced error term are all negative.

In sharp contrast to the semiparametric long-run variance estimators considered in Newey and West (1987) and Andrews (1991), and the autoregressive spectral estimate in Berk (1974), the computation of $\widehat{V}_{T_M}$ proposed by Robinson (1998) has nothing to do with the selection of an AR lag length, a kernel function, or a bandwidth parameter. Most importantly, no matter whether some of the regressors are nonstationary or non-invertible fractionally-integrated processes, each element of β in (4) can be tested with the usual t-ratio statistic with the assistance of the easy-to-implement but elegant long-run variance estimator of Robinson (1998). Of course, a Wald-type statistic can be used to test the value of β jointly. Building on the seminal work of Robinson (1998), this paper proposes a unifying framework to deal with the inference problems of the regression model in (4).

3. Numerical illustrations

3.1. The choice of M

This subsection considers the efficiency issues surrounding the choice of M in (5). Theorem 1 reveals that the efficiency of $\widehat{\beta}_{MD}$ is inversely related to the asymptotic value of $\widehat{D}$ in Theorem 1, i.e., a $(k \times k)$ finite and positive definite matrix D. Under the case where there is only one stochastic regressor $Z_{t,j}$, we have:

$$D = \gamma_j(0)^{-2} \sum_{i=-\infty}^{\infty} \gamma_j(i)\gamma_U(i), \tag{11}$$

where $\gamma_j(i)$ denotes the autocovariance function of $\triangle^M Z_{t,j}$ at lag i, and $\gamma_U(i)$ denotes the autocovariance function of $\triangle^M U_t$ at lag i, respectively. Thus, the magnitude of D in (11) sheds light on the relationship between the efficiency of $\widehat{\beta}_{MD}$ and the value of M in (5).

Although a closed-form solution for the value of D in (11) is not available in this paper, even when $\triangle^M Z_{t,j}$ and $\triangle^M U_t$ are the simplest ARFIMA$(0, d, 0)$ processes, theoretically we can well approximate the value of D in (11) with the following formula:

$$\widetilde{D} = \gamma_j(0)^{-2} \sum_{i=-h}^{h} \gamma_j(i)\gamma_U(i),$$

if the value of h is large enough and we can choose a proper value of M to make the value of $\gamma_j(i)\gamma_U(i)$ in (11) decay at a fast enough rate of $2(d_U + d_{Z_j}) - 4M - 2$. To demonstrate the preceding argument, we assume the data is generated as follows:

$$U_t = \triangle^{-d_U} v_t, \qquad Z_{t,j} = \triangle^{-d_{Z_j}} w_{t,j}, \qquad d_U \in (-1, 0), \qquad d_{Z_j} \in (-1, 0), \tag{12}$$

where v_t and $w_{t,j}$ both are zero-mean n.i.d. processes. For ease of comparison, we also assume that:

$$E(v_t^2) = 1, \qquad E(w_{t,j}^2) = 1, \tag{13}$$

i.e., both U_t and $Z_{t,j}$ in (12) are intermediate memory processes, implying that we do not need to difference the data to fulfill the regularity conditions in Theorem 1. In other words, $M = 0$ can be used under the specifications in (12). Our task is to find out what will happen if we choose $M = 1$ instead.

When $M = 0$ is adopted (no difference at all), we calculate a truncated version of D in (11) as:

$$\widetilde{D} = \gamma_j(0)^{-2} \sum_{i=-h}^{h} \gamma_j(i)\gamma_U(i), \qquad h = 800, \tag{14}$$

and present the results in Table 1. Under the configurations specified in Table 1, we note that the value of $\gamma_j(i)\gamma_U(i)$ in (14) decays at a rate of $2(d_U + d_{Z_j}) - 2 < -2$ when $M = 0$ is chosen. We also find that the value of $\widetilde{D}$ in (14) remains unchanged when the value of h increases from 500 to 800 for any configuration considered in Table 1. These two observations support that $\widetilde{D}$ in (14) with $h = 800$ is a good approximation to D under the specifications in Table 1.

To illustrate the impact of 'over-differencing' on the efficiency of the MD estimator, we compute the corresponding value of $\widetilde{D}$ when $M = 1$ is employed and present the results in Table 2. Please note that the reason we label the word 'over-differencing' is to signify that we do not need to difference the data under the specifications in Table 1, but we on purpose do it in Table 2. We find that the value of $\widetilde{D}$ in Table 2 is larger than that of the corresponding $\widetilde{D}$ in Table 1 when $M = 0$ is replaced with $M = 1$. In other words, provided that the conditions in Theorem 1 are satisfied, 'over-differencing' incurs efficiency loss under the specifications in Table 1 where only the stationary ARFIMA$(0, d, 0)$ processes are considered.

To further explore the preceding issue, we extend our experiment to the following more general processes:

$$U_t = \triangle^{-d_U} v_t^*, \qquad Z_{t,j} = \triangle^{-d_{Z_j}} w_{t,j}^*, \qquad d_U \in (-1, 0), \qquad d_{Z_j} \in (-1, 0), \tag{15}$$

where

$$\begin{aligned} (1 - \phi_1 L)v_t^* &= (1 + \theta_1 L)v_t, \\ (1 - \phi_{2,1} L)w_{t,j}^* &= (1 + \theta_{2,1} L)w_{t,j}. \end{aligned} \tag{16}$$

All the AR and MA parameters in (16) are set as follows:

$$\phi_1 = \theta_1 = \phi_{2,1} = \theta_{2,1} = \frac{1}{2}. \tag{17}$$

The corresponding values of $\widetilde{D}$ when $M = 0$ is chosen are presented in Table 3, while those of $\widetilde{D}$ as $M = 1$ is selected are presented in Table 4. As compared to the results in Table 3 where $M = 0$ is used, the corresponding value of $\widetilde{D}$ in Table 4 does not always increase when $M = 1$ is employed, indicating that

'over-differencing' may not always result in efficiency loss of the MD estimator. This also implies that, when there is only one stochastic regressor, we may try different values of M and pick up the one producing a minimum value of $\widehat{D}$ in Theorem 1 as the choice of M for the MD estimator.

The preceding results are based on the asymptotic variance of the MD estimator. If the finite sample performance of the MD estimator is satisfactory, then the value of $T \times \text{MSE} = D^{\dagger}$ should be very close to the value of $\widetilde{D}$ in Tables 1-4.[4] To verify this conjecture and check the finite sample performance of the MD estimator, we further compute the MSE of the MD estimator under the specifications in Tables 1-4 based on 1,000 replications with $T = 200$.[5] We find that the value of $D^{\dagger}$ is very close to that of $\widetilde{D}$ in Tables 1-4, even though the sample size is only 200, indicating that the MD estimator not only can be easily implemented, but also possesses excellent finite sample performance.

3.2. Asymptotic local power analysis

This subsection explores the asymptotic local power of the difference-based method via a sequence of local DGP's in what Davidson and MacKinnon (1985) call "regression direction".[6] A limited Monte Carlo experiment based on 10,000 replications is conducted to verify the theoretical finding generated from the asymptotic local power analysis. The DGP considered in Tables 6 and 7 is arranged as follows:

$$C_t = \beta_1 Z_{t,1} + \beta_2 Z_{t,2} + U_t, \qquad \text{where} \qquad \beta_1 = \beta_2, \tag{18}$$

$$\begin{cases} U_t = \triangle^{-d_U} v_t^*, \quad Z_{t,2} = \triangle^{-d_{Z_2}} w_{t,2}^*, \ d_U = 0.2, \ d_{Z_2} = \{0.1, 0.3\}, \\ Z_{t,1} = \triangle^{-q} b_{t,1}^{\#}, \quad b_{t,1}^{\#} = \left(\triangle^{-d_{Z_1^*}} w_{t,1}^*\right) \mathbf{1}(t \geq 1), \ d_{Z_1^*} \in \left[-\frac{1}{2}, \frac{1}{2}\right), \ q \geq 1, \end{cases} \tag{19}$$

[4] MSE denotes the simulated mean-squared error (MSE) of the MD estimator generated from a Monte Carlo experiment.

[5] Throughout this paper, a $(T \times 1)$ value of a stationary I(d) process is generated with the Cholesky decomposition algorithm suggested by McLeod and Hipel (1978) and Hosking (1984), and two hundred additional values are generated in order to obtain random starting values. For a Type I nonstationary $I(q+d)$ process, we just partially sum a stationary $I(d)$ q times. All the programs are written in GAUSS language.

[6] Please also see Godfrey and Wickens (1982) for the related literature.

$$
\begin{aligned}
(1-\frac{1}{2}L)v_t^* &= (1+\frac{1}{2}L)v_t, \\
(1-\frac{1}{2}L)w_{t,1}^* &= (1+\frac{1}{2}L)w_{t,1}, \\
(1-\frac{1}{2}L)w_{t,2}^* &= (1+\frac{1}{2}L)w_{t,2},
\end{aligned} \tag{20}
$$

where v_t, $w_{t,1}$, and $w_{t,1}$ are all zero-mean n.i.d. processes with unit variance as that defined in (13), and v_t, $w_{t,1}$, and $w_{t,2}$ are independent with each other. It follows that D is a diagonal matrix throughout this subsection.

We include both nonstationary and stationary regressors in (18), (19), and (20) to demonstrate the coverage of Theorem 1. For clarity of illustration, $Z_{t,1}$ is always generated as a nonstationary process, and $Z_{t,2}$ is certainly stationary, i.e., $d_{Z_1} \geq \frac{1}{2}$ and $d_{Z_2} < \frac{1}{2}$ are imposed throughout the numerical investigation. As shown previously, intermediate memory regressors are allowed by the regression model in (4) but are not considered in the simulation studies. Since the maximum value of d_{Z_1} used in the experiment is 1.9, $M = 2$ is chosen for the entire simulation studies, even though $M = 1$ is large enough to deal with the cases where $d_{Z_1} < 1$. The flexibility of choosing the value of M is clearly borne out in our experiment.

Theorem 1 shows that each element of β, β_j, can be tested with the t-ratio statistic generated from the results in Theorem 1 and the null hypothesis tested is displayed as follows:

$$
H_0 : \beta_j = \beta_{0,j}, \qquad \text{for} \quad j = 1, 2, \tag{21}
$$

where $\beta_{0,j}$ is the j-th element of a column vector β_0. These two individual t-ratio statistics are jointly named as the t_{MD} test hereafter. To better understand the finite sample performance of the t_{MD} test, we follow Davidson and MacKinnon (1985) to examine the asymptotic local power of the t_{MD} test when the DGP is a sequence of the form:

$$
\triangle^M C_t = \triangle^M Z_t' \beta_0 + \triangle^M Z_t' \left(\delta_0 / \sqrt{T} \right) + \triangle^M U_t, \qquad t = M+1, M+2, \ldots, T, \tag{5'}
$$

where δ_0 is a $(k \times 1)$ finite constant.

Note that the t_{MD} test for the j-th regressor is named as the $t_{MD,j}$ statistic and is calculated as follows:

$$
t_{MD,j} \equiv \frac{\sqrt{T}\left(\widehat{\beta}_{MD,j} - \beta_{0,j} \right)}{\sqrt{\widehat{D}_{j,j}}}, \qquad \text{for} \qquad j = 1, 2,
$$

where $\widehat{D}_{j,j}$ denotes the row j, column j element of $\widehat{D}$ in Theorem 1. Given that β_j is generated as $\beta_{0,j} + T^{-1/2}\delta_{0,j}$ as implied in (5′), $\delta_{0,j}$ is the j-th element of δ_0, and then asymptotically:

$$\widehat{D}_{j,j} \xrightarrow{p} D_{j,j}, \qquad t_{MD,j} \xrightarrow{d} N(D_{j,j}^{-1/2}\delta_{0,j}, 1), \tag{22}$$

where $\xrightarrow{p}$ denotes convergence in probability. This implies that the power of the $t_{MD,j}$ test grows with the increasing value of $|\delta_{0,j}|$ for a fixed sample size T and a fixed $D_{j,j}$. Of course, the power of the $t_{MD,j}$ test is equivalent to its own size as $\delta_{0,j} = 0$. Furthermore, for a fixed value of $\delta_{0,j}$ and a fixed T, the power of the $t_{MD,j}$ test will increase with a decreasing value of $D_{j,j}$. To investigate the finite sample performance of the MD estimator more closely, we present the corresponding values of $\widetilde{D}_{j,j}$ in Table 5.

To confirm the preceding local power properties of the $t_{MD,j}$ test, we investigate the small sample performance of the $t_{MD,j}$ test at the 5% level of significance, and the simulation results are contained in Tables 6 and 7. Four different values of coefficients (i.e., $\beta_1 = \beta_2 = 1.0$, 0.9, 0.8, and 0.7) in (18) are considered in the experimental studies. The values of $\beta_{0,j}$ in (21) are always set to be 1 for all $j = 1, 2$ considered in this subsection.

The simulation results correspond exactly to the size of the $t_{MD,j}$ test, when the data are generated by the condition $\beta_1 = \beta_2 = 1$. These tables reveal that the size control of the $t_{MD,j}$ test is quite well even when the sample size is only 100. The worst size distortion found within these tables is only 1.55%.[7] When T increases to be 200, the size control of the $t_{MD,j}$ test improves significantly, no matter if the regressor is a nonstationary or stationary process. This phenomenon supports the prediction made in Theorem 1 which claims that $\widehat{\beta}_{MD}$ is asymptotically normally distributed and converges at a rate of $\sqrt{T}$, which we frequently encounter with the model consisting of weakly dependent regressors and the error term. This also reveals that Robinson's (1998) long-run variance estimator works effectively in controlling the nuisance parameters embodied in the intermediate memory processes generated from the multiple-differenced transformation.

We now consider the power performance of the $t_{MD,j}$ test under three alternatives, i.e., $\beta_1 = \beta_2 = 0.7, 0.8$, and 0.9. According to the local power properties of the $t_{MD,j}$ test discussed above, we note that

[7] Please confirm the configuration $d_{Z_1} = 0.9$ and $d_{Z_2} = 0.10$ in Table 6 for testing the value of β_1.

$T^{-1/2}\delta_{0,j} = -0.1$ when $\beta_j = 0.9$, $T^{-1/2}\delta_{0,j} = -0.2$ when $\beta_j = 0.8$, and $T^{-1/2}\delta_{0,j} = -0.3$ when $\beta_j = 0.7$. Thus, for a fixed alternative ($\beta_j = 0.9$, or 0.8, or 0.7), *ceteris paribus*, the value of $|\delta_{0,j}|$ increases with an increasing value of T. At the same time, the power of the $t_{MD,j}$ test should improve as T becomes larger. This prediction is completely supported in Tables 6 and 7. As compared to the results in Table 6 where $T = 100$, the power of the $t_{MD,j}$ test is found to increase for any configuration in Table 7 where the sample size is now 200.

Provided that $D_{j,j}$ can be consistently estimated with Robinson's (1998) long-run variance estimator, other things being equal, the power of the $t_{MD,j}$ test is inversely related to the value of $\widetilde{D}_{j,j}$ in Table 5. This theoretical finding is strongly supported by Table 6 and Table 7. In particular, the power of the $t_{MD,j}$ test always reaches its maximum when d_{Z_j} is 1.9 under all alternatives and sample sizes considered, while the occurrence of the minimum power is always located at $d_{Z_j} = 0.9$ under all alternatives and the sample size is 200. We find only two cases where the pattern of power change is not what we expect, i.e., when testing the hypothesis $\beta_1 = 1$, but the true value of β_1 is 0.9 and $T = 100$, the power of the MD estimator as $d_{Z_1} = 0.9$ is found to be higher than that of the MD estimator as $d_{Z_1} = 1.1$ in Table 6. Overall, with the assistance of Robinson's (1998) long-run variance estimator, the difference-based method works effectively for a large variety of DGP considered in the experiment.

4. Conclusions

This paper considers a multiple regression model allowing the joint presence of stationary and nonstationary fractionally-integrated processes, and it provides a unifying framework to deal with the inference problems associated with this model. Provided that the conditions in Theorem 1 are satisfied, we show that each regression coefficient can be tested with the usual t-ratio statistic generated from our Theorem 1, no matter if the targeted regressor is nonstationary or non-invertible. Most importantly, we do not need to simulate different sets of critical values for various combinations of d_Z, d_U, and sample size which are frequently encountered in the inference problems associated with the presence of nonstationary processes in the regression model. This feature is especially valuable, because in practice the fractional differencing

parameters of the regressors and that of the error term cannot be estimated without bias, especially when the sample size is relatively moderate. Moreover, building on the easy-to-implement but elegant long-run variance estimator of Robinson (1998), the computational cost of our inference method is extremely mild and can be conducted with standard statistics packages. The simulations also reveal that the size control of our method is very good, although the sample size is only 100, and the pattern of power performance is highly consistent with the theoretical finding from the asymptotic local power analysis conducted in this paper.

REFERENCES

Anderson, T.W., 1971. The Statistical Analysis of Time Series. New York: Wiley.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent matrix estimation. Econometrica 59, 817-858.

Berk, K.N., 1974. Consistent autoregressive spectral estimates. Annals of Statistics 2, 489-502.

Davidson, R., MacKinnon, J.G., 1985. The interpretation of test statistics. Canadian Journal of Economics 18, 38-57.

Godfrey, L.G., Wickens, M.R., 1982. Tests of misspecification using locally equivalent alternative models. In: Chow, G.C., Corsi, P. (Eds.), Evaluating the Reliability of Econometric Models. John Wiley, New York, 71-103.

Hansen, B.E., 1992. Consistent covariance matrix estimation for dependent heterogeneous processes. Econometrica 60, 967-972.

Hidalgo, J., 2000. Nonparametric test for causality with long-range dependence. Econometrica 68, 1465-1490.

Hosking, J.R.M., 1984. Modeling persistence in hydrological time series using fractional differencing. Water Resources Research 20, 1898-1908.

Marinucci, D., 2000. Spectral regression for cointegration time series with long-memory innovations. Journal of Time Series Analysis 21, 685-705.

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.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-708.

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

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with $I(1)$ processes. Review of Economic Studies 57, 99-125.

Robinson, P.M., 1998. Inference-without-smoothing in the presence of nonparametric autocorrelation. Econometrica 66, 1163-1182.

Robinson, P.M., 2005. The distance between rival nonstationary fractional processes. Journal of Econometrics 128, 283-300.

Robinson, P.M., Hidalgo, F.J., 1997. Time series regression with long range dependence. Annals of Statistics 25, 77-104.

Sowell, F.B., 1990. The fractional unit root distribution. Econometrica 58, 495-505.

Tsay, W. J., 2000. Estimating trending variables in the presence of fractionally integrated errors. Econometric Theory 16, 324-346.

Velasco, C., Robinson, P.M., 2000. Whittle pseudo-maximum likelihood estimation for nonstationary time series. Journal of the American Statistical Association 95, 1229-1243.

Table 1. The Approximate Asymptotic Variance of MD Estimator and the Corresponding Finite Sample Values when DGP are ARFIMA$(0, d, 0)$ processes

		d_{Z_j}								
d_U		-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9
-0.1	$\widetilde{D}$	1.0226	1.0015	0.9614	0.9096	0.8513	0.7899	0.7281	0.6673	0.6088
	$D^\dagger$	1.0386	1.0071	0.9588	0.9004	0.8369	0.7717	0.7071	0.6446	0.5852
-0.2	$\widetilde{D}$	1.0779	1.0681	1.0346	0.9859	0.9281	0.8653	0.8007	0.7364	0.6738
	$D^\dagger$	1.0951	1.0745	1.0325	0.9769	0.9134	0.8464	0.7786	0.7122	0.6484
-0.3	$\widetilde{D}$	1.1496	1.1495	1.1214	1.0749	1.0166	0.9516	0.8836	0.8149	0.7476
	$D^\dagger$	1.1681	1.1570	1.1201	1.0662	1.0020	0.9323	0.8607	0.7896	0.7207
-0.4	$\widetilde{D}$	1.2372	1.2459	1.2226	1.1774	1.1180	1.0501	0.9778	0.9042	0.8313
	$D^\dagger$	1.2573	1.2548	1.2224	1.1695	1.1037	1.0307	0.9544	0.8779	0.8030
-0.5	$\widetilde{D}$	1.3410	1.3583	1.3392	1.2948	1.2337	1.1621	1.0849	1.0054	0.9261
	$D^\dagger$	1.3632	1.3689	1.3405	1.2881	1.2201	1.1429	1.0612	0.9784	0.8968
-0.6	$\widetilde{D}$	1.4619	1.4878	1.4728	1.4288	1.3653	1.2894	1.2063	1.1201	1.0336
	$D^\dagger$	1.4866	1.5007	1.4760	1.4236	1.3529	1.2708	1.1827	1.0926	1.0033
-0.7	$\widetilde{D}$	1.6013	1.6362	1.6252	1.5813	1.5148	1.4337	1.3440	1.2501	1.1553
	$D^\dagger$	1.6292	1.6519	1.6308	1.5781	1.5039	1.4161	1.3208	1.2225	1.1244
-0.8	$\widetilde{D}$	1.7611	1.8056	1.7986	1.7544	1.6844	1.5973	1.5000	1.3973	1.2932
	$D^\dagger$	1.7926	1.8245	1.8072	1.7538	1.6755	1.5811	1.4776	1.3699	1.2619
-0.9	$\widetilde{D}$	1.9433	1.9982	1.9955	1.9508	1.8766	1.7827	1.6766	1.5641	1.4493
	$D^\dagger$	1.9792	2.0211	2.0077	1.9533	1.8703	1.7685	1.6555	1.5373	1.4181

Notes: $\widetilde{D}$ denotes the approximated value of D in (11) with the formula in (14). The corresponding finite sample observation, $D^\dagger$, is based on the value $T \times$ MSE generated from the 1,000 replications of the simulated data defined in (12) and (13) with sample size $T = 200$ and $M = 0$ is chosen.

Table 2. The Approximate Asymptotic Variance of MD Estimator and the Corresponding Finite Sample Values when DGP are ARFIMA$(0, d, 0)$ processes

d_U		d_{Z_j} -1.1	-1.2	-1.3	-1.4	-1.5	-1.6	-1.7	-1.8	-1.9
-1.1	$\widetilde{D}$	1.5494	1.4174	1.2920	1.1739	1.0636	0.9613	0.8669	0.7802	0.7009
	$D^\dagger$	1.5191	1.3838	1.2561	1.1369	1.0263	0.9245	0.8312	0.7461	0.6688
-1.2	$\widetilde{D}$	1.7448	1.5975	1.4571	1.3248	1.2010	1.0861	0.9799	0.8823	0.7930
	$D^\dagger$	1.7164	1.5646	1.4212	1.2871	1.1625	1.0476	0.9422	0.8460	0.7585
-1.3	$\widetilde{D}$	1.9664	1.8016	1.6444	1.4959	1.3570	1.2277	1.1083	0.9984	0.8977
	$D^\dagger$	1.9407	1.7704	1.6091	1.4580	1.3175	1.1878	1.0687	0.9598	0.8608
-1.4	$\widetilde{D}$	2.2177	2.0331	1.8568	1.6902	1.5340	1.3886	1.2540	1.1302	1.0167
	$D^\dagger$	2.1957	2.0044	1.8229	1.6526	1.4941	1.3475	1.2128	1.0896	0.9773
-1.5	$\widetilde{D}$	2.5028	2.2958	2.0979	1.9106	1.7349	1.5712	1.4196	1.2799	1.1518
	$D^\dagger$	2.4855	2.2705	2.0662	1.8742	1.6952	1.5295	1.3771	1.2375	1.1103
-1.6	$\widetilde{D}$	2.8261	2.5939	2.3716	2.1609	1.9631	1.7786	1.6077	1.4501	1.3054
	$D^\dagger$	2.8151	2.5732	2.3431	2.1264	1.9242	1.7369	1.5643	1.4062	1.2619
-1.7	$\widetilde{D}$	3.1931	2.9323	2.6822	2.4451	2.2222	2.0142	1.8214	1.6434	1.4800
	$D^\dagger$	3.1897	2.9175	2.6581	2.4136	2.1851	1.9731	1.7777	1.5984	1.4348
-1.8	$\widetilde{D}$	3.6095	3.3164	3.0350	2.7679	2.5166	2.2820	2.0642	1.8632	1.6785
	$D^\dagger$	3.6158	3.3092	3.0167	2.7405	2.4822	2.2423	2.0209	1.8177	1.6320
-1.9	$\widetilde{D}$	4.0824	3.7525	3.4356	3.1346	2.8512	2.5863	2.3403	2.1131	1.9042
	$D^\dagger$	4.1003	3.7549	3.4248	3.1128	2.8206	2.5490	2.2982	2.0677	1.8569

Notes: $\widetilde{D}$ denotes the approximated value of D in (11) with the formula in (14). The corresponding finite sample observation, $D^\dagger$, is based on the value $T \times$ MSE generated from the 1,000 replications of the simulated data defined in (12) and (13) with sample size $T = 200$ and $M = 1$ is chosen.

Table 3. The Approximate Asymptotic Variance of MD Estimator and the Corresponding Finite Sample Values when DGP are ARFIMA$(1, d, 1)$ processes

d_U		d_{Z_j} -0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9
-0.1	$\widetilde{D}$	2.0011	2.1282	2.1798	2.1681	2.1064	2.0077	1.8841	1.7454	1.5997
	$D^\dagger$	2.1045	2.2292	2.2778	2.2571	2.1808	2.0648	1.9248	1.7734	1.6195
-0.2	$\widetilde{D}$	1.6336	1.7849	1.8710	1.8988	1.8779	1.8189	1.7317	1.6255	1.5080
	$D^\dagger$	1.7376	1.8862	1.9699	1.9908	1.9575	1.8826	1.7792	1.6590	1.5309
-0.3	$\widetilde{D}$	1.3701	1.5320	1.6385	1.6929	1.7013	1.6718	1.6129	1.5324	1.4375
	$D^\dagger$	1.4714	1.6319	1.7376	1.7870	1.7851	1.7414	1.6667	1.5715	1.4645
-0.4	$\widetilde{D}$	1.1760	1.3417	1.4608	1.5336	1.5637	1.5571	1.5204	1.4607	1.3843
	$D^\dagger$	1.2738	1.4398	1.5597	1.6293	1.6511	1.6317	1.5801	1.5055	1.4160
-0.5	$\widetilde{D}$	1.0299	1.1962	1.3234	1.4096	1.4564	1.4678	1.4493	1.4067	1.3459
	$D^\dagger$	1.1244	1.2924	1.4217	1.5063	1.5464	1.5467	1.5142	1.4569	1.3823
-0.6	$\widetilde{D}$	0.9182	1.0837	1.2164	1.3129	1.3729	1.3991	1.3957	1.3676	1.3201
	$D^\dagger$	1.0093	1.1778	1.3137	1.4099	1.4649	1.4814	1.4651	1.4228	1.3612
-0.7	$\widetilde{D}$	0.8318	0.9960	1.1329	1.2376	1.3087	1.3474	1.3570	1.3414	1.3054
	$D^\dagger$	0.9195	1.0877	1.2287	1.3343	1.4018	1.4323	1.4301	1.4010	1.3511
-0.8	$\widetilde{D}$	0.7646	0.9277	1.0680	1.1797	1.2603	1.3100	1.3310	1.3265	1.3008
	$D^\dagger$	0.8485	1.0164	1.1616	1.2754	1.3538	1.3967	1.4071	1.3899	1.3506
-0.9	$\widetilde{D}$	0.7121	0.8745	1.0180	1.1361	1.2253	1.2849	1.3162	1.3218	1.3054
	$D^\dagger$	0.7918	0.9597	1.1088	1.2301	1.3183	1.3725	1.3946	1.3884	1.3590

Notes: $\widetilde{D}$ denote the approximated value of D in (11) with the formula in (14). The corresponding finite sample observation, $D^\dagger$, is based on the value $T \times$ MSE generated from the 1,000 replications of the simulated data defined in (13), (15), (16), and (17) with sample size $T = 200$ and $M = 0$ is chosen.

Table 4. The Approximate Asymptotic Variance of MD Estimator and the Corresponding Finite Sample Values when DGP are ARFIMA$(1, d, 1)$ processes

d_U		d_{Z_j} -1.1	-1.2	-1.3	-1.4	-1.5	-1.6	-1.7	-1.8	-1.9
-1.1	$\widetilde{D}$	1.2840	1.2345	1.1753	1.1093	1.0392	0.9670	0.8947	0.8234	0.7544
	$D^\dagger$	1.3212	1.2592	1.1893	1.1145	1.0373	0.9597	0.8832	0.8089	0.7377
-1.2	$\widetilde{D}$	1.3264	1.2813	1.2253	1.1613	1.0921	1.0201	0.9470	0.8744	0.8036
	$D^\dagger$	1.3684	1.3107	1.2434	1.1698	1.0928	1.0144	0.9363	0.8600	0.7864
-1.3	$\widetilde{D}$	1.3767	1.3358	1.2827	1.2205	1.1520	1.0797	1.0056	0.9315	0.8585
	$D^\dagger$	1.4233	1.3698	1.3051	1.2327	1.1556	1.0762	0.9964	0.9177	0.8413
-1.4	$\widetilde{D}$	1.4353	1.3985	1.3481	1.2874	1.2194	1.1466	1.0712	0.9951	0.9196
	$D^\dagger$	1.4863	1.4371	1.3750	1.3037	1.2265	1.1458	1.0639	0.9826	0.9030
-1.5	$\widetilde{D}$	1.5026	1.4698	1.4220	1.3627	1.2949	1.2214	1.1444	1.0659	0.9876
	$D^\dagger$	1.5576	1.5129	1.4535	1.3834	1.3058	1.2238	1.1396	1.0552	0.9721
-1.6	$\widetilde{D}$	1.5792	1.5504	1.5052	1.4472	1.3794	1.3048	1.2258	1.1448	1.0633
	$D^\dagger$	1.6379	1.5978	1.5413	1.4723	1.3944	1.3108	1.2240	1.1363	1.0493
-1.7	$\widetilde{D}$	1.6658	1.6410	1.5985	1.5415	1.4736	1.3977	1.3165	1.2324	1.1474
	$D^\dagger$	1.7279	1.6927	1.6391	1.5713	1.4930	1.4077	1.3181	1.2267	1.1354
-1.8	$\widetilde{D}$	1.7632	1.7427	1.7027	1.6468	1.5785	1.5011	1.4174	1.3299	1.2408
	$D^\dagger$	1.8282	1.7982	1.7477	1.6812	1.6025	1.5154	1.4228	1.3274	1.2314
-1.9	$\widetilde{D}$	1.8724	1.8563	1.8189	1.7640	1.6952	1.6160	1.5294	1.4382	1.3446
	$D^\dagger$	1.9400	1.9153	1.8683	1.8031	1.7240	1.6349	1.5390	1.4393	1.3383

Notes: $\widetilde{D}$ denotes the approximated value of D in (11) with the formula in (14). The corresponding finite sample observation, $D^\dagger$, is based on the value $T \times$ MSE generated from the 1,000 replications of the simulated data defined in (13), (15), (16), and (17) with sample size $T = 200$ and $M = 1$ is chosen.

Table 5. The Approximate Long-Run Variance $\widetilde{D}_{j,j}$ when $Z_{t,j}$ and U_t, are Generated by (18), (19), and (20)

d_{Z_j}									
0.1	0.3	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9
1.2408	1.4174	1.5785	1.7027	1.7632	1.7320	1.5872	1.3245	0.9704	0.5870

Notes: The results are all based on the model in (18), $M = 2$, and U_t and $Z_{t,j}$ are generated according to (19) and (20). $\widetilde{D}_{j,j} = \sum_{i=-800}^{800} \gamma_j(i)\gamma_U(i)$, where $\gamma_j(i)$ denotes the autocovariance function of $\triangle^M Z_{t,j}$ at lag i, and $\gamma_U(i)$ denotes the autocovariance function of $\triangle^M U_t$ at lag i, respectively.

Table 6. Rejection Percentages of the t_{MD} Test at the 5% Level of Significance

					d_{Z_1}				
$d_U = 0.2$	$T = 100$ Alternatives	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9
$d_{Z_2} = 0.30$									
True null	$\beta_1 = 1.0$	6.36	6.49	6.51	6.47	6.06	5.93	5.56	4.80
	$\beta_1 = 0.9$	14.07	13.41	13.08	12.97	13.52	14.65	18.27	25.51
	$\beta_1 = 0.8$	36.73	34.48	33.05	33.68	35.96	41.38	52.39	70.42
	$\beta_1 = 0.7$	65.89	62.73	61.51	62.09	65.51	72.52	82.84	93.60
True null	$\beta_2 = 1.0$	6.15	6.27	6.28	6.20	6.24	6.22	6.22	6.19
	$\beta_2 = 0.9$	15.79	15.71	15.71	15.74	15.71	15.68	15.64	15.63
	$\beta_2 = 0.8$	41.13	41.09	41.13	41.18	41.14	41.18	41.11	41.21
	$\beta_2 = 0.7$	71.59	71.58	71.61	71.54	71.55	71.48	71.43	71.49
$d_{Z_2} = 0.10$									
True null	$\beta_1 = 1.0$	6.47	6.46	6.55	6.48	6.10	5.86	5.44	4.79
	$\beta_1 = 0.9$	14.05	13.35	13.07	12.88	13.56	14.69	18.25	25.65
	$\beta_1 = 0.8$	36.71	34.47	33.09	33.72	36.13	41.49	52.35	70.48
	$\beta_1 = 0.7$	65.88	62.69	61.46	62.09	65.46	72.59	82.99	93.61
True null	$\beta_2 = 1.0$	6.26	6.26	6.24	6.27	6.29	6.35	6.34	6.25
	$\beta_2 = 0.9$	17.11	17.03	16.88	16.82	16.85	16.84	16.76	16.68
	$\beta_2 = 0.8$	45.47	45.49	45.59	45.56	45.52	45.45	45.47	45.38
	$\beta_2 = 0.7$	76.68	76.67	76.69	76.69	76.60	76.54	76.54	76.66

Notes: The results are all based on 10,000 replications with $M = 2$. The data are generated based on (18), (19), and (20). The null hypotheses $\beta_1 = 1$ and $\beta_2 = 1$ are tested separately.

Table 7. Rejection Percentages of the t_{MD} Test at the 5% Level of Significance

		d_{Z_1}							
$d_U = 0.2$	$T = 200$ Alternatives	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9
$d_{Z_2} = 0.30$									
True null	$\beta_1 = 1.0$	6.05	6.12	6.04	5.85	6.08	5.84	5.63	5.07
	$\beta_1 = 0.9$	20.52	19.31	18.76	18.90	20.32	23.19	29.83	44.44
	$\beta_1 = 0.8$	60.57	57.02	55.15	56.21	59.93	67.57	79.70	93.40
	$\beta_1 = 0.7$	91.13	89.03	87.88	88.20	90.58	94.27	98.27	99.75
True null	$\beta_2 = 1.0$	5.80	5.84	5.74	5.76	5.81	5.80	5.82	5.83
	$\beta_2 = 0.9$	23.73	23.70	23.59	23.55	23.51	23.49	23.44	23.54
	$\beta_2 = 0.8$	67.00	67.02	66.97	67.00	66.91	66.85	66.92	66.91
	$\beta_2 = 0.7$	94.20	94.15	94.18	94.10	94.10	94.13	94.12	94.16
$d_{Z_2} = 0.10$									
True null	$\beta_1 = 1.0$	5.95	6.12	6.01	5.81	6.08	5.90	5.62	5.08
	$\beta_1 = 0.9$	20.49	19.35	18.90	18.91	20.34	23.08	29.87	44.40
	$\beta_1 = 0.8$	60.58	57.13	55.22	56.22	60.00	67.49	79.72	93.44
	$\beta_1 = 0.7$	91.10	89.08	87.89	88.29	90.59	94.29	98.33	99.75
True null	$\beta_2 = 1.0$	5.76	5.80	5.85	5.89	5.84	5.78	5.74	5.78
	$\beta_2 = 0.9$	26.06	26.08	25.96	25.82	26.03	26.00	25.95	26.00
	$\beta_2 = 0.8$	72.20	72.30	72.40	72.28	72.28	72.26	72.22	72.28
	$\beta_2 = 0.7$	96.40	96.38	96.39	96.37	96.39	96.30	96.35	96.36

Notes: The results are all based on 10,000 replications with $M = 2$. The data are generated based on (18), (19), and (20). The null hypotheses $\beta_1 = 1$ and $\beta_2 = 1$ are tested separately.