Fractional Dickey-Fuller tests under heteroskedasticity

Hsein Kew and David Harris
Department of Economics
University of Melbourne

May 29, 2007

Abstract

In a recent paper, Dolado, Gonzalo and Mayoral (2002) introduce a fractional Dickey-Fuller (FD-F) t-statistic for testing a unit root against the alternative of a mean reverting fractional unit root process. This t-statistic is based on the assumption that the errors are unconditionally homoskedastic. However, Busetti and Taylor (2003), McConnell and Perez-Quiros (2000), and van Dijk et al. (2002) have found compelling evidence that such an assumption is unlikely to hold in many macroeconomic and financial time series, especially those obtained at a longer time span. In this paper, we investigate the finite-sample properties of the FD-F statistic when the errors are unconditionally heteroskedastic. We find that, depending on the form of heteroskedasticity, the FD-F statistic suffers from substantial size distortions. In order to correct for such distortions, we propose the use of White standard errors (White (1980)) when computing the FD-F statistic. This yields a test that is robust to heteroskedasticity of unknown form. We demonstrate that the FD-F statistic that employs White standard error has a standard normal limiting distribution under the unit root null hypothesis as in the FD-F statistic with homoskedastic errors. Monte Carlo results suggest that: (i) White's correction is effective in reducing the size distortions; and (ii) the power loss of using White standard error in the case of homoskedasticity is very small. These results suggest that it is prudent to use the White robust standard errors regardless of whether the errors are heteroskedastic or not.

Keywords: Fractional unit root tests, heteroskedasticity, structural breaks
J.E.L. Classification: C30, C32

1 Introduction

Several articles have provided Monte Carlo evidence on the performance of the unit-root tests, stationary tests and persistence change tests under unconditional heteroskedasticity. The general finding from these articles is that these tests suffer from severe size distortions when there is an abrupt change in the error variance. In standard unit root testing against stationarity alternatives, Hamori and Tokihisa (1997) study the Dickey-Fuller (D-F) test, Cavaliere and Taylor (2007) study the M test of Stock (1999) and Beare (2007) investigates the Phillips-Perron test. In the stationarity testing framework, where the null hypothesis of stationarity is tested against a unit root alternative, Cavaliere (2004) examines the properties of the KPSS test, and Busetti and Taylor (2003) examines the properties of the test statistics due to Busetti and Harvey (2001). In the persistence change testing framework, where statistical procedures are developed to test the null hypothesis of a constant unit root process against the alternative of a change in persistence either from a stationary process to a unit root process or vice versa, Cavaliere and Taylor (2006) investigate the properties of the ratio-based tests introduced by Kim (2000), Kim *et al.* (2002) and Busetti and Taylor (2004).

However, to the best of our knowledge, no study has examined the effects of unconditional heteroskedasticity on existing statistical procedures designed to test the null hypothesis of a unit root process against the alternative of a mean reverting fractional process. It is well known that although the D-F test is consistent against such an alternative, the power of D-F test is generally quite low. To circumvent this problem, Dolado, Gonzalo and Mayoral (2002, hereinafter DGM) recently introduced the Fractional Dickey-Fuller (FD-F) test for a unit root that has high power against fractional alternatives. Let $\Delta = (1 - L)$, where L is the lag operator. The FD-F statistical procedure is based on the usual t-statistic of ϕ in the ordinary least squares (OLS) regression $\Delta y_t = \phi \Delta^{d_1} y_{t-1} + e_t$. The null hypothesis that y_t has a unit root is then $H_0 : \phi = 0$ against $H_1 : \phi < 0$. If $\phi = 0$, then $\Delta y_t = e_t$ and so y_t has a unit root. Under the alternative that $\phi < 0$, it can shown that y_t is fractionally integrated of order d_1. The regression cannot be made operational without a value of d_1. DGM demonstrate that if the value of d_1 is obtained by using a $\sqrt{T}$-consistent estimator of the true fractional differencing parameter, the resulting t-statistic of ϕ has a limiting standard normal distribution under the unit root null hypothesis.

In DGM, the error term e_t is assumed to be independent and unconditionally homoskedastic. In this paper, we investigate the robustness of this t-statistic in the presence of unconditional heteroskedastic errors. Our Monte Carlo experiments show that, depending on the form of heteroskedasticity, the FD-F statistic suffers from substantial size distortions. In order to correct for such distortions, we propose to replace the usual OLS homoskedastic standard errors by White's heteroskedasticity robust standard errors when constructing the t-statistic (White (1980)). We demonstrate that the FD-F statistic that employs White standard errors still retains its standard normal limiting distribution. We obtain this result by exploiting the fact that under the unit root null hypothesis the FD-F t-statistic, unlike the standard unit root tests, possesses standard limit theory.

White's results are established under the assumption that the regressors are exogenous. Nicholls and Pagan (1983) and more recently Phillips and Xu (2006) show that White's results remain valid in a dynamic regression model, provided that the time series are covariance stationary. Our results therefore illustrate the general applicability of White's results in unit root testing literature.

Robustified versions of the unit root tests mentioned above have been suggested in the time series literature. Kim, Leybourne and Newbold (2002) suggest pre-estimating the variance break point together with the pre- and post-break variances. These estimates are then employed in modified variants of the Perron-type unit root tests (Perron (1989, 1990)).

The critical values for these modified tests are given in Perron (1989), and they depend on the location of the break. Kim, Leybourne and Newbold (2002) assume the existence of a *single* break in the error variance process. In a recent paper, Cavaliere and Taylor (2007), based on the work of Cavaliere (2004), relax this assumption. Their procedure does not require a parametric specification of the error variance, but only requires the error variance to be uniformly bounded and to display a countable number of jumps. They demonstrate that heteroskedasticity induces a time-deformation in the limiting distribution of the unit root statistics. Using a consistent non-parametric estimator of this time-deformation, the resulting limiting distribution can then be simulated to obtain asymptotically valid critical values. Thus the correct critical values to use depend on the precise nature of the heteroskedasticity. More recently, Beare (2007) suggests an alternative approach that does not require case-by-case numerical tabulation of critical values. Beare's approach is to transform the data in such a way that the transformed data are approximately homoskedastic. The Phillips-Perron unit root test is then applied to the homoskedastic transformed data. This yields, under conditions more restrictive than those used by Cavaliere and Taylor (2007), a unit root test which has a pivotal asymptotic null distribution.

Since the FD-F test involves standard limit theory, our solution to the size distortion problem is very simple to implement in practice, is robust against unknown heteroskedasticity, and does not require pre-estimation of some non-parametric functions of the heteroskedasticity. Unlike Cavaliere and Taylor (2007), our error variance is not required to display countable number of jumps. We only assume that the error variance is uniformly bounded away from zero and infinity. We note here that the procedures suggested by Kim, Leybourne and Newbold (2002) and Cavaliere and Taylor (2007) allow for the presence of deterministic components in the data generating process, while our procedure assumes that the series contains no deterministic components. Nevertheless, our procedure could be extended in a straightforward manner to include deterministic regressors.

Throughout this paper, the symbols $\xrightarrow{p}$ and $\xrightarrow{d}$ denote, respectively, convergence in probability and in distribution. We write $X_t = O_p(1)$ to denote a stochastic sequence $\{X_t\}$ that is uniformly bounded in probability for all t. We write $X_t = o_p(1)$ to denote that X_t converges in probability to zero. The indicator function $\mathbf{1}_{(t \geq 1)}$ is equal to one if $t \geq 1$ or to zero otherwise. The notation $\lfloor x \rfloor$ denotes the largest integer below x. The notation $\Gamma(x)$ denotes the Euler gamma function defined for any real value of x except negative integers. We use $a \vee b$ as shorthand for $\max(a, b)$.

The rest of this paper is organised as follows. Section 2 presents the model and assumptions. Section 3 describes the FD-F statistical procedure and analyses the properties of this test under unconditional heteroskedasticity. The Monte Carlo experiments presented in Section 3.2 serve to illustrate that the presence of unconditional heteroskedasticity in the error term can cause invalid statistical inferences. Section 4 suggests a modified version of the FD-F t statistic that is robust to the presence of unconditional heteroskedasticity. It also establishes the asymptotic theory of this modified test. Monte Carlo evidence on the small-sample properties of this modified test are presented in Section 5. The final section concludes. The proofs are collected in the Appendix.

2 Model and Assumptions

Suppose the time series $\{y_t\}$ is generated by the fractionally integrated model:

$$\Delta^{d_0} y_t = e_t \quad (1)$$

$$e_t = \sigma_{Tt}\varepsilon_t \mathbf{1}_{(t\geq 1)} \quad (2)$$

where $d_0 \in [0,1]$. Since the fractional parameter d_0 can take on any real value rather than only an integer value, the fractional filter Δ^{d_0} in (1) can be expanded to obtain a truncated autoregressive representation for $\{y_t\}$:

$$\Delta^{d_0} y_t = \sum_{j=0}^{t-1} \pi_j(d_0)\, y_{t-j}$$

where $\pi_j(d_0) = \Gamma(j-d_0)/(\Gamma(j+1)\Gamma(-d_0))$. Starting at $\pi_0(d_0) = 1$, $\pi_j(d_0)$ can be written recursively as $\pi_j(d_0) = (\pi_{j-1}(d_0)(j-d_0-1))/j$ for $j \geq 1$. In (2), σ_{Tt} and ε_t are assumed to satisfy the following conditions.

Assumption V
σ_{Tt} is non-stochastic and satisfies for all t

$$0 < \underline{\sigma} < \sigma_{Tt} \leq \bar{\sigma} < \infty$$

where $\underline{\sigma}$ and $\bar{\sigma}$ are strictly positive constants.

Assumption E
ε_t is an i.i.d. process with $E(\varepsilon_t) = 0$, $E\left(\varepsilon_t^2\right) = 1$ and $\sup_t E|\varepsilon_t|^{2\delta} \leq B < \infty$ for some $\delta > 2$.

Under Assumption E, the error term e_t in (1) has zero unconditional mean. Assumption E normalises the variance of ε_t to unity so that (2) implies that $E\left(e_t^2\right) = \sigma_{Tt}^2$, signifying explicitly that the unconditional variance of the error term is not constant over time. The functional form of the error variance σ_{Tt}^2 is treated as unknown, and thus, our framework is non-parametric. By Liapunov's inequality and Assumption E, it follows that for $r \leq 4$, the error term e_t satisfies

$$\sup_t E|e_t|^r = \sup_t E|\sigma_{Tt}\varepsilon_t|^r = \sup_t \sigma_{Tt}^r E|\varepsilon_t|^r < \bar{\sigma}^r B^{\frac{r}{2\delta}} < \infty. \quad (3)$$

Under Assumptions V and E, the series $\{y_t\}$ in (1)-(2) is generated by a fractionally integrated model with time-varying innovation variances.

Since σ_{Tt} depends on T, formally we should write $e_t = e_{Tt}$ and thus $y_t = y_{Tt}$, so that the time series in $(1)-(2)$ form a triangular array. However, this extra subscript T does not play any role in the development of asymptotic theory, and is thus suppressed for notational simplicity.

Assumption V is slightly weaker than that of Cavaliere (2004) and Cavaliere and Taylor (2007). We do not require the error variance σ_t^2 to be imbedded in a variance function, to display a countable number of jumps, and to satisfy a Lipschitz condition. The independence assumption imposed in Assumption E is stronger than necessary and is adopted for simplicity. This assumption rules out the presence of conditional heteroskedasticity of the type introduced by Engle (1982) and Bollerslev (1986).

3 The Fractional Dickey-Fuller Test

The FD-F test is a regression based statistical procedure. It involves OLS estimation of the following regression equation:

$$\Delta y_t = \phi \Delta^{d_1} y_{t-1} + e_t. \tag{4}$$

For a given value of $d_1 \in [0,1)$, under the Type 2 definition of the fractionally integrated model (see Robinson and Marinucci (2001)), the regressor can be computed as $\Delta^{d_1} y_{t-1} = \sum_{j=0}^{t-2} \pi_j (d_1) y_{t-1-j}$. Notice that when $d_1 = 0$, equation (4) is the standard D-F regression. Like the D-F test, the FD-F test is the conventional t-statistic for testing the hypothesis $H_0 : \phi = 0$ against $H_1 : \phi < 0$ in (4).

Under the null, the process y_t has a unit root because the regression in (4) becomes $\Delta y_t = e_t$. Under the alternative that $\phi < 0$, the process y_t is fractionally integrated of order d_1, because DGM show that the regression in (4) can be rewritten as

$$\Delta^{d_1} y_t = C(L) e_t$$

where $C(L) = \left(\Delta^{1-d_1} - \phi L\right)^{-1}$ and they show that $C(L)$ does not contain unit roots. If we have some a priori knowledge as to the value of d_1, then it can be seen immediately that testing for the significance of ϕ in (4) is equivalent to testing the null hypothesis

$$H_0 : d_0 = 1$$

against the simple alternative

$$H_1 : d_0 = d_1. \tag{5}$$

Note that the standard D-F unit root testing procedure imposes a priori restriction on d_1 by setting $d_1 = 0$. This is theoretically justified only when the process y_t is taken to follow either an $I(1)$ process or an $I(0)$ process. The FD-F test eliminates such a restriction by generalising the D-F test to explicitly allow for the possibility that y_t is a mean reverting fractional integrated process $I(d)$. If the true data generating process for $\{y_t\}$ is indeed an $I(d_0)$ process with $0 < d_0 < 1$, it is expected that the FD-F will yield better finite sample power properties than the D-F test. In a realistic case in which d_1 is typically unspecified in practice, the simple alternative hypothesis in (5) can be replaced by a composite alternative

$$H_1 : 0 \leq d_0 < 1.$$

Without the value of d_1, the OLS regression in (4) is not feasible and in order to make the FD-F test operational, DGM suggest replacing the unknown parameter d_1 in (4) by a $\sqrt{T}$-consistent estimate of d_0 in (1).

The OLS estimator of ϕ for equation (4) is

$$\hat{\phi} = \frac{T^{-1} \sum_{t=2}^{T} \Delta^{d_1} y_{t-1} \Delta y_t}{T^{-1} \sum_{t=2}^{T} \left(\Delta^{d_1} y_{t-1}\right)^2}. \tag{6}$$

In DGM, the error term $\{e_t\}$ is assumed to be i.i.d. with zero mean and unknown variance σ^2. Under these assumptions, the standard error of $\hat{\phi}$ (denoted as $\mathrm{SE}\left(\hat{\phi}\right)$) is given by

$$SE\left(\hat{\phi}\right) = \frac{\hat{\sigma}}{\left(\sum_{t=2}^{T} \left(\Delta^{d_1} y_{t-1}\right)^2\right)^{1/2}}$$

where $\hat{\sigma}^2$ is the usual OLS estimator of the unknown error variance σ^2 :

$$\hat{\sigma}^2 = T^{-1} \sum_{t=2}^{T} \left(\Delta y_t - \hat{\phi} \Delta^{d_1} y_{t-1}\right)^2.$$

The t-ratio for testing $\phi = 0$ is therefore given by

$$t\left(d_1\right) = \frac{\hat{\phi}}{SE\left(\hat{\phi}\right)}. \tag{7}$$

3.1 Normality under the null hypothesis

Under the unit root null hypothesis, Theorem 1 of DGM establishes that the OLS estimate $\hat{\phi}$ is consistent, converges at the standard asymptotic rate of $\sqrt{T}$ and is asymptotically normal only if d_1 lies in the interval $1/2 < d_1 < 1$. An interesting feature of this result is that the $t\left(d_1\right)$-ratio given in (7) has a standard normal limiting null distribution. This is in contrast to the D-F unit root test which has a non standard asymptotic distribution. The intuitive reasoning behind this result is quite simple. When the $H_0 : d_0 = 1$ is true, the data generating process for $\{y_t\}$ is

$$\Delta y_t = e_t$$

and hence the standardised and centred OLS estimator in (6) becomes

$$\sqrt{T} \hat{\phi} = \frac{T^{-1/2} \sum_{t=2}^{T} z_{t-1} e_t}{T^{-1} \sum_{t=2}^{T} z_{t-1}^2} \tag{8}$$

where $z_{t-1} = \Delta^{d_1} y_{t-1}$. The sequence $\{z_{t-1}\}$ is fractionally integrated of order $1 - d_1$. To see this, we pre-multiply both sides of (1) by $\Delta^{d_1 - 1}$ to obtain

$$\begin{aligned} \Delta^{d_1} y_{t-1} &= \Delta^{d_1 - 1} e_{t-1} \\ \Delta^{1-d_1} \left(z_{t-1}\right) &= e_{t-1}. \end{aligned}$$

Thus, $\{z_{t-1}\} \sim I(1-d_1)$ as claimed.

If d_1 is restricted to lie in the interval $(1/2, 1)$, the process $\{z_t\}$ is asymptotically stationary since $0 < 1 - d_1 < 1/2$. This indicates that the process $\{z_t\}$ may possess standard asymptotic properties. Although, under the Type 2 model of fractional integration, the series $\{z_t\}$ is non-stationary for any values of d_1, Lemma 1 of DGM shows that

$$T^{-1}\sum_{t=1}^{T}\left(z_t^2 - (z_t^s)^2\right) \xrightarrow{p} 0 \tag{9}$$

where z_t^s is the non-truncated fractionally integrated process of order $1-d_1$ corresponding to z_t and is expressed as an infinite order moving average of the innovations:

$$z_t^s = z_t + \sum_{j=t}^{\infty}\pi_j(d_1-1)\, e_{t-j}.$$

The result in equation (9) states that the difference between z_t^2 and $(z_t^s)^2$, when suitably centred, disappears asymptotically. This is the key ingredient in deriving the formula for the asymptotic variance of $\hat{\phi}$. Using (9) together with the fact that z_t^s is a stationary ergodic process (see Sowell (1990)), the probability limit of the denominator in (8) can be obtained simply by appealing to the WLLN for stationary and ergodic process:

$$T^{-1}\sum_{t=2}^{T} z_{t-1}^2 \xrightarrow{p} E\left(z_{t-1}^s\right)^2. \tag{10}$$

Note that $E(z_t^s)^2$ is the variance function of z_t^s, which can be obtained from Hosking (1981) by replacing the parameter d in equation (3.2) of Hosking with $1-d_1$:

$$E(z_t^s)^2 = \frac{\sigma^2\Gamma(1-2(1-d_1))}{\Gamma^2(1-(1-d_1))} = \frac{\sigma^2\Gamma(2d_1-1)}{\Gamma^2(d_1)}. \tag{11}$$

As for the numerator in (8), Lemma 2 of DGM shows that it has the normal asymptotic distribution; that is

$$T^{-1/2}\sum_{t=2}^{T} z_{t-1}e_t \xrightarrow{d} N\left(0, E\left(z_{t-1}^s e_t\right)^2\right). \tag{12}$$

Putting together Cramer's Theorem with (10) and (12) yields the result given in DGM Theorem 1; that is under the null hypothesis,

$$\sqrt{T}\hat{\phi} \xrightarrow{d} N\left(0, AVar\left(\hat{\phi}\right)\right)$$

where $AVar\left(\hat{\phi}\right)$ denotes the asymptotic variance of $\hat{\phi}$ and is given by

$$AVar\left(\hat{\phi}\right) = \left(E\left(z_{t-1}^s\right)^2\right)^{-1} E\left(z_{t-1}^s e_t\right)^2 \left(E\left(z_{t-1}^s\right)^2\right)^{-1}. \tag{13}$$

Since the error term e_t is i.i.d., the formula for $AVar(\phi)$ can be refined further by applying the LIE to $E(z_t^s e_t)^2$ and noting that $E(e_t^2|\Im_{t-1}) = E(e_t^2) = \sigma^2$. It therefore follows that the asymptotic variance for $\hat{\phi}$, under the null hypothesis, becomes

$$\begin{aligned} AVar\left(\hat{\phi}\right) &= \left(E\left(z_{t-1}^s\right)^2\right)^{-1} E\left[E\left(\left(z_{t-1}^s\right)^2 e_t^2|\Im_{t-1}\right)\right]\left(E\left(z_{t-1}^s\right)^2\right)^{-1} \\ &= \left(E\left(z_{t-1}^s\right)^2\right)^{-1} E\left[\left(z_{t-1}^s\right)^2 E\left(e_t^2|\Im_{t-1}\right)\right]\left(E\left(z_{t-1}^s\right)^2\right)^{-1} \\ &= \sigma^2\left(E\left(z_{t-1}^s\right)^2\right)^{-1}. \end{aligned} \tag{14}$$

Substituting for $E\left(z_{t-1}^s\right)^2$ using equation (11), we obtain the required formula for $AVar\left(\hat{\phi}\right)$ as given in Lemma 2 of DGM:

$$AVar\left(\hat{\phi}\right) = \frac{\Gamma^2(d_1)}{\Gamma(2d_1 - 1)}.$$

In the presence of heteroskedasticity, the formula for the asymptotic variance given in (14) is incorrect because the third equality in (14) is obtained by assuming homoskedasticity. This will yield incorrect estimate of OLS standard errors. As a result, the FD-F t-statistic computed using this homoskedastic OLS standard errors will give misleading statistical inferences.

3.2 The effects of unconditional heteroskedasticity

In order to investigate how heteroskedasticity can affect the size properties of the FD-F test, we conduct a simple Monte Carlo study. Under the null hypothesis that $\phi = 0$ in (4), the data generating process is $(1) - (2)$ with $d_0 = 1$. The $\{\varepsilon_t\}$ are standardised normal random variables i.i.d.$N(0,1)$ and were generated using GAUSS normal random number generator *rndn*. We concentrate on the case of a single abrupt structural change in the error variance:

$$\sigma_t^2 = \gamma_1^2 \mathbf{1}(t \leq \lfloor \tau T \rfloor) + \gamma_2^2 \mathbf{1}(t > \lfloor \tau T \rfloor) \tag{15}$$

with $\tau \in (0,1)$ gives the location of the break point. In this case, the error variance shifts from γ_1^2 to γ_2^2 at time $\lfloor \tau T \rfloor$. Let $\delta = \gamma_2/\gamma_1$ be the parameter that measures the magnitude of the shift. Without loss of generality, we set $\gamma_1^2 = 1$, then, $\delta > 1$ $(\delta < 1)$ corresponds to a positive (negative) shift. The further the value of δ differs from unity, the larger the magnitude of the shift. For a positive shift, we consider $\delta = 2, 5, 10$ and for a negative shift, we consider $\delta = 0.1, 0.2, 0.5$. The values of τ are in steps of 0.1. The number of Monte Carlo replications is 10,000. We calculate the empirical size of the $t(d_1)$-statistic in (7) with $d_1 = 0.9$ for a sample size of $T = 250$. The critical value is obtained from the standard normal distribution $N(0,1)$ which is -1.645 for 5% significance level.

Figure 1 reports the real size against the nominal size at the 5% level of significance. On the horizontal axis nine break points are marked and the vertical axis gives the corresponding

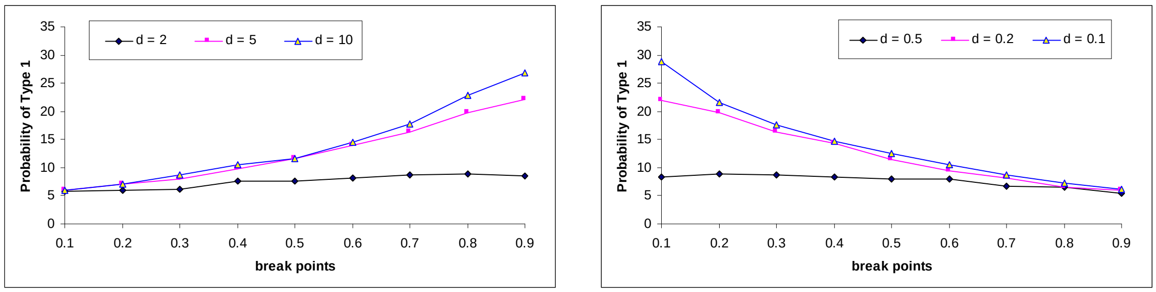

Figure 1: Empirical Sizes at the 5 percent significance level

percentage of rejections under the unit root null hypothesis. The left graph presents the case of an upward shift and the right graph presents the case of a downward shift.

DGM have shown that the empirical size of the $t(d_1)$-test, under homoskedasticity, are reasonably close to the 5% nominal level. In the upward shift case (left graph), as τ moves from 0.1 to 0.9, rejection frequency increases from about 5% to about 30%. The opposite is true for the case of a downward shift (right graph). Notice also that as the magnitude of the shift increases, the rejection frequency increases. The degree of size distortion varies with the direction, timing and magnitude of the break. Without knowledge of the timing and magnitude of the break, the t-ratio given in (7) cannot be relied upon to give valid inferences.

4 Testing for a unit root under unconditional heteroskedasticity

It is important to note that in the heteroskedastic case we cannot claim that the correct formula for $AVar\left(\hat{\phi}\right)$ is given in (13) unless we re-establish the results stated in $(9), (10)$ and (12) under Assumption V. In order to correct for the size distortion shown in the Monte Carlo study we follow White's suggestion. The White's heteroskedasticity robust t-ratio (denoted as $t_W(d_1)$) for testing the significance of ϕ in (4) is given by

$$t_W(d_1) = \frac{\hat{\phi}}{WSE\left(\hat{\phi}\right)} \tag{16}$$

where $WSE\left(\hat{\phi}\right)$ is the White standard error given by

$$WSE\left(\hat{\phi}\right) = \sqrt{\frac{\sum_{t=2}^{T}\left(\Delta^{d_1}y_{t-1}\right)^2 \widehat{e}_t^2}{\left(\sum_{t=2}^{T}\left(\Delta^{d_1}y_{t-1}\right)^2\right)^2}}.$$

The advantage of using White standard errors is that applied researchers are not required to specify, a priori, a model for the error variance process. This is especially useful since in many practical situations such model is unknown a priori. In the case of a single structural break in the error variance, the Monte Carlo evidence presented in the previous section reveals that the t-test is affected by the location of the break in the sample. The use of White standard errors to correct for heteroskedasticity is particularly useful since it does not require any information regarding the timing of the break.

4.1 Asymptotic Distribution of the FD-F Test with a fixed d_1

The following theorem is concerned with the asymptotic properties of the $t_W(d_1)$-ratio defined in (16) for a fixed d_1.

Theorem 1 *Let the time series process* $\{y_t\}_{t=1}^{T}$ *be defined by* $(1)-(2)$. *Under Assumptions E and V, the asymptotic properties of the* $t_W(d_1)$*-ratio with* $1/2 < d_1 < 1$ *for testing* $\phi = 0$ *in* (4) *are given by:*

$$t_W(d_1) \xrightarrow{d} N(0,1)$$

under the null hypothesis that $d_0 = 1$;

$$t_W(d_1) \xrightarrow{p} -\infty$$

under the alternative hypothesis that $0 \leq d_0 < 1$.

The first part of Theorem 1 states that the $t_W(d_1)$-ratio defined in (16) is pivotal and has a standard limiting null distribution. The second part states that the $t_W(d_1)$-ratio is able to discriminate between the null hypothesis of a random walk process and the alternative hypothesis of a fractional integrated process with probability one in large samples. Thus for a given value of $d_0 \in [0,1)$, the $t_W(d_1)$-ratio, computed using any value of d_1 in the interval $1/2 < d_1 < 1$, tends to negative infinity and so the power of the $t_W(d_1)$-ratio tends to unity as $T \to \infty$. This means that the $t_W(d_1)$-ratio is a consistent test statistic even if d_1 is chosen differently from d_0.

Establishing Theorem 1 is more difficult than in the homoskedastic case where the ergodic stationary WLLN is a key ingredient. This is because unconditional heteroskedasticity renders invalid the application of ergodic stationary WLLN for the various sample moments appearing in $\hat{\phi}$ and $t_W(d_1)$. The proof of Theorem 1 is given in the Appendix.

4.2 Asymptotic Distribution of the FD-F Test with an Estimated d_1

So far the value of d_1 is specified a priori under the simple alternative hypothesis. In the absence of such a specification, as is usually the case, Theorem 2 shows that any available consistent estimator of d_0 in (1) can be used to make the FD-F operational. Let $\hat{d}$ be a consistent estimator such that $\hat{d} - d_0 = o_p(1)$. Since FD-F regression requires the value of d_1 to be strictly less than one, we follow DGM's suggession and define a trimming rule for $\hat{d}_1$:

$$\hat{d}_1 = \begin{cases} \hat{d} & \text{if } \hat{d} < 1-c \\ 1-c & \text{if } \hat{d} \geq 1-c \end{cases} \tag{17}$$

where c is a fixed constant such that $0 < c < 1/2$. The least squares regression in (4) can now be made operational by replacing the unspecified d_1 by $\hat{d}_1$:

$$\Delta y_t = \phi \Delta^{\hat{d}_1} y_{t-1} + e_t. \tag{18}$$

The heteroskedasticity robust t-statistic, computed in the same way as before except using the regression (18), is now denoted as $t_W\left(\hat{d}_1\right)$. The following Theorem establishes that the $t_W\left(\hat{d}_1\right)$-ratio has a standard normal limiting null distribution under heteroskedastic errors. The proof of the Theorem is given in the Appendix.

Unlike DGM, we do not require the estimator $\hat{d}$ to converge in probability to d_0 at a $\sqrt{T}$-rate. Instead of relying only on parametric estimation, which has a standard $\sqrt{T}$-rate of convergence, Theorem 2 opens up the possibility that semi-parametric estimation, which converges slower than the $\sqrt{T}$-rate, can be used to make the FD-F test operational.

Theorem 2 *Let $\hat{d}_1$ satisfy the trimming rule in* (17) *with $\hat{d} - d_0 = o_p(1)$. Suppose Assumptions E and V hold. Under the null hypothesis that y_t is generated by* (1) − (2) *with $d_0 = 1$, the asymptotic distribution of the $t_W\left(\hat{d}_1\right)$-ratio of the OLS estimator of ϕ in the regression*

$$\Delta y_t = \phi \Delta^{\hat{d}_1} y_{t-1} + e_t$$

is given by

$$t_W\left(\hat{d}_1\right) \xrightarrow{d} N(0,1).$$

We follow the DGM approach and consider the parametric GMD estimation method. Harris and Kew (2007) examine the GMD estimation in the presence of unconditional heteroskedasticity in a fractional integrated model. The GMD estimation procedure relies only on the absence of autocorrelations and not of heteroskedasticity in the residuals and hence this estimator can be potentially robust to the presence of heteroskedasticity. We next describe the GMD estimator.

4.3 GMD estimation method

This section describes the GMD estimation procedure. To describe this estimator, we define the residuals $e_t(d)$ for some d as

$$e_t(d) = \Delta^d y_t = \sum_{j=0}^{t-1} \pi_j(d) y_{t-j}, \qquad t = 1, ..., T. \tag{19}$$

The Minimum Distance Estimation (MDE) of d_0 is defined as

$$\hat{d} = \arg\min_{d \in [0,1]} \sum_{m=1}^{k} \hat{\rho}_m^2(d)$$

where $\hat{\rho}_m(d)$ is the m'th sample autocorrelation of the residuals $e_t(d)$, for $m = 1, 2, ..., k$. The notation arg min denotes the value of d such that the argument of $\sum_{m=1}^{k} \hat{\rho}_m^2(d)$ is minimised Given the residuals $e_1(d), ..., e_T(d)$, the $\hat{\rho}_m(d)$ is a function of the parameter d and it can be calculated as follows:

$$\hat{\rho}_m(d) = \frac{T^{-1} \sum_{t=m+1}^{T} e_t(d)\, e_{t-m}(d)}{T^{-1} \sum_{t=1}^{T} e_t^2(d)}.$$

Note that the residual in (19) can be re-written as

$$\begin{aligned} e_t(d) &= \Delta^{d-d_0+d_0} y_t \\ &= \Delta^{d-d_0} \Delta^{d_0} y_t \\ &= \Delta^{d-d_0} e_t \end{aligned}$$

and the population counterpart of $\hat{\rho}_m(d)$ is defined to be

$$\rho_m(d) = \frac{T^{-1} \sum_{t=m+1}^{T} E\left(e_t(d)\, e_{t-m}(d)\right)}{T^{-1} \sum_{t=1}^{T} E\left(e_t^2(d)\right)}. \tag{20}$$

Despite the fact that $e_t(d_0)$ are unconditionally heteroskedastic, the population autocorrelations, $\rho_m(d)$, evaluated at the true parameter d_0 can still be zero since

$$\rho_m(d_0) = \frac{T^{-1} \sum_{t=m+1+p}^{T} E\left(e_t(d_0)\, e_{t-m}(d_0)\right)}{T^{-1} \sum_{t=1+p}^{T} E\left(e_t^2(d_0)\right)} = \frac{T^{-1} \sum_{t=m+1+p}^{T} E\left(e_t e_{t-m}\right)}{T^{-1} \sum_{t=1+p}^{T} E\left(e_t^2\right)} = 0.$$

In order for the GMD method to be a reliable estimator under unconditional heteroskedasticity, the population autocorrelation $\rho_m(d)$ should not equal zero if $d \neq d_0$ and therefore the residuals, $e_t(d)$, are autocorrelated. To verify this, we write the numerator of $\rho_m(d)$ in (20) as

$$\begin{aligned} E\left(e_t(d)\, e_{t-m}(d)\right) &= E\left(\Delta^{d-d_0} e_t \Delta^{d-d_0} e_{t-m}\right) \\ &= E\left(\sum_{j=0}^{t-1} \pi_j(d-d_0)\, e_{t-j} \sum_{i=0}^{t-m-1} \pi_i(d-d_0)\, e_{t-m-i}\right) \\ &= \sum_{j=0}^{t-1} \sum_{i=m}^{t-1} \pi_j(d-d_0)\, \pi_{i-m}(d-d_0)\, E\left(e_{t-j} e_{t-i}\right) \\ &= \sum_{i=m}^{t-1} \pi_{i-m}^2(d-d_0)\, \sigma_{t-i}^2 \end{aligned}$$

and so $\rho_m(d)$ is not equal zero when $d \neq d_0$.

It is clear that the GMD estimator can be potentially robust to an unknown form of unconditional heteroskedasticity. The assumption of constant unconditional variances over time is unnecessarily restrictive for the consistency of the GMD estimator. Harris and Kew (2007) show that the GMD estimator is consistent and converges at $\sqrt{T}$-rate under Assumption V. This result implies that the GMD estimator turns out to be very useful in implementing a feasible FD-F statistic under heteroskedasticity.

5 Monte Carlo Simulation Results

This section uses Monte Carlo (MC) experiments to examine the finite sample performance of the FD-F tests with White standard errors when the errors are unconditionally heteroskedastic. The simulated data set for $\{y_t\}$ is generated according to (1) - (2). The pseudo random numbers for ε_t are generated using the *rndn* function in Gauss 7. For all of the MC experiments, the number of replications is 10000 and the seed used for the *rndn* function is 999. The sample sizes considered are $T = 250, 500$ and 1000. The larger values of T are chosen since empirical studies of structural breaks in the error variance use data collected over an extended period of time.

Apart from the single structural break model considered in Section 3.2, we consider two additional models for the error variance process, σ_t^2. These heteroskedasticity models are based on that used by Cavaliere (2004), Cavaliere and Taylor (2007) and Phillips and Xu (2006).

Double Variance Shifts. Two abrupt shifts in the error variance at first from γ_1^2 to γ_2^2 occurring at time $\lfloor \tau_1 T \rfloor$ and then follow by another abrupt shift from γ_2^2 to γ_3^2 at time $\lfloor \tau_2 T \rfloor$. The dynamics of σ_t^2 can be written as

$$\sigma_t^2 = \gamma_1^2 \mathbf{1} \left(t \leq \lfloor \tau_1 T \rfloor \right) + \gamma_2^2 \mathbf{1} \left(\lfloor \tau_1 T \rfloor < t \leq \lfloor \tau_2 T \rfloor \right) + \gamma_3^2 \mathbf{1} \left(\lfloor \tau_2 T \rfloor < t \leq T \right)$$

where $\tau_1, \tau_2 \in (0, 1)$. A special case arises when $\tau_2 = 1 - \tau_1$ and $\gamma_3^2 = \gamma_1^2$. In this case, the multiple variance shifts are symmetric and and hence the double variance shifts model reduces to

$$\sigma_t^2 = \gamma_1^2 \mathbf{1} \left(t \leq \lfloor \tau T \rfloor \right) + \gamma_2^2 \mathbf{1} \left(\lfloor \tau T \rfloor < t \leq 1 - \lfloor \tau T \rfloor \right) + \gamma_1^2 \mathbf{1} \left(1 - \lfloor \tau T \rfloor < t \leq T \right) \qquad (21)$$

where $\tau \in (0, 1)$.

Trending Variances. The variance of the innovations changes monotonically from γ_1^2 at time $t = 0$ to γ_2^2 at time $t = T$. Note that, the variance may not necessarily trends linearly. The dynamics of σ_t^2 can be written as

$$\sigma_t^2 = \gamma_1^2 + \left(\gamma_2^2 - \gamma_1^2 \right) \left(\frac{t}{T} \right)^m$$

where $m = 1, 2, .. < \infty$. The variance changes continuously in a linear fashion when $m = 1$ and it changes in a non-linear way otherwise.

For each of the variance models, a wide range of parameter settings are used to generate the different patterns of variance dynamics. Following Section 3.2, we define $\delta = \gamma_2/\gamma_1$ and normalise $\gamma_1 = 1$.

As for the single structural break model in (15), we allow the shift to occur towards the beginning, middle and end of the sample by setting $\tau = 0.1, 0.5$ and 0.9. Two values of δ were used: $\delta = 5$ (positive shift) and $\delta = 0.2$ (negative shift). We consider early positive shift ($\tau = 0.1$ and $\delta = 5$), early negative shift ($\tau = 0.1$ and $\delta = 0.2$), positive shift occurring mid-way through the sample ($\tau = 0.5$ and $\delta = 5$), negative shift occurring mid-way through the sample ($\tau = 0.5$ and $\delta = 0.2$), late positive shift ($\tau = 0.9$ and $\delta = 5$), and late negative shift ($\tau = 0.9$ and $\delta = 0.2$).

As for the double variance shifts model, we consider the special case where the double breaks occur symmetrically. We let $\tau = 0.05, 0.45$ and $\delta = 5, 0.2$. Here, four different types of multiple break points are generated: early positive break then followed by late negative break ($\tau = 0.05$ and $\delta = 5$); early negative break then followed by late positive break ($\tau = 0.05$ and $\delta = 0.2$); positive shift occurring near the middle of the sample then immediately followed by a negative shift ($\tau = 0.45$ and $\delta = 5$) and a negative shift occurring near the middle of the sample then immediately followed by a positive shift ($\tau = 0.45$ and $\delta = 0.2$).

In the trending variances model, we allow the trending variance to increase continuously in a linear and non-linear fashion. In the linear case where $m = 1$, we consider both upward ($\delta = 5$) and downward ($\delta = 0.2$) trends. In the non-linear case ($m = 2$), we consider both upward and downward trending variances.

5.1 Size Properties with known d_1

Tables 1 to 3 report the percentage of rejections under the null hypothesis (empirical size) when $d_0 = 1$ in (1) for the $t(d_1)$- and $t_W(d_1)$-tests with $d_1 = 0.6, 0.7, 0.8, 0.9, 0.95$ at the nominal 5% level. Since both tests have standard normal limiting distributions, the critical value, at the 5% significance level, is -1.645. All other things equal, the $t(d_1)$ and $t_W(d_1)$ tests display roughly the same rejection frequencies for all values of d_1, although it is worth nothing that size distortions are slightly smaller the closer d_1 is to 1.

For comparison purposes, we include the actual sizes of the $t(d_1)$ test under homoskedastic error and these are reported in the first row of Table 1. In terms of empirical size, the $t_W(d_1)$ test performs just as well as the $t(d_1)$ test in the absence of heteroskedasticity.

We will discuss first the performance of the $t(d_1)$ test in the presence of unconditional heteroskedasticity. Tables 1 to 3 clearly show that the $t(d_1)$ tests are not robust to departures from the homoskedasticity assumption. The degree of size distortions can vary, depending on the variance structure. For example, in the case of a single abrupt shift (see Table 1), substantial size distortion occurs when the abrupt shift is either an early negative shift or a late positive shift, while the opposite is true when the abrupt shift is either an early positive shift or a late negative shift. When the positive or negative shift occurs towards the middle of the sample, the $t(d_1)$ tests overreject moderately with empirical sizes of around 11%.

As for the double variance shifts model (see Table 2), serious size distortions arise when an early negative shift is followed by a late positive shift of the same magnitude as the earlier shift ($\tau = 0.05$, $\delta = 0.2$), and when a positive shift occurring towards the middle of the sample is followed immediately by a negative shift of the same magnitude ($\tau = 0.45$, $\delta = 5$). In these cases, the proportion of rejections is about 25% when the nominal size is 5%. However, the $t(d_1)$ tests appear to have approximately correct size when an early positive shift is followed by a late negative shift ($\tau = 0.05, \delta = 5$), and when a negative shift occurring towards the middle of the sample is followed immediately by a positive shift ($\tau = 0.45$, $\delta = 0.2$).

When the error variance follows a polynomial trend (see Table 3), the $t(d_1)$ display smaller size distortions than those in Tables 1 and 2. In this case, the $t(d_1)$ tests moderately overreject the unit root null hypothesis with empirical sizes varying from 7 to 10 per cent.

Tables 1 − 3 about here

The $t_W(d_1)$ test which uses White standard errors is very effective in reducing the observed size distortions. Take for example the case where there is an early negative break (i.e. $\tau = 0.1$ and $\delta = 0.2$). In this case, when $d_1 = 0.90$ and $T = 250$, Table 1 shows that the White correction can reduce the empirical size from 23.14% to 6.98%. This 6.98% empirical size of the $t_W(d_1)$ test is seen, as expected, to fall towards the 5% nominal size as T grows. When the sample size is relatively large ($T = 1000$), empirical sizes of $t_W(d_1)$ tests are always reasonably close to the 5% nominal level in all of the heteroskedastic models considered. Thus, Monte Carlo evidence reveals that White's correction works well in practice for a wide range of models of unconditional heteroskedasticity.

Tables 4 to 7 report the raw power of the $t_W(d_1)$ test against the alternative of fractionally integrated processes given in (1) with values of d_0 chosen from $[0.55, 0.95]$ in steps of 0.05. The $t_W(d_1)$ test is computed under the assumption that d_1 is known a priori by setting $d_1 = d_0$. Under homoskedasticity (i.e. $\delta = 1$), Table 4 shows that the rejection frequencies of the heteroskedastic standard errors are comparable to those of the homoskedastic standard errors. This suggests there is no loss in power from using White's correction when the errors are homoskedastic. Table 4 therefore can be used as a benchmark to compare the finite sample power results of the $t_W(d_1)$ under heteroskedasticity. From Tables 5 to 7, there are cases where the power of the $t_W(d_1)$ tests under heteroskedasticity are considerably lower than those under homoskedasticity. In those cases, it turns out that the FD-F tests without White standard errors suffer from severe size distortions. To illustrate, take an example of an early negative break ($\tau = 0.1, \delta = 0.2$). As noted before, the $t(d_1)$ tests tends to overreject substantially. For this same variance model, Table 5 shows that the power of the $t_W(d_1)$ test is considerably less than that observed in Table 4, where the errors are homoskedastic. In some cases where the $t(d_1)$ tests suffer from severe size distortions, the White correction loses power relative to the homoskedastic case. In all cases, as expected, power increases as T increases, and as d_0 moves away from 1.

Tables 4 − 7 about here

So far we have been treating the value of d_1 as if it is pre-specified. Now we use the GMD estimation procedure described earlier to estimate d_0. Using the trimming rule defined in (17), the resulting estimate is then used to replace d_1 with $\hat{d}_1$. Under homoskedasticity, DGM show via Monte Carlo experiments that replacing the value of d_1 with $\hat{d}_1$ has very little impact on the size and power properties of the FD-F test without White's correction. We now check whether these results continue to hold under heteroskedasticity. Tables 9 to 11 present the size and power properties of the FD-F tests with and without White's correction. The $t\left(\hat{d}_1\right)$ and $t_W\left(\hat{d}_1\right)$ tests are calculated in the same way as before, but using the input $\hat{d}_1$ as defined in (17). Following DGM, we set $c = 0.02$ and thus $\hat{d}_1 \leq 0.98$. By comparing the results in Tables 8 - 11 with the corresponding results in Tables 1-7, the rejection frequencies when d_1 is replaced with $\hat{d}_1$ are broadly similar to those when d_1 is specified a priori. Thus, the estimation of d_1 using the GMD estimator under unconditional heteroskedasticity will not affect the performance of the $t_W\left(\hat{d}_1\right)$ tests.

Tables 9 − 11 about here

6 Conclusion

We have shown that, via Monte Carlo simulations, the OLS t-statistics suffer from substantial size distortions when the errors are heteroskedastic. Thus, the FD-F t-statistics computed using OLS standard errors derived under the assumption of homoskedasticity will give misleading statistical inferences. We suggest FD-F t-statistic that uses White heteroskedasticity robust standard errors to account for the presence of unconditional heteroskedasticity of unknown form. We demonstrate that White's version of the FD-F statistics has a standard limiting null distribution unaffected by unconditional heteroskedasticity. A Monte Carlo study shows that the proposed method is effective in reducing the size distortion. In the absence of heteroskedasticity, the power loss due to the use of White standard errors instead of homoskedastic standard errors turns out to be very small.

References

[1] Anderson, T.W. (1971) The statistical analysis of time series. *Wiley New York.*

[2] Beare, B. (2004) Robustifying unit root test to permanent changes in innovation variance. Yale University, mimeographed.

[3] Beran, J. (1995) Maximum likelihood estimation of the differencing parameter for invertible and short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society,* 57, 659-672.

[4] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

[5] Busetti, F. and A.C. Harvey (2001) Testing for the presence of a random walk in series with structural breaks. *Journal of Time Series Analysis*, 22, 127-150.

[6] Busetti, F. and A.M.R. Taylor (2003) Variance shifts, structural breaks, and stationarity tests. *Journal of Business and Economic Statistics,* 21, 510-531.

[7] Busetti, F. and A.M.R. Taylor (2004) Tests of stationarity against a change in persistence. *Journal of Econometrics*, 123, 33-66.

[8] Cavaliere, G. (2004) Unit root tests under time-varying variances. *Econometric Reviews*, 23, 259-292.

[9] Cavaliere, G. and A.M.R. Taylor (2005) Stationarity tests under time-varying second moments. *Econometric Theory,* 21, 1112-1129.

[10] Cavaliere, G. and A.M.R. Taylor (2006) Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics*, forthcoming.

[11] Cavaliere, G. and A.M.R. Taylor (2006) Testing for a change in persistence in the presence of a volatility shift. Oxford Bulletin of Economics and Statistics, 68, 761-781.

[12] Davidson, J. (1994) Stochastic limit theory: an introduction for econometricians. *Oxford University Press.*

[13] Dolado, J., Gonzalo, J. and L. Mayoral (2002) A fractional Dickey-Fuller test for unit roots. *Econometrica*, 70, 1963-2006.

[14] Engle, R. F. (1982) Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. *Econometrica*, 50, 987-1008.

[15] Hamori, S. and A. Tokihisa (1997) Testing for a unit root in the presence of a variance shift. *Economics Letters,* 57, 245-253.

[16] Harris D. and H. Kew (2007) Root-T consistent estimator of fractionally integrated model with unconditional heteroskedastic errors. Melbourne University, mimeographed.

[17] Hosking, J.R.M. (1981) Fractional differencing, *Biometrica,* 68, 165-176.

[18] Kim, J. (2000) Detection of change in persistence of a linear time series. *Journal of Econometrics*, 95, 97-116.

[19] Kim, J., J. Belaire Franch and R. Badillo Amador (2002) Corrigendum to "Detection of change in persistence of a linear time series". *Journal of Econometrics*, 109, 389-392.

[20] Kim, T.H., S. Leybourne and P. Newbold (2002) Unit root tests with a break in innovation variance. *Journal of Econometrics,* 109, 365-387.

[21] Mayoral M. (2007) Minimum distance estimation of stationary and non-stationary ARFIMA processes. *Econometrics Journal,* 10, 124-148.

[22] McConnell, M.M. and G. Perez-Quiros (2000) Output fluctuations in the United States: what has changed since the early 1980s? *American Economic Review,* 90, 1464-1476.

[23] Nicholls, D.F. and A.R. Pagan (1983) Heteroskedasticity in models with lagged dependent variables. *Econometrica,* 51, 1233-1242.

[24] Perron, P. (1989) The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57, 1361-1401.

[25] Perron, P. (1990) Testing for a unit root in a time series with a changing mean. *Journal of Business and Economic Statistics,* 8, 153-162.

[26] Phillips, P.C.B. and K.-L. Xu (2006) Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis,* 27, 289-308.

[27] Robinson, P.M. and D. Marinucci (2001) Narrow-band analysis of nonstationary processes. *Annals of Statistics*, 29, 947-986.

[28] Sowell, F.B. (1990) The fractional unit root distribution. *Econometrica,* 58, 495-505.

[29] Stock, J.H., (1999) A class of tests for integration and cointegration. In: Engle, R.F., White, H. (Eds.), Cointegration, causality and forecasting. A Festschrift in honour of Clive W.J. Granger. *Oxford Univeristy Press*, Oxford, pp. 137-167.

[30] Tanaka, K. (1999) The nonstationary fractional unit root. *Econometric Theory,* 15, 549-582.

[31] van Dijk, D., D.R. Osborn and M. Sensier (2002) Changes in variability of the business cycle in the G7 countries. Econometric Institute Report EI 2002-28, Eramus University Rotterdam.

[32] White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica,* 48, 817-838.

[33] White, H. (1984) Asymptotic theory for econometricians. *Academic Press.*

Appendix: Proofs

Lemmas 3 and 4 below give the weak law of large numbers and the Central Limit Theorem for fractional integrated processes under Assumptions V and E, respectively.

Lemma 3 *Let $\{e_t\}$ be a sequence of random variables generated according to* (2) *with σ_t and $\{\varepsilon_t\}$ satisfy Assumptions V and E respectively. Consider the following fractionally integrated processes:*

$$\Delta^{\delta} z_t = e_t, \qquad \delta < 1/2$$

and

$$\Delta^{\gamma} x_t = e_t, \qquad \gamma < 1/2.$$

Then
(i) $\sup_{1 \leq t \leq T} E\left|z_t\right|^r < \infty$ *for* $r \leq 4$*;*
(ii) $0 < \underline{B} \leq T^{-1} \sum_{t=2}^{T} E\left(z_{t-1}^2\right)$*; and*
(iii) $T^{-1} \sum_{t=2}^{T} \left(x_{t-m} z_{t-n} - E\left(x_{t-m} z_{t-n}\right)\right) \xrightarrow{p} 0$ *for* $m, n = 0, 1.$

It seems useful to state the following results here. The stochastic sequence $T^{-1} \sum_{t=2}^{T} z_{t-1}^2$ does not converge to zero in probability. To see this, we write

$$T^{-1} \sum_{t=2}^{T} z_{t-1}^2 = T^{-1} \sum_{t=2}^{T} \left(z_{t-1}^2 - E\left(z_{t-1}^2\right)\right) + T^{-1} \sum_{t=2}^{T} E\left(z_{t-1}^2\right). \tag{22}$$

The first term on the right-hand side of the above equation converges in probability to zero by Lemma 3(iii) with $x_{t-1} = z_{t-1}$. The second term is bounded away from zero uniformly in T by Lemma 3(ii).

Lemma 4 *Let $\{z_t\}$ and $\{e_t\}$ be defined as in Lemma 3. If the conditions of Lemma 3 are satisfied, then as $T \to \infty$,*

$$\left(T^{-1} \sum_{t=2}^{T} E\left(z_{t-1}^2 e_t^2\right)\right)^{-1/2} T^{-1/2} \sum_{t=2}^{T} z_{t-1} e_t \xrightarrow{d} N\left(0, 1\right).$$

Proof of Lemma 3

Part (i) Under the Type 2 model of fractional integration, the series z_t can be written as

$$z_t = \Delta^{-\delta} e_t 1_{(t>0)} = \sum_{j=0}^{t-1} \pi_j\left(-\delta\right) e_{t-j} \tag{23}$$

and since $\delta < 1/2$, the coefficient $\pi_j(-\delta)$ is square summable. Choosing $r = 4$, we have

$$\begin{aligned} & E\left(z_t^4\right) \\ = \ & E\left(\sum_{j=0}^{t-1} \pi_j(-\delta) e_{t-j}\right)^4 \\ = \ & \sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\sum_{k=0}^{t-1}\sum_{l=0}^{t-1} \pi_i(-\delta)\pi_j(-\delta)\pi_k(-\delta)\pi_l(-\delta) E\left(e_{t-i}e_{t-j}e_{t-k}e_{t-l}\right). \end{aligned} \tag{24}$$

First observe that

$$\begin{aligned} & E\left(e_{t-i}e_{t-j}e_{t-k}e_{t-l}\right) \\ = \ & E\left(e_{t-i}e_{t-j}\right)E\left(e_{t-k}e_{t-l}\right) + E\left(e_{t-i}e_{t-k}\right)E\left(e_{t-j}e_{t-l}\right) \\ & + E\left(e_{t-i}e_{t-l}\right)E\left(e_{t-j}e_{t-k}\right) + \kappa_t\left(t-i, t-j, t-k, t-l\right) \end{aligned} \tag{25}$$

where $\kappa_t(.,.,.,.)$ denotes the joint fourth order cumulants of e_t. It follows that equation (24) can be rewritten as

$$\begin{aligned} & E\left(z_t^4\right) \\ = \ & 3\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\sum_{k=0}^{t-1}\sum_{l=0}^{t-1} \pi_i(-\delta)\pi_j(-\delta)\pi_k(-\delta)\pi_l(-\delta) E\left(e_{t-i}e_{t-j}\right)E\left(e_{t-k}e_{t-l}\right) \\ & + \sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\sum_{k=0}^{t-1}\sum_{l=0}^{t-1} \pi_i(-\delta)\pi_j(-\delta)\pi_k(-\delta)\pi_l(-\delta)\kappa_t\left(t-i, t-j, t-k, t-l\right). \end{aligned} \tag{26}$$

We will show that each term on the right hand side of equation (26) is uniformly bounded in $1 \le t \le T$.

Since $\{e_t\}$ are independent and $E(e_t) = 0$, non-zero expectations arise in the following three pairs: (i) $i = j$ and $k = l$; (ii) $i = k$ and $j = l$; (iii) $i = l$ and $j = k$. The first term on the right side of equation (26) is uniformly bounded in $1 \le t \le T$ since

$$\begin{aligned} & 3\sum_{i=0}^{t-1}\sum_{k=0}^{t-1} \pi_i^2(-\delta)\pi_k^2(-\delta) E\left(e_{t-i}^2\right)E\left(e_{t-k}^2\right) \\ = \ & 3\sum_{i=0}^{t-1}\sum_{k=0}^{t-1} \pi_i^2(-\delta)\pi_k^2(-\delta)\sigma_{t-i}^2\sigma_{t-k}^2 \\ \le \ & 3\sum_{i=0}^{t-1}\sum_{k=0}^{t-1} \pi_i^2(-\delta)\pi_k^2(-\delta)\bar{\sigma}^4 \\ \le \ & 3\bar{\sigma}^4 \sum_{i=0}^{\infty} \pi_i^2(-\delta) \sum_{k=0}^{\infty} \pi_k^2(-\delta) < \infty. \end{aligned} \tag{27}$$

Regarding the second term on the right hand side of equation (26), the independence assumption for e_t implies that the fourth order cumulants are zero except when $i = j =$

$k = l$. To see this, consider the case where $i = j \neq k = l$. Then, equation (25) becomes

$$\begin{aligned} \kappa_t (t-i, t-i, t-k, t-k) &= E\left(e_{t-i}^2 e_{t-k}^2\right) - E\left(e_{t-i}^2\right) E\left(e_{t-k}^2\right) \\ &= E\left(e_{t-i}^2 e_{t-k}^2\right) - E\left(e_{t-i}^2\right) E\left(e_{t-k}^2\right) \\ &= 0. \end{aligned}$$

Similar arguments follow for cases where $i = k \neq j = l$ and $i = l \neq j = k$. However, if $i = j = k = l$, equation (25) becomes

$$\kappa (t-i, t-i, t-i, t-i) = E\left(e_{t-i}^4\right) - 3E\left(e_{t-i}^2\right)^2.$$

In this case, the fourth order cumulants are zero only when the process $\{e_t\}$ is Gaussian because $E\left(e_{t-i}^4\right) = 3\sigma_{t-i}^4$. Therefore, the second term in (26) is uniformly bounded in $1 + m \leq t \leq T$ under Assumption V since

$$\begin{aligned} & \sum_{i=0}^{t-1} \pi_i^4 (-\delta) \left|\kappa_t (t-i, t-i, t-i, t-i)\right| \\ = \; & \sum_{i=0}^{t-1} \pi_i^4 (-\delta) \left| E\left(e_{t-i}^4\right) - 3E\left(e_{t-i}^2\right)^2 \right| \\ \leq \; & \sum_{i=0}^{t-1} \pi_i^4 (-\delta) \left(\left| E\left(e_{t-i}^4\right)\right| + \left|3\sigma_{t-i}^4\right| \right) \\ \leq \; & \sum_{i=0}^{t-1} \pi_i^4 (-\delta) \left(\left| \sup_t E\left(e_t^4\right)\right| + \left|3\bar{\sigma}^4\right| \right) \\ \leq \; & \left(\left| \sup_t E\left(e_t^4\right)\right| + \left|3\bar{\sigma}^4\right| \right) \sum_{i=0}^{\infty} \pi_i^4 (-\delta) < \infty. \end{aligned} \tag{28}$$

Combining equations (27) and (28) yields

$$\sup_{1 \leq t \leq T} E\left(z_t^4\right) \leq B < \infty.$$

Then (i) follows directly from the Liapunov inequality.

Part (ii) To show (22), we write

$$\begin{aligned} T^{-1} \sum_{t=2}^{T} E\left(z_{t-1}^2\right) &= T^{-1} \sum_{t=2}^{T} E\left(\Delta^{-\delta} e_{t-1}\right)^2 \\ &= T^{-1} \sum_{t=2}^{T} \sum_{j=1}^{t-1} \pi_{j-1}^2 (-\delta) \, \sigma_{t-j}^2 \\ &\geq \underline{\sigma}^2 T^{-1} \sum_{t=2}^{T} \sum_{j=1}^{t-1} \pi_{j-1}^2 (-\delta). \end{aligned}$$

Since $\underline{\sigma}^2 > 0$, it is only required to show that

$$0 < \underline{B} \leq \lim_{T\to\infty} T^{-1} \sum_{t=2}^{T} \sum_{j=1}^{t-1} \pi_{j-1}^2 \left(-\delta\right) < \infty.$$

In order to show this, we write

$$\begin{aligned} & T^{-1} \sum_{t=2}^{T} \sum_{j=1}^{t-1} \pi_{j-1}^2 \left(-\delta\right) \\ = \quad & T^{-1} \sum_{t=2}^{T} \sum_{j=1}^{T-1} \pi_{j-1}^2 \left(-\delta\right) - T^{-1} \sum_{t=2}^{T-1} \sum_{j=t}^{T-1} \pi_{j-1}^2 \left(-\delta\right) \\ = \quad & \left(1 - \frac{1}{T}\right) \sum_{j=1}^{T-1} \pi_{j-1}^2 \left(-\delta\right) - T^{-1} \sum_{t=2}^{T} \sum_{j=t}^{T-1} \pi_{j-1}^2 \left(-\delta\right) \\ = \quad & \sum_{j=1}^{T-1} \pi_{j-1}^2 \left(-\delta\right) - \frac{1}{T} \sum_{j=1}^{T-1} \pi_{j-1}^2 \left(-\delta\right) - T^{-1} \sum_{t=2}^{T} \sum_{j=t}^{T-1} \pi_{j-1}^2 \left(-\delta\right). \end{aligned}$$

As $T \to \infty$, the first term approaches a finite *positive* limiting value since the sequence $\pi_j\left(-\delta\right)$ is square summable, that is

$$0 < \underline{B} \leq \lim_{T\to\infty} \sum_{j=1}^{T-1} \pi_{j-1}^2 \left(-\delta\right) < \infty.$$

This thus implies that the second term converges to zero. The third term converges to zero by Cesaro summation. This completes the proof for part (iii).

Part (iii) We show that the sequence $\{x_t z_{t-1} - E\left(x_t z_{t-1}\right)\}$ satisfies the conditions of the Chebyshev Law of Large Numbers (see Davidson 2000 pg 42). The first condition, which requires the sequence has zero mean, is satisfied trivially. The second condition requires that

$$\lim_{T\to\infty} E\left(T^{-1} \sum_{t=2}^{T} \left(x_t z_{t-1} - E\left(x_t z_{t-1}\right)\right)\right)^2 = 0. \tag{29}$$

Like the series z_t in (23), x_t can be written as

$$x_t = \Delta^{-\gamma} e_t 1_{(t>0)} = \sum_{j=0}^{t-1} \pi_j \left(-\gamma\right) e_{t-j}. \tag{30}$$

In order to simplify notation, we will rewrite the coefficient $\pi_i\left(-\gamma\right)$ in (30) as α_i and the coefficient $\pi_j\left(-\delta\right)$ in (23) as β_i. Under the condition of Lemma 3 ($\gamma < 1/2$ and $\delta < 1/2$), the coefficients α_i and β_i are square summable.

To show (29), we write

$$
\begin{aligned}
&\left| E\left(T^{-1}\sum_{t=2}^{T}\left(x_t z_{t-1} - E\left(x_t z_{t-1}\right)\right)\right)^2 \right| \\
\leq\ & 2T^{-2}\sum_{t=2}^{T}\sum_{s=0}^{t-2}\left| E\left(\left(x_t z_{t-1} - E\left(x_t z_{t-1}\right)\right)\left(x_{t-s} z_{t-s-1} - E\left(x_{t-s} z_{t-s-1}\right)\right)\right)\right| \\
\leq\ & 2T^{-2}\sum_{t=2}^{T}\sum_{s=0}^{t-2}\left| E \begin{array}{c} \left(\sum_{i=0}^{t-1}\sum_{j=0}^{t-2}\alpha_i\beta_j\left(e_{t-i}e_{t-1-j} - E\left(e_{t-i}e_{t-1-j}\right)\right)\right) \\ \left(\sum_{k=0}^{t-s-1}\sum_{l=0}^{t-s-2}\alpha_k\beta_l\left(e_{t-s-k}e_{t-s-1-l} - E\left(e_{t-s-k}e_{t-s-1-l}\right)\right)\right) \end{array} \right| \\
=\ & 2T^{-2}\sum_{t=2}^{T}\sum_{s=0}^{t-2}\left| E \begin{array}{c} \left(\sum_{i=0}^{t-1}\sum_{j=1}^{t-1}\alpha_i\beta_{j-1}\left(e_{t-i}e_{t-j} - E\left(e_{t-i}e_{t-j}\right)\right)\right) \\ \left(\sum_{k=s}^{t-1}\sum_{l=s+1}^{t-1}\alpha_{k-s}\beta_{l-s-1}\left(e_{t-k}e_{t-l} - E\left(e_{t-k}e_{t-l}\right)\right)\right) \end{array} \right|. \qquad (31)
\end{aligned}
$$

Because of independence property of e_t, the following three pairs have non-zero expected values: (a) $i = j$ and $k = l$; (b) $i = k$ and $j = l$; and (c) $i = l$ and $j = k$. In each of these cases, we will show that the term converges to zero as $T \to \infty$.

Case (a). The term in equation (31) for which $i = j$ and $k = l$ simplifies as follows:

$$
\begin{aligned}
& 2T^{-2}\sum_{t=2}^{T}\sum_{s=0}^{t-2}\left|\sum_{j=1}^{t-1}\sum_{l=s+1}^{t-1}\alpha_j\beta_{j-1}\alpha_{l-s}\beta_{l-s-1}E\left(e_{t-j}^2 - E\left(e_{t-j}^2\right)\right)\left(e_{t-l}^2 - E\left(e_{t-l}^2\right)\right)\right| \\
\leq\ & 2T^{-2}\sum_{t=2}^{T}\sum_{s=0}^{t-2}\sum_{l=s+1}^{t-1}\left|\alpha_l\right|\left|\beta_{l-1}\right|\left|\alpha_{l-s}\right|\left|\beta_{l-s-1}\right|\left|E\left(e_{t-l}^2 - \sigma_{t-l}^2\right)^2\right| \qquad (32) \\
& +2T^{-2}\sum_{t=2}^{T}\sum_{s=0}^{t-2}\sum_{j=1}^{t-1}\sum_{l=s+1,l\neq j}^{t-1}\left|\alpha_j\right|\left|\beta_{j-1}\right|\left|\alpha_{l-s}\right|\left|\beta_{l-s-1}\right|\left|E\left(e_{t-j}^2 - E\left(e_{t-j}^2\right)\right)\left(e_{t-l}^2 - E\left(e_{t-l}^2\right)\right)\right|.
\end{aligned}
$$

The inequality is triangle inequality. The second term is zero since for $l \neq j$

$$
E\left(e_{t-j}^2 - E\left(e_{t-j}^2\right)\right)\left(e_{t-l}^2 - E\left(e_{t-l}^2\right)\right) = E\left(e_{t-j}^2 - E\left(e_{t-j}^2\right)\right)E\left(e_{t-l}^2 - E\left(e_{t-l}^2\right)\right) = 0.
$$

The term (32) converges to zero as $T \to \infty$ since

$$\begin{aligned}
& 2T^{-2} \sum_{t=2}^{T} \sum_{s=0}^{t-2} \sum_{l=s+1}^{t-1} |\alpha_l| \left|\beta_{l-1}\right| |\alpha_{l-s}| \left|\beta_{l-s-1}\right| \left| E\left(e_{t-l}^2 - \sigma_{t-l}^2\right)^2 \right| \\
= \;& 2T^{-2} \sum_{t=2}^{T} \sum_{l=1}^{t-1} |\alpha_l| \left|\beta_{l-1}\right| \sum_{s=1}^{l} |\alpha_s| \left|\beta_{s-1}\right| \left| E\left(e_{t-l}^4\right) - \sigma_{t-l}^4 \right| \\
\leq \;& 2T^{-2} \sum_{t=2}^{T} \sum_{l=1}^{t-1} |\alpha_l| \left|\beta_{l-1}\right| \sum_{s=1}^{l} |\alpha_s| \left|\beta_{s-1}\right| \left(\left| E\left(e_{t-l}^4\right) \right| + \left|\sigma_{t-l}^4\right| \right) \\
\leq \;& 2T^{-2} \sum_{t=2}^{T} \sum_{l=1}^{t-1} |\alpha_l| \left|\beta_{l-1}\right| \sum_{s=1}^{l} |\alpha_s| \left|\beta_{s-1}\right| \left(\left| \sup_t E\left(e_t^4\right) \right| + \left|\bar{\sigma}^4\right| \right) \\
= \;& 2 \left(\left| \sup_t E\left(e_t^4\right) \right| + \left|\bar{\sigma}^4\right| \right) T^{-2} \sum_{t=2}^{T} \sum_{l=1}^{t-1} |\alpha_l| \left|\beta_{l-1}\right| \sum_{s=1}^{l} |\alpha_s| \left|\beta_{s-1}\right| \\
\leq \;& 2 \left(\left| \sup_t E\left(e_t^4\right) \right| + \left|\bar{\sigma}^4\right| \right) T^{-2} \sum_{t=1}^{T} \left(\sum_{l=1}^{\infty} \alpha_l^2 \right) \left(\sum_{l=0}^{\infty} \beta_l^2 \right) \left(\sum_{s=1}^{\infty} \alpha_s^2 \right) \left(\sum_{s=0}^{\infty} \beta_s^2 \right) \\
= \;& 2 \left(\left| \sup_t E\left(e_t^4\right) \right| + \left|\bar{\sigma}^4\right| \right) T^{-1} \left(\sum_{l=1}^{\infty} \alpha_l^2 \right)^2 \left(\sum_{l=0}^{\infty} \beta_l^2 \right)^2 .
\end{aligned}$$

The first inequality follows from the triangle inequality, the second inequality follows from equation (3), the third inequality holds because the coefficients $\alpha_l \left(= \pi_l(-\gamma)\right)$ and $\beta_l (= \pi_l(-\delta))$ are square summable.

Case (b). The term in equation (31) for which $i = k$ and $j = l$ simplifies as follows:

$$\begin{aligned}
& 2\, T^{-2} \sum_{t=2}^{T} \sum_{s=0}^{t-2} \sum_{k=s}^{t-1} \sum_{l=s+1}^{t-1} |\alpha_k| \left|\beta_{l-1}\right| |\alpha_{k-s}| \left|\beta_{l-1-s}\right| \left| E\left(e_{t-k}^2 e_{t-l}^2\right) \right| \\
\leq \;& 2\, T^{-2} \sum_{t=2}^{T} \sum_{s=0}^{t-2} \sum_{k=s}^{t-1} |\alpha_k| |\alpha_{k-s}| \sum_{l=s+1}^{t-1} \left|\beta_{l-1}\right| \left|\beta_{l-1-s}\right| \left| \sqrt{E\left(e_{t-k}^4\right) E\left(e_{t-l}^4\right)} \right| \\
\leq \;& 2T^{-2} \sum_{t=2}^{T} \sum_{s=0}^{t-2} \sum_{k=s}^{t-1} |\alpha_k| |\alpha_{k-s}| \sum_{l=s}^{t-2} |\beta_l| \left|\beta_{l-s}\right| \left| \sqrt{\sup_t E\left(e_t^4\right) \sup_t E\left(e_t^4\right)} \right| \\
\leq \;& 2 \sup_t E\left(e_t^4\right) \sqrt{\left(\sum_{k=0}^{\infty} \alpha_k^2 \right) \left(\sum_{l=0}^{\infty} \beta_l^2 \right)} T^{-2} \sum_{t=1}^{T} \sum_{s=0}^{T} \sqrt{\left(\sum_{k=s}^{\infty} \alpha_k^2 \right) \left(\sum_{l=s}^{\infty} \beta_l^2 \right)} \\
= \;& 2 \sup_t E\left(e_t^4\right) \sqrt{\left(\sum_{k=0}^{\infty} \alpha_k^2 \right) \left(\sum_{l=0}^{\infty} \beta_l^2 \right)} T^{-1} \sum_{s=0}^{T} \sqrt{\left(\sum_{k=s}^{\infty} \alpha_k^2 \right) \left(\sum_{l=s}^{\infty} \beta_l^2 \right)} .
\end{aligned}$$

The first inequality follows from Cauchy-Schwartz inequality, the second inequality follows from equation (3) and the third inequality holds since the coefficients α_k and β_l are square summable. From the last equality, the terms $\left(\sum_{k=s}^{\infty} \alpha_k^2\right)$ and $\left(\sum_{l=s}^{\infty} \beta_l^2\right)$ go to zero as $s \to \infty$. Thus by Cesaro summation, as $T \to \infty$, it follows that

$$T^{-1} \sum_{s=0}^{T} \sqrt{\left(\sum_{k=s}^{\infty} \alpha_k^2 \right) \left(\sum_{l=s}^{\infty} \beta_l^2 \right)} \to 0.$$

Case (c). The case when $i = l$ and $j = k$, disappears similarly as in case (b). This completes the proof for part (iii).

Proof of Lemma 4

Define

$$X_{Tt} = s_T^{-1} z_{t-1} e_t$$

where

$$s_T^2 = \sum_{t=2}^{T} E\left(z_{t-1}^2 e_t^2\right).$$

The sequence $\{X_{Tt}\}$ is a martingale difference sequence, since

$$\begin{aligned} E\left(X_{Tt} | \mathcal{F}_{T(t-1)}\right) &= E\left(s_T^{-1} z_{t-1} e_t | \mathcal{F}_{T(t-1)}\right) \\ &= s_T^{-1} z_{t-1} E\left(e_t | \mathcal{F}_{T(t-1)}\right) \\ &= s_T^{-1} z_{t-1} E\left(e_t\right) = 0 \ a.s. \end{aligned}$$

Theorem 6.2.3 of Davidson gives conditions under which the process $\{X_{Tt}\}$ obeys the central limit theorem for martingale differences; that is

$$\sum_{t=2}^{T} X_{Tt} = \frac{T^{-1/2} \sum_{t=2}^{T} z_{t-1} e_t}{\sqrt{T^{-1} \sum_{t=2}^{T} E\left(z_{t-1}^2 e_t^2\right)}} \xrightarrow{d} N(0,1).$$

The first condition requires the square sequence X_{Tt} obeys the weak law of large numbers or

$$\sum_{t=2}^{T} X_{Tt}^2 \xrightarrow{p} 1 \tag{33}$$

and the second condition requires

$$\max_{2 \le t \le T} |X_{Tt}| \xrightarrow{p} 0. \tag{34}$$

Regarding the first condition, note that

$$\sum_{t=2}^{T} E\left(X_{Tt}^2\right) = \frac{\sum_{t=2}^{T} E\left(z_{t-1}^2 e_t^2\right)}{\sum_{t=2}^{T} E\left(z_{t-1}^2 e_t^2\right)} = 1$$

and thus equation (33) can be re-written as

$$\sum_{t=2}^{T} \left(X_{Tt}^2 - E\left(X_{Tt}^2\right)\right) \xrightarrow{p} 0. \tag{35}$$

Even though X_{Tt} is a martingale difference sequence, the squared sequence, X_{Tt}^2, is not a martingale difference sequence. Consequently, the WLLN for martingale difference sequence cannot be used to imply that equation (35) holds. We proceed by rewriting the squared sequence as

$$\begin{aligned}
& \sum_{t=2}^{T}\left(X_{Tt}^2 - E\left(X_{Tt}^2\right)\right) \\
= & \frac{T^{-1}\sum_{t=2}^{T}\left(z_{t-1}^2 e_t^2 - E\left(z_{t-1}^2 e_t^2\right)\right)}{T^{-1}s_T^2} \\
= & \left(T^{-1}s_T^2\right)^{-1}\left(T^{-1}\sum_{t=2}^{T} z_{t-1}^2\left(e_t^2 - \sigma_t^2\right) - T^{-1}\sum_{t=2}^{T}\sigma_t^2\left(z_{t-1}^2 - E\left(z_{t-1}^2\right)\right)\right).
\end{aligned}$$

Thus it suffices to show that equation (35) is true if:

$$T^{-1}\sum_{t=2}^{T}\left(z_{t-1}^2 e_t^2 - z_{t-1}^2\sigma_t^2\right) \xrightarrow{p} 0; \tag{36}$$

and

$$T^{-1}\sum_{t=2}^{T}\sigma_t^2\left(z_{t-1}^2 - E\left(z_{t-1}^2\right)\right) \xrightarrow{p} 0 \tag{37}$$

and

$$0 < \underline{B} \leq T^{-1}s_T^2 \leq \bar{B} < \infty. \tag{38}$$

To show (36), we note that $\left\{z_{t-1}^2 e_t^2 - z_{t-1}^2\sigma_t^2\right\}$ is a martingale difference sequence as

$$\begin{aligned}
E\left(z_{t-1}^2 e_t^2 - z_{t-1}^2\sigma_t^2|\mathcal{F}_{t-1}\right) &= E\left(z_{t-1}^2 e_t^2|\mathcal{F}_{t-1}\right) - E\left(z_{t-1}^2\sigma_t^2|\mathcal{F}_{t-1}\right) \\
&= z_{t-1}^2 E\left(e_t^2|\mathcal{F}_{t-1}\right) - z_{t-1}^2\sigma_t^2 \\
&= z_{t-1}^2\sigma_t^2 - z_{t-1}^2\sigma_t^2 = 0 \quad a.s.
\end{aligned}$$

By Minkowski inequality, Lemma 3(i), and (3), we have

$$\begin{aligned}
E\left(z_{t-1}^2 e_t^2 - z_{t-1}^2\sigma_t^2\right)^2 &\leq \left(\sqrt{E\left(z_{t-1}^4 e_t^4\right)} + \sqrt{E\left(z_{t-1}^4\sigma_t^4\right)}\right)^2 \\
&= \left(\sqrt{E\left(z_{t-1}^4\right)E\left(e_t^4\right)} + \sigma_t^2\sqrt{E\left(z_{t-1}^4\right)}\right)^2 \\
&\leq \left(\sqrt{\sup_{2\leq t\leq T} E\left(z_{t-1}^4\right)\sup_t E\left(e_t^4\right)} + \bar{\sigma}^2\sqrt{\sup_{2\leq t\leq T} E\left(z_{t-1}^4\right)}\right)^2 < \infty.
\end{aligned}$$

Thus by WLLN for martingale differences (Theorem 6.2.2 of Davidson) equation (36) holds.

For the consistency in (37) to be valid, we have to check that the sequence $\left\{\sigma_t^2\left(z_{t-1}^2 - E\left(z_{t-1}^2\right)\right)\right\}$ satisfies the conditions of the Chebyshev Law of Large Numbers (see Davidson 2000 pg 42).

The first condition, which requires the sequence has zero mean, is satisfied trivially. The second condition requires that the variance of the sum tends to zero as $T \to \infty$; that is

$$\lim_{T\to\infty} E\left(T^{-1}\sum_{t=2}^{T}\sigma_t^2\left(z_{t-1}^2 - E\left(z_{t-1}^2\right)\right)\right)^2 = 0. \tag{39}$$

Unlike the above proof, this time the sequence $\left\{\sigma_t^2\left(z_{t-1}^2 - E\left(z_{t-1}^2\right)\right)\right\}$ is not a martingale difference sequence and therefore the variance of the sum of terms is not equal to the sum of the variances. In order to simplify the notation, we will express z_t^2 as

$$\begin{aligned} z_{t-1}^2 &= \left(\Delta^{-\delta} e_{t-1} 1_{(t-1>0)}\right)^2 \\ &= \left(\sum_{i=0}^{t-2}\pi_i\left(-\delta\right)e_{t-1-i}\right)^2 \\ &= \sum_{i=0}^{t-2}\sum_{j=0}^{t-2}\beta_i\beta_j e_{t-1-i}e_{t-1-j}, \end{aligned}$$

where $\beta_i = \pi_i\left(-\delta\right)$. To show (39), we write

$$\begin{aligned} &\left|E\left(T^{-1}\sum_{t=2}^{T}\sigma_t^2\left(z_{t-1}^2 - E\left(z_{t-1}^2\right)\right)\right)^2\right| \\ \leq\ & 2T^{-2}\sum_{t=2}^{T}\sum_{s=0}^{t-2}\sigma_t^2\sigma_{t-s}^2\left|E\left(\left(z_{t-1}^2 - E\left(z_{t-1}^2\right)\right)\left(z_{t-1-s}^2 - E\left(z_{t-1-s}^2\right)\right)\right)\right| \\ \leq\ & \bar{\sigma}^4 2T^{-2}\sum_{t=2}^{T}\sum_{s=0}^{t-2}\left|E\begin{array}{c}\left(\sum_{i=0}^{t-2}\sum_{j=0}^{t-2}\beta_i\beta_j\left(e_{t-1-i}e_{t-1-j} - E\left(e_{t-1-i}e_{t-1-j}\right)\right)\right) \\ \left(\sum_{k=0}^{t-s-2}\sum_{l=0}^{t-s-2}\beta_k\beta_l\left(e_{t-s-1-k}e_{t-s-1-l} - E\left(e_{t-s-1-k}e_{t-s-1-l}\right)\right)\right)\end{array}\right|. \end{aligned}$$

By letting $\alpha_i = \beta_i$ and $\alpha_k = \beta_k$, the above equation tends to zero by arguments similar to those used in proving Lemma 3(i). Thus by the Chebyshev Law of Large Numbers,

$$T^{-1}\sum_{t=2}^{T}\sigma_t^2\left(z_{t-1}^2 - E\left(z_{t-1}^2\right)\right) \xrightarrow{p} 0,$$

as required.

To show (38), we first show that

$$0 < \underline{B} \leq T^{-1}\sum_{t=2}^{T}E\left(z_{t-1}^2 e_t^2\right). \tag{40}$$

To see this, we write

$$\begin{aligned} T^{-1}\sum_{t=2}^{T}E\left(z_{t-1}^{2}e_{t}^{2}\right) &= T^{-1}\sum_{t=2}^{T}E\left(e_{t}^{2}\right)E\left(z_{t-1}^{2}\right) \\ &= T^{-1}\sum_{t=2}^{T}\sigma_{t}^{2}\sum_{j=1}^{t-1}\pi_{j-1}^{2}\left(-\delta\right)\sigma_{t-j}^{2} \\ &= \sum_{j=1}^{T-1}\pi_{j-1}^{2}\left(-\delta\right)T^{-1}\sum_{t=1}^{T-j}\sigma_{t}^{2}\sigma_{t+j}^{2} \\ &\geq \underline{\sigma}^{4}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\pi_{j-1}^{2}\left(-\delta\right). \end{aligned} \tag{41}$$

Since the coefficient $\pi_j\left(-\delta\right)$ is square summable, which means that $\sum_{j=0}^{T-1}\pi_j^2\left(-\delta\right)$ is convergent, Lemma 8.3.1 of Anderson (1971) implies that

$$0<\underline{B}\leq\lim_{T\rightarrow\infty}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\pi_{j-1}^{2}\left(-\delta\right)=\sum_{j=0}^{\infty}\pi_{j}^{2}\left(-\delta\right)<\infty,$$

where $\sum_{j=0}^{\infty}\pi_j^2\left(-\delta\right)>0$. Given that $\underline{\sigma}^4$ is a strictly positive constant, equation (41) is bounded away from zero for all $T\geq 2$. Next for (38) we show

$$T^{-1}\sum_{t=2}^{T}E\left(z_{t-1}^{2}e_{t}^{2}\right)\leq\bar{B}<\infty. \tag{42}$$

Following from the arguments in equation (41), we write

$$\begin{aligned} T^{-1}\sum_{t=2}^{T}E\left(z_{t-1}^{2}e_{t}^{2}\right) &\leq \bar{\sigma}^{4}T^{-1}\sum_{t=2}^{T}\sum_{j=1}^{t-1}\pi_{j-1}^{2}\left(-\delta\right) \\ &\leq \bar{\sigma}^{4}T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{\infty}\pi_{j}^{2}\left(-\delta\right) \\ &= \bar{\sigma}^{4}\sum_{j=0}^{\infty}\pi_{j}^{2}\left(-\delta\right)=\bar{B}<\infty. \end{aligned} \tag{43}$$

Thus (38) has been proved.

Equations (38), (36) and (37) imply that the first condition of the central limit theorem stated in (33) holds. Regarding the second condition for the central limit theorem (see equation (34)), we note that for any $\eta>0$ and for some $\delta>1$,

$$P\left(\max_{2\leq t\leq T}\left|X_{Tt}\right|>\eta\right)\leq\sum_{t=2}^{T}P\left(\left|X_{Tt}\right|>\eta\right)\leq\sum_{t=2}^{T}\frac{E\left|X_{Tt}\right|^{\delta}}{\eta^{\delta}}.$$

Set $\delta = 4$ and note that $\{e_t\}$ are independent, we obtain

$$\begin{aligned}
\sum_{t=2}^{T} E\left(X_{Tt}^4\right) &= s_T^{-4} \sum_{t=2}^{T} E\left(z_{t-1}^4 e_t^4\right) \\
&= \frac{T^{-2}\sum_{t=2}^{T} E\left(z_{t-1}^4 e_t^4\right)}{\left(T^{-1}\sum_{t=2}^{T} E\left(z_{t-1}^2 e_t^2\right)\right)^2} \\
&= \frac{T^{-2}\sum_{t=2}^{T} E\left(z_{t-1}^4\right) E\left(e_t^4\right)}{\left(T^{-1}\sum_{t=2}^{T} E\left(z_{t-1}^2 e_t^2\right)\right)^2} \\
&\leq \frac{T^{-1}\left(\sup_{2\leq t\leq T} E\left(z_{t-1}^4\right)\right)\left(\sup_t E\left(e_t^4\right)\right)}{\left(T^{-1}\sum_{t=2}^{T} E\left(z_{t-1}^2 e_t^2\right)\right)^2} \to 0
\end{aligned}$$

as $T \to \infty$. This follows because Lemma 3(i) and equation (3) uniformly bound the numerator and (38) uniformly bounds the denominator away from zero and infinity.

Proof of Theorem 1

Part (a) Asymptotic Distribution. In order to simplify notation we write z_{t-1} for $\Delta^{d_1} y_{t-1}$. Following from the discussion in section 3.1, $z_t \sim I(1-d_1)$ under the unit root null hypothesis. The values of $(1-d_1)$ will lie in the interval $(0, 1/2)$ since $d_1 \in (1/2, 1)$. Using this and noting that $\Delta y_t = e_t$, the t_W-ratio in (16) can be re-written as

$$\begin{aligned}
t_W(d_1) &= \frac{\sum\limits_{t=2}^{T} z_{t-1} e_t}{\sqrt{\sum\limits_{t=2}^{T} z_{t-1}^2 \hat{e}_t^2}} \\
&= \frac{T^{-\frac{1}{2}} \sum\limits_{t=2}^{T} z_{t-1} e_t}{\sqrt{T^{-1} \sum\limits_{t=2}^{T} E\left(z_{t-1} e_t\right)^2}} \left(\frac{T^{-1} \sum\limits_{t=2}^{T} z_{t-1}^2 \hat{e}_t^2}{T^{-1} \sum\limits_{t=2}^{T} E\left(z_{t-1} e_t\right)^2} \right)^{-\frac{1}{2}}.
\end{aligned}$$

Since $1/2 < d_1 < 1$, the series $\{z_t\}$ is asymptotically stationary and by Lemma 4, it follows that

$$\frac{T^{-\frac{1}{2}} \sum\limits_{t=2}^{T} z_{t-1} e_t}{\left(\sqrt{T^{-1} \sum\limits_{t=2}^{T} E\left(z_{t-1} e_t\right)^2}\right)} \xrightarrow{d} N(0,1).$$

It remains to show that

$$\frac{T^{-1}\sum_{t=2}^{T} z_{t-1}^2 \hat{e}_t^2}{T^{-1}\sum_{t=2}^{T} E\left(z_{t-1}e_t\right)^2} \xrightarrow{p} 1$$

or equivalently

$$\frac{T^{-1}\sum_{t=2}^{T}\left(z_{t-1}^2 \hat{e}_t^2 - E\left(z_{t-1}e_t\right)^2\right)}{T^{-1}\sum_{t=2}^{T} E\left(z_{t-1}e_t\right)^2} \xrightarrow{p} 0. \tag{44}$$

Equation (38) both bounds the denominator uniformly away from zero and ensures that it is finite. It suffices to show that the numerator converges in probability to zero as $T \to \infty$. We note that the residuals $\hat{e}_t$ can be written as $\hat{e}_t = \Delta y_t - \hat{\phi}\Delta^{d_1} y_{t-1} = e_t - \hat{\phi} z_{t-1}$. Using this, we write the numerator in (44) as

$$\begin{aligned} & T^{-1}\sum_{t=2}^{T}\left(z_{t-1}^2\left(e_t - \hat{\phi} z_{t-1}\right)^2 - E\left(z_{t-1}^2 e_t^2\right)\right) \\ = \;& T^{-1}\sum_{t=2}^{T}\left(z_{t-1}^2\left(e_t^2 - 2\hat{\phi} z_{t-1}e_t + \hat{\phi}^2 z_{t-1}^2\right) - \sigma_t^2 E\left(z_{t-1}^2\right)\right) \\ = \;& T^{-1}\sum_{t=2}^{T}\left(z_{t-1}^2 e_t^2 - \sigma_t^2 E\left(z_{t-1}^2\right)\right) - 2\hat{\phi} T^{-1}\sum_{t=2}^{T} z_{t-1}^3 e_t + \hat{\phi}^2 T^{-1}\sum_{t=2}^{T} z_{t-1}^4 \end{aligned} \tag{45}$$

The aim is to show that all the three terms in (45) converge in probability to zero. The first term converges in probability to zero by equations (36) and (37). As for the second and third terms in (45), we will first show that $\hat{\phi} \xrightarrow{p} 0$. Under the null hypothesis, $\hat{\phi}$ in (6) can be written as

$$\hat{\phi} = \frac{T^{-1}\sum_{t=2}^{T} z_{t-1}e_t}{T^{-1}\sum_{t=2}^{T} z_{t-1}^2}.$$

The numerator $T^{-1}\sum_{t=2}^{T} z_{t-1}e_t$ converges to zero in probability by Lemma 3(iii) with $\gamma = 0$. As for the denominator, we have already shown in equation (22) that $T^{-1}\sum_{t=2}^{T} z_{t-1}^2$ does not converge to zero in probability. Therefore $\hat{\phi} \xrightarrow{p} 0$.

Now, coming back to the second term in (45) , since $\hat{\phi} \xrightarrow{p} 0$, it will converge in probability to zero if

$$T^{-1}\sum_{t=2}^{T} z_{t-1}^3 e_t = O_p\left(1\right). \tag{46}$$

Before proving equation (46), we state the Holder's inequality. For any random variables X and Y and for $a > 0$, if $E\left|X\right|^a < \infty$ and $E\left|Y\right|^{a/(a-1)} < \infty$ then

$$E\left|XY\right| \leq \left(E\left|X\right|^a\right)^{1/a}\left(E\left|Y\right|^{a/(a-1)}\right)^{(a-1)/a}.$$

Now to establish (46), by Holder's inequality (with $a = 4$) and Lemma 3(i), we have

$$\begin{aligned} E\left|T^{-1}\sum_{t=2}^{T} z_{t-1}^3 e_t\right| &\leq T^{-1}\sum_{t=2}^{T} E\left|z_{t-1}^3 e_t\right| \\ &\leq T^{-1}\sum_{t=2}^{T}\left(E\left|z_{t-1}^3\right|^{\frac{4}{3}}\right)^{\frac{3}{4}}\left(E\left|e_t\right|^4\right)^{\frac{1}{4}} \\ &= T^{-1}\sum_{t=2}^{T}\left(E\left(z_{t-1}^4\right)\right)^{\frac{3}{4}}\left(E\left(e_t^4\right)\right)^{\frac{1}{4}} \\ &\leq \left(\sup_{2\leq t\leq T} E\left(z_{t-1}^4\right)\right)^{\frac{3}{4}}\left(\sup_t E\left(e_t^4\right)\right)^{\frac{1}{4}} < \infty. \end{aligned}$$

Thus by Markov inequality, equation (46) is true.

In view of Lemma 3(i), the third term in (45) converges in probability to zero by arguments similar to those used in proving the second term.

Part (b) Consistency In order to simplify notation we write z_{t-1} for $\Delta^{d_1} y_{t-1}$ and x_t for Δy_t. Under the alternative hypothesis, the data generating process for $\{y_t\}$ is

$$\Delta^{d_0} y_t = e_t. \tag{47}$$

The sequence $\{z_{t-1}\}$ is fractionally integrated of order $d_0 - d_1$. To see this, we pre-multiply both sides of (47) by $\Delta^{d_1 - d_0}$ to obtain

$$\begin{aligned} \Delta^{d_1} y_{t-1} &= \Delta^{d_1 - d_0} e_{t-1} \\ \Delta^{d_0 - d_1}\left(z_{t-1}\right) &= e_{t-1} \end{aligned} \tag{48}$$

and so $\{z_{t-1}\} \sim I\left(d_0 - d_1\right)$ as claimed. The values of $\left(d_0 - d_1\right)$ will lie in the interval $(-1, 0.5)$ since $d_0 \in [0, 1)$ and $1/2 < d_1 < 1$.

The sequence $\{x_t\}$ is fractionally integrated of order $d_0 - 1$. To see this, we pre-multiply (47) by Δ^{1-d_0} to obtain

$$\begin{aligned} \Delta y_t &= \Delta^{1-d_0} e_t \\ \Delta^{d_0 - 1}\left(x_t\right) &= e_t \end{aligned} \tag{49}$$

and so $\{x_t\} \sim I\left(d_0 - 1\right)$ as claimed. The values of $\left(d_0 - 1\right)$ will lie in the interval $[-1, 0)$.

Using these representations, the $t_W\left(d_1\right)$-ratio in (16) under the alternative hypothesis can be re-written as

$$t_W\left(d_1\right) = T^{1/2}\frac{T^{-1}\sum_{t=2}^{T} z_{t-1} x_t}{\sqrt{T^{-1}\sum_{t=2}^{T}\left(z_{t-1}^2 \hat{e}_t^2\right)}}. \tag{50}$$

The residuals $\hat{e}_t$ can be expressed as

$$\hat{e}_t = \Delta y_t - \hat{\phi}\Delta^{d_1} y_{t-1} = x_t - \hat{\phi} z_{t-1}. \tag{51}$$

Substitute (51) into (50) yields

$$t_W(d_1) = T^{1/2} \frac{T^{-1} \sum_{t=2}^{T} z_{t-1} x_t}{\sqrt{T^{-1} \sum_{t=2}^{T} \left(z_{t-1}^2 \left(x_t - \hat{\phi} z_{t-1} \right)^2 \right)}}. \tag{52}$$

To prove the Theorem, we need to: (a) establish that the numerator in (52) converges in probability to a negative constant; and (ii) establish that the denominator in (52) is uniformly bounded in probability. Then as $T \to \infty$, the $t_W(d_1)$-ratio will diverge to negative infinity, implying that the $t_W(d_1)$-ratio is consistent and this proves the second part of Theorem 1. We organise the presentation of the proof as follows. First we consider the case when $d_1 \neq d_0$ and then we consider the case when $d_1 = d_0$. A separate treatment for the latter case is necessary because the process $\{z_t\}$ is no longer a fractionally integrated process but a short memory process with $z_t = e_t$.

1. $d_1 \neq d_0$. The numerator in (52) can be written as

$$T^{-1} \sum_{t=2}^{T} \left(z_{t-1} x_t - E\left(z_{t-1} x_t \right) \right) + T^{-1} \sum_{t=2}^{T} E\left(z_{t-1} x_t \right).$$

The first term converges in probability to zero by Lemma 3(iii) with $m = 0$. As for the second term, we will show that it converges to a negative constant as $T \to \infty$. To see this, we write

$$\begin{aligned}
T^{-1} \sum_{t=2}^{T} E\left(z_{t-1} x_t \right) &= T^{-1} \sum_{t=2}^{T} E\left(\Delta^{d_1 - d_0} e_{t-1} \Delta^{1-d_0} e_t \right) \\
&= T^{-1} \sum_{t=2}^{T} E\left(\sum_{i=1}^{t-1} \pi_i \left(d_1 - d_0 \right) e_{t-i} \sum_{j=0}^{t-1} \pi_j \left(1 - d_0 \right) e_{t-j} \right) \\
&= T^{-1} \sum_{t=2}^{T} \sum_{i=1}^{t-1} \sum_{j=0}^{t-1} \pi_{i-1} \left(d_1 - d_0 \right) \pi_j \left(1 - d_0 \right) E\left(e_{t-i} e_{t-j} \right) \\
&= T^{-1} \sum_{t=2}^{T} \sum_{i=1}^{t-1} \pi_i \left(1 - d_0 \right) \pi_{i-1} \left(d_1 - d_0 \right) \sigma_{t-i}^2 \\
&= \sum_{i=1}^{T-1} \pi_i \left(1 - d_0 \right) \pi_{i-1} \left(d_1 - d_0 \right) T^{-1} \sum_{t=i+1}^{T} \sigma_{t-i}^2 \\
&= \sum_{i=1}^{T-1} \pi_i \left(1 - d_0 \right) \pi_{i-1} \left(d_1 - d_0 \right) T^{-1} \sum_{t=1}^{T-i} \sigma_t^2 \\
&= \left(\sum_{i=1}^{T-1} \pi_i \left(1 - d_0 \right) \pi_{i-1} \left(d_1 - d_0 \right) \right) \left(T^{-1} \sum_{t=1}^{T} \sigma_t^2 \right) \\
&\quad - \sum_{i=1}^{T-1} \pi_i \left(1 - d_0 \right) \pi_{i-1} \left(d_1 - d_0 \right) T^{-1} \sum_{t=T-i+1}^{T} \sigma_t^2 \\
&= R_{1T} - R_{2T}.
\end{aligned}$$

As $T \to \infty$, we will show that R_{1T} converges to a negative constant; that is

$$\lim_{T\to\infty} R_{1T} = \left(\lim_{T\to\infty} \sum_{i=1}^{T-1} \pi_i\left(1-d_0\right)\pi_{i-1}\left(d_1-d_0\right)\right)\left(\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sigma_t^2\right) = -C \tag{53}$$

for some $0 < C < \infty$ and R_{2T} converges to zero; that is

$$\lim_{T\to\infty} R_{2T} = 0. \tag{54}$$

Regarding R_{1T}, we will first show that, as $T \to \infty$,

$$\lim_{T\to\infty}\left|\sum_{i=1}^{T-1} \pi_i\left(1-d_0\right)\pi_{i-1}\left(d_1-d_0\right)\right| < \infty \tag{55}$$

and then we will show that its limiting value is strictly negative; that is

$$\lim_{T\to\infty} \sum_{i=1}^{T-1} \pi_i\left(1-d_0\right)\pi_{i-1}\left(d_1-d_0\right) < 0. \tag{56}$$

Equations (55) and (56), taken together, imply that equation (53) holds.

To show (55) , note that for large i, the sequence

$$\pi_i\left(1-d_0\right)\pi_{i-1}\left(d_1-d_0\right) \sim i^{-(1-d_0)-1} i^{-(d_1-d_0)-1} \sim i^{-3+2d_0-d_1}.$$

For the above sequence to be summable, we require that $-3+2d_0-d_1 < -1$ or $d_0 < 1+\frac{1}{2}d_1$, which holds since $d_1 > 1/2$ and $d_0 < 1$ under the alternative hypothesis. Thus equation (55) holds. Next to show (56) , we note that the coefficient $\pi_i\left(1-d_0\right)$ is negative for all $i \geq 1$ but the coefficient $\pi_i\left(d_1-d_0\right)$ is not necessary positive. We consider the following two cases:

(i) Assume that $d_1 < d_0$. Then, the coefficient $\pi_i\left(d_1-d_0\right)$ is positive for all $i \geq 0$. Thus, the product of $\pi_i\left(1-d_0\right)$ and $\pi_i\left(d_1-d_0\right)$ will be negative for all $i \geq 1$. Hence (56) holds as required.

(ii) Assume that $d_1 > d_0$. Then, $\pi_0\left(d_1-d_0\right) = 1$ but for all $i \geq 1$ the coefficient $\pi_i\left(d_1-d_0\right)$ is negative. In contrast to the previous case, now $\pi_i\left(1-d_0\right)\pi_{i-1}\left(d_1-d_0\right) > 0$. In order to show that equation (56) is true, we need only show that

$$\pi_1\left(1-d_0\right) < -\left(\sum_{i=2}^{\infty} \pi_i\left(1-d_0\right)\pi_{i-1}\left(d_1-d_0\right)\right). \tag{57}$$

DGM have shown that the absolute value of the right hand term of the above equation can be

$$\begin{aligned}\left|\sum_{i=2}^{\infty} \pi_i\left(1-d_0\right)\pi_{i-1}\left(d_1-d_0\right)\right| &\leq \sup_{j\in[2,\infty)}\left|\pi_j\left(1-d_0\right)\right| \sum_{i=1}^{\infty}\left|\pi_i\left(d_1-d_0\right)\right| \\ &= \left|\pi_2\left(1-d_0\right)\right|.\end{aligned}$$

The last equality is obtained by noting that $\sum_{i=1}^{\infty}|\pi_i(d_1-d_0)|=1$. Since $|\pi_2(1-d_0)|<|\pi_1(1-d_0)|$, it follows that

$$\left|\sum_{i=2}^{\infty}\pi_i(1-d_0)\pi_{i-1}(d_1-d_0)\right|<|\pi_1(1-d_0)|$$

as required to show that the inequality in (57) is true.

To show (54), we write

$$\begin{aligned}
&\left|\sum_{i=1}^{T-1}\pi_i(1-d_0)\pi_{i-1}(d_1-d_0)T^{-1}\sum_{t=T-i+1}^{T}\sigma_t^2\right| \\
\leq\;& \sum_{i=1}^{T-1}|\pi_i(1-d_0)|\,|\pi_{i-1}(d_1-d_0)|\,T^{-1}\sum_{t=T-i+1}^{T}\bar{\sigma}^2 \\
=\;& T^{-1}\sum_{i=1}^{T-1}|\pi_i(1-d_0)|\,|\pi_{i-1}(d_1-d_0)|\,i\bar{\sigma}^2 \\
\leq\;& B^*\bar{\sigma}^2T^{-1}\sum_{i=1}^{T-1}i^{-(1-d_0)-1}i^{-(d_1-d_0)-1}i \\
=\;& B^*\bar{\sigma}^2T^{-1}\sum_{i=1}^{T-1}i^{2d_0-2-d_1}\to 0
\end{aligned}$$

(for some $B^*>0$) as $T\to\infty$ by Cesaro Summation since $2d_0-2-d_1<0$ under the alternative hypothesis. This therefore completes the proof that the numerator of the $t_W(d_1)$-ratio in (52) converges in probability to a negative constant.

The denominator of the $t_W(d_1)$-ratio in (52) can be written as

$$\begin{aligned}
&T^{-1}\sum_{t=2}^{T}\left(z_{t-1}^2\left(x_t-\hat{\phi}z_{t-1}\right)^2\right) \\
=\;& T^{-1}\sum_{t=2}^{T}z_{t-1}^2x_t^2-2\hat{\phi}T^{-1}\sum_{t=2}^{T}z_{t-1}^3x_t+\hat{\phi}^2T^{-1}\sum_{t=2}^{T}z_{t-1}^4. \qquad (58)
\end{aligned}$$

We will show that each term in (58) is uniformly bounded in probability or $O_p(1)$. As for the first term, by Cauchy-Schwarz (CS) inequality and Lemma 3(i), it follows that

$$\begin{aligned}
E\left|T^{-1}\sum_{t=2}^{T}z_{t-1}^2x_t^2\right| &\leq T^{-1}\sum_{t=2}^{T}E\left|z_{t-1}^2x_t^2\right| \\
&\leq T^{-1}\sum_{t=2}^{T}\sqrt{E\left(z_{t-1}^4\right)E\left(x_t^4\right)} \\
&\leq \sqrt{\left(\sup_{2\leq t\leq T}E\left(z_{t-1}^4\right)\right)\left(\sup_{t}E\left(x_t^4\right)\right)}<\infty.
\end{aligned}$$

Therefore the first term in (58) is $O_p(1)$ by Markov inequality.

As for the second term and third terms in (58), we first show that $\hat{\phi} = O_p(1)$. Under the alternative hypothesis and using expressions (48) and (49), the $\hat{\phi}$ in (6) can be rewritten as

$$\hat{\phi} = \frac{T^{-1} \sum_{t=2}^{T} z_{t-1} x_t}{T^{-1} \sum_{t=2}^{T} z_{t-1}^2}.$$

The numerator of $\hat{\phi}$ is $O_p(1)$ since it is the same expression as that given in the numerator of the $t_W(d_1)$-ratio, see (52). We have already shown in equation (22) that the denominator does not converge to zero in probability. Therefore $\hat{\phi} = O_p(1)$.

Now coming back to the second term in (58), all that is required is to show that

$$T^{-1} \sum_{t=2}^{T} z_{t-1}^3 x_t = O_p(1)$$

since $\hat{\phi} = O_p(1)$. To see this, by Holder's inequality and Lemma A (i), it follows that

$$\begin{aligned} E\left|T^{-1} \sum_{t=2}^{T} z_{t-1}^3 x_t\right| &\leq T^{-1} \sum_{t=2}^{T} E\left|z_{t-1}^3 x_t\right| \\ &\leq T^{-1} \sum_{t=2}^{T} \left(E\left|z_{t-1}^3\right|^{\frac{4}{3}}\right)^{\frac{3}{4}} \left(E\left|x_t\right|^4\right)^{\frac{1}{4}} \\ &= T^{-1} \sum_{t=2}^{T} \left(E\left(z_{t-1}^4\right)\right)^{\frac{3}{4}} \left(E\left(x_t^4\right)\right)^{\frac{1}{4}} \\ &\leq \left(\sup_{2\leq t\leq T} E\left(z_{t-1}^4\right)\right)^{\frac{3}{4}} \left(\sup_t E\left(x_t^4\right)\right)^{\frac{1}{4}} < \infty. \end{aligned}$$

The third term in (58) is $O_p(1)$ by using similar arguments as above. Therefore the denominator in (58) is $O_p(1)$ and this completes the proof for the case when $d_1 \neq d_0$.

2. $d_1 = d_0$. In this case $z_{t-1} = e_{t-1}$ and thus the numerator in (52) can be written as

$$T^{-1} \sum_{t=2}^{T} \left(e_{t-1} x_t - E\left(e_{t-1} x_t\right)\right) + T^{-1} \sum_{t=2}^{T} E\left(e_{t-1} x_t\right). \tag{59}$$

The first term converges in probability to zero by Lemma 3(iii) with $\delta = 0$ and $n = 1$.

Regarding the second term, we write

$$\begin{aligned}T^{-1}\sum_{t=2}^{T}E\left(e_{t-1}x_t\right) &= T^{-1}\sum_{t=2}^{T}E\left(e_{t-1}\Delta^{1-d_0}e_t\right)\\ &= T^{-1}\sum_{t=2}^{T}\sum_{j=0}^{t-1}\pi_j\left(1-d_0\right)E\left(e_{t-j}e_{t-1}\right)\\ &= T^{-1}\sum_{t=2}^{T}\pi_1\left(1-d_0\right)\sigma_{t-1}^2\\ &= \pi_1\left(1-d_0\right)\left(T^{-1}\sum_{t=1}^{T-1}\sigma_t^2\right).\end{aligned}$$

Under Assumption V, the term $T^{-1}\sum_{t=1}^{T}\sigma_t^2$ is uniformly bounded away from zero and infinity for all T and since the coefficient $\pi_1\left(1-d_0\right)<0$, the second term in (59) converges to a negative constant as $T\to\infty$. This completes the proof that the numerator in (52) converges in probability to a negative constant. The denominator in (52) is $O_p\left(1\right)$ by similar argument to the $d_1\neq d_0$ case. This completes the proof for the case where $d_1=d_0$.

Proof of Theorem 2

Under the null hypothesis, $\hat{d}$ is a consistent estimator of $d_0=1$. The trimming rule defined in (17) implies that the pre-estimated value of d_1 $\left(\hat{d}_1\right)$ is also a consistent estimator of $(1-c)$; that is

$$\hat{d}_1\xrightarrow{p}\left(1-c\right).$$

Following DGM, we will show that

$$t_W\left(\hat{d}_1\right)-t_W\left(1-c\right)=o_p\left(1\right). \tag{60}$$

Since $0<c<1/2$, part (a) of Theorem 1 shows that $t_W\left(1-c\right)\xrightarrow{d}N\left(0,1\right)$ and thus an asymptotic equivalence argument implies that $t_W\left(\hat{d}_1\right)\xrightarrow{d}N\left(0,1\right).$

To show (60), we apply the mean value theorem to $t_W\left(\hat{d}_1\right)$ around $(1-c)$ to obtain

$$t_W\left(\hat{d}_1\right)-t_W\left(1-c\right)=\left.\frac{\partial t_W\left(d_1\right)}{\partial d_1}\right|_{d_1=d^*}\left(\hat{d}_1-\left(1-c\right)\right)$$

where $\hat{d}_1\leq d^*\leq\left(1-c\right)$. Since $\hat{d}_1-\left(1-c\right)=o_p\left(1\right)$, it suffices to show that

$$\left.\frac{\partial t_W\left(d_1\right)}{\partial d_1}\right|_{d_1=d^*}=O_p\left(1\right). \tag{61}$$

To show (61), we show

$$\frac{\partial t_W\left(d_1\right)}{\partial d_1}=O_p\left(1\right) \tag{62}$$

for all $d_1 \in [1-c-\varepsilon, 1-c]$ with any $0 < \varepsilon < 1/2 - c$. Then (61) follows since $d^* \xrightarrow{p} (1-c)$, that is

$$\Pr\left(d^* \in [1-c-\varepsilon, 1-c]\right) \to 1.$$

The remainder of the proof is concerned with showing (62).

We next evaluate the first derivative of $t_W(d_1)$. Under H_0 we have $\Delta y_t = e_t$ and $\Delta^{d_1} y_{t-1} = \Delta^{d_1-1} e_t$. The $t_W(d_1)$ ratio can thus be written as

$$t_W(d_1) = \frac{\sum_{t=2}^{T} e_t \Delta^{d_1-1} e_{t-1}}{\sqrt{\sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^2 \left(e_t - \hat{\phi}\Delta^{d_1-1} e_{t-1}\right)^2}} \tag{63}$$

with

$$\hat{\phi} = \frac{\sum_{t=2}^{T} e_t \Delta^{d_1-1} e_{t-1}}{\sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^2}. \tag{64}$$

When we substitute (64) into (63), the denominator can be written as

$$\begin{aligned}
& \sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^2 \left(e_t - \hat{\phi}\Delta^{d_1-1} e_{t-1}\right)^2 \\
= & \sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^2 e_t^2 - 2\hat{\phi} \sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^3 e_t + \hat{\phi}^2 \sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^4 \\
= & \sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^2 e_t^2 - 2\left(\sum_{t=2}^{T} e_t \Delta^{d_1-1} e_{t-1}\right)\left(\sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^3 e_t\right)\left(\sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^2\right)^{-1} \\
& + \left(\sum_{t=2}^{T} e_t \Delta^{d_1-1} e_{t-1}\right)^2 \left(\sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^4\right)\left(\sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^2\right)^{-2}.
\end{aligned}$$

To simplify notations, we let:

$$N_T(d_1) = \sum_{t=2}^{T} e_t \Delta^{d_1-1} e_{t-1}, \tag{65}$$

$$D_{1T}(d_1) = \sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^3 e_t, \tag{66}$$

$$D_{2T}(d_1) = \sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^2, \tag{67}$$

$$D_{3T}(d_1) = \sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^4, \text{ and} \tag{68}$$

$$D_{4T}(d_1) = \sum_{t=2}^{T} \left(\Delta^{d_1-1} e_{t-1}\right)^2 e_t^2. \tag{69}$$

Using equations $(65) - (69)$, the $t_W(d_1)$ in (63) can then be written as

$$t_W(d_1) = N_T(d_1) D_T(d_1)^{-1/2} \tag{70}$$

where $D_T(d_1)$ is

$$\begin{aligned} D_T(d_1) &= D_{4T}(d_1) - 2N_T(d_1) D_{1T}(d_1) D_{2T}(d_1)^{-1} \\ &\quad + N_T(d_1)^2 D_{3T}(d_1) D_{2T}(d_1)^{-2}. \end{aligned}$$

The first derivative of $t_W(d_1)$ is given by

$$\begin{aligned} &\frac{\partial t_W(d_1)}{\partial d_1} \\ &= \frac{\partial N_T(d_1)}{\partial d_1} D_T(d_1)^{-\frac{1}{2}} + \left(\frac{\partial \left(D_T(d_1)^{-1/2} \right)}{\partial d_1} \right) N_T(d_1). \end{aligned} \tag{71}$$

Therefore equation (62) holds if, for all $d_1 \in [1-c-\varepsilon, 1-c]$ with any $0 < \varepsilon < 1/2 - c$,

$$\left(T^{-1/2} \frac{\partial N_T(d_1)}{\partial d_1} \right) \left(T^{-1} D_T(d_1) \right)^{-\frac{1}{2}} = O_p(1) \tag{72}$$

and

$$\left(T^{1/2} \frac{\partial \left(D_T(d_1)^{-1/2} \right)}{\partial d_1} \right) \left(T^{-1/2} N_T(d_1) \right) = O_p(1). \tag{73}$$

For equation (72), Lemma 5 below shows that

$$T^{-1/2} \frac{\partial N_T(d_1)}{\partial d_1} = O_p(1).$$

Given that $D_T(d_1)$ appears in the denominator, we wish to show that $T^{-1} D_T(d_1)$ does not converge to zero in probability. To show this we write

$$\begin{aligned} &T^{-1} D_T(d_1) \\ &= T^{-1} D_{4T}(d_1) - 2T^{-1} N_T(d_1) T^{-1} D_{1T}(d_1) \left(T^{-1} D_{2T}(d_1) \right)^{-1} \\ &\quad + \left(T^{-1} N_T(d_1) \right)^2 T^{-1} D_{3T}(d_1) \left(T^{-2} D_{2T}(d_1) \right)^{-2}. \end{aligned} \tag{74}$$

Let $\Delta^{d_1 - 1} e_{t-1} = z_{t-1}$. Since $d_1 \in [1-c-\varepsilon, 1-c]$, we can apply Lemma 3 to the terms in (xx). The first term in (74) does not converge to zero in probability. To see this, we write

$$\begin{aligned} &T^{-1} D_{4T}(d_1) \\ &= T^{-1} \sum_{t=2}^{T} \left(z_{t-1}^2 e_t^2 - E\left(z_{t-1}^2 e_t^2 \right) \right) + T^{-1} \sum_{t=2}^{T} E\left(z_{t-1}^2 e_t^2 \right). \end{aligned} \tag{75}$$

the first term in (75) converges to zero in probability by (36) and (37). The second term in (75) is bounded away from zero by (38).

We will show that the second and third terms in (74) converge to zero in probability. To see this, we note the following convergence results for $T^{-1}D_{2T}(d_1), T^{-1}N_T(d_1), T^{-1}D_{1T}(d_1)$ and $T^{-1}D_{3T}(d_1)$. Given that $T^{-1}D_{2T}(d_1) = T^{-1}\sum_{t=2}^{T} z_{t-1}^2$ appears in the denominator of (74), we have shown that equation (22) does not converge to zero in probability. We have

$$\begin{aligned} T^{-1}N_T(d_1) &= T^{-1}\sum_{t=2}^{T} e_t\Delta^{d_1-1}e_{t-1} \\ &= T^{-1}\sum_{t=2}^{T} e_t z_{t-1} = o_p(1) \end{aligned} \tag{76}$$

by Lemma 3(ii) with $\delta = 0$. We have

$$\begin{aligned} T^{-1}D_{1T}(d_1) &= T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^3 e_t \\ &= T^{-1}\sum_{t=2}^{T} z_{t-1}^3 e_t = O_p(1) \end{aligned} \tag{77}$$

by equation (46). We have

$$T^{-1}D_{3T}(d_1) = T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^4 = O_p(1) \tag{78}$$

by equation (**??**). Thus (72) is shown.

As for equation (73), since Lemma 4 and (38) imply that

$$T^{-1/2}N_T(d_1) = T^{-1/2}\sum_{t=2}^{T} e_t\Delta^{d_1-1}e_{t-1} = O_p(1)$$

it suffices to show that

$$T^{1/2}\frac{\partial\left(D_T(d_1)^{-1/2}\right)}{\partial d_1} = O_p(1). \tag{79}$$

To show (79), we write

$$\begin{aligned} & T^{1/2}\frac{\partial\left(D_T(d_1)^{-1/2}\right)}{\partial d_1} \\ = \; & -\frac{1}{2}\left(T^{-1}D_T(d_1)\right)^{-3/2}\left(T^{-1}\frac{\partial D_{4T}(d_1)}{\partial d_1} - 2T^{-1}\frac{\partial\left(N_T(d_1)\,D_{1T}(d_1)\,D_{2T}(d_1)^{-1}\right)}{\partial d_1}\right. \\ & \left. + T^{-1}\frac{\partial\left(N_T(d_1)^2\,D_{3T}(d_1)\,D_{2T}(d_1)^{-2}\right)}{\partial d_1}\right) \end{aligned}$$

We have already shown in the proof of (72) that $D_T(d_1)$ does not converge to zero in probability. From Lemma 5 below,

$$T^{-1}\frac{\partial D_{4T}(d_1)}{\partial d_1} = O_p(1).$$

We will show that

$$T^{-1}\frac{\partial\left(N_T(d_1)\,D_{1T}(d_1)\,D_{2T}(d_1)^{-1}\right)}{\partial d_1} \xrightarrow{p} 0 \tag{80}$$

and

$$T^{-1}\frac{\partial\left(N_T(d_1)^2\,D_{3T}(d_1)\,D_{2T}(d_1)^{-2}\right)}{\partial d_1} \xrightarrow{p} 0. \tag{81}$$

To show (80), we write

$$\begin{aligned}
&T^{-1}\frac{\partial\left(N_T(d_1)\,D_{1T}(d_1)\,D_{2T}(d_1)^{-1}\right)}{\partial d_1}\\
=\;& T^{-1}D_{1T}(d_1)\left(T^{-1}D_{2T}(d_1)\right)^{-1}T^{-1}\frac{\partial N_T(d_1)}{\partial d_1}\\
&+T^{-1}N_T(d_1)\left(T^{-1}D_{2T}(d_1)\right)^{-1}T^{-1}\frac{\partial D_{1T}(d_1)}{\partial d_1}\\
&-T^{-1}N_T(d_1)\,T^{-1}D_{1T}(d_1)\left(T^{-1}D_{2T}(d_1)\right)^{-2}T^{-1}\frac{\partial D_{2T}(d_1)}{\partial d_1}.
\end{aligned} \tag{82}$$

As noted previously, $T^{-1}D_{2T}(d_1)$ does not converge to zero in probability. The first term on the right-hand side of (82) is $o_p(1)$ by (77) and Lemma 5(d) below. The second term is $o_p(1)$ by (76) and Lemma 5(e) below. The third term is $o_p(1)$ by (76), (77) and Lemma 5(f) below.

To show (81), we write

$$\begin{aligned}
&T^{-1}\frac{\partial\left(N_T(d_1)^2\,D_{3T}(d_1)\,D_{2T}(d_1)^{-2}\right)}{\partial d_1}\\
=\;& 2T^{-1}N_T(d_1)\,T^{-1}D_{3T}(d_1)\left(T^{-1}D_{2T}(d_1)\right)^{-2}T^{-1}\frac{\partial N_T(d_1)}{\partial d_1}\\
&+\left(T^{-1}N_T(d_1)\right)^2\left(T^{-1}D_{2T}(d_1)\right)^{-2}T^{-1}\frac{\partial D_{3T}(d_1)}{\partial d_1}\\
&-2\left(T^{-1}N_T(d_1)\right)^2T^{-1}D_{3T}(d_1)\left(T^{-1}D_{2T}(d_1)\right)^{-3}T^{-1}\frac{\partial D_{2T}(d_1)}{\partial d_1}.
\end{aligned}$$

By similar arguments as above, the first term is $o_p(1)$ by (76), (78) and Lemma 5(d) below. The second term is $o_p(1)$ by (76) and Lemma 5(g). The third term is $o_p(1)$ by (76), (78) and 5(f) below.

The next Lemma gives the asymptotic properties of the first-order derivatives for the expressions (65)-(69) defined above.

Lemma 5 *Let $\{e_t\}$ be a sequence of random variables generated according to (2) with σ_t and $\{\varepsilon_t\}$ satisfy Assumptions V and E respectively. Let $0 < c < 1/2$. If $d_1 \in [1-c-\varepsilon, 1-c]$ for any $0 < \varepsilon < 1/2 - c$ then:*
(a) $\sup_{2\leq t\leq T} E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2 < B_c < \infty$
(b) $\sup_{2\leq t\leq T} E\left(\Delta^{d_1-1}e_{t-1}\right)^2\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2 < B_c < \infty$
(c) $\sup_{d_1} T^{-1/2}\frac{\partial N_T(d_1)}{\partial d_1} = \sup_{d_1} T^{-1/2}\sum_{t=2}^{T} e_t\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right) = O_p(1);$
(d) $\sup_{d_1} T^{-1}\frac{\partial N_T(d_1)}{\partial d_1} = \sup_{d_1} T^{-1}\sum_{t=2}^{T} e_t\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right) = o_p(1);$
(e) $\sup_{d_1} T^{-1}\frac{\partial D_{1T}(d)}{\partial d} = \sup_{d_1} 3T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^2 e_t\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right) = o_p(1);$
(f) $\sup_{d_1} T^{-1}\frac{\partial D_{2T}(d)}{\partial d} = \sup_{d_1} 2T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right) = O_p(1);$
(g) $\sup_{d_1} T^{-1}\frac{\partial D_{3T}(d)}{\partial d} = \sup_{d_1} 4T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^3\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right) = O_p(1);$
(h) $\sup_{d_1} T^{-1}\frac{\partial D_{4T}(d)}{\partial d} = \sup_{d_1} 2T^{-1}\sum_{t=2}^{T} e_t^2\left(\Delta^{d_1-1}e_{t-1}\right)\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right) = O_p(1).$

Proof of Lemma 5

Throughout the proof, we use the formula

$$\log\Delta = \log(1-L) = -\sum_{i=1}^{\infty}\frac{1}{i}L^i. \tag{83}$$

According to DGM, (83) holds because the function $\log(1-L)$ is analytic in the convergence disc $|z| < 1$.

(a) We write

$$\begin{aligned}
& E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2 \\
= & \left|E\left(\sum_{i=1}^{t-2}\frac{1}{i}\Delta^{d_1-1}e_{t-1-i}\right)^2\right| \\
= & \left|\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}E\left(\left(\Delta^{d_1-1}e_{t-1-i}\right)\left(\Delta^{d_1-1}e_{t-1-j}\right)\right)\right| \\
= & \left|\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\sum_{k=i}^{t-2}\sum_{l=j}^{t-2}\pi_{k-i}(d_1-1)\pi_{l-j}(d_1-1)E\left(e_{t-1-k}e_{t-1-l}\right)\right| \\
\leq & \sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\sum_{k=i\vee j}^{t-2}\left|\pi_{k-i}(d_1-1)\right|\left|\pi_{k-j}(d_1-1)\right|\sigma_{t-1-k}^2 \\
\leq & \bar{\sigma}^2\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{i}\frac{1}{j}\sum_{k=i\vee j}^{\infty}\left|\pi_{k-i}(d_1-1)\right|\left|\pi_{k-j}(d_1-1)\right| = B_{d_1} < B_c < \infty.
\end{aligned}$$

The second last inequality follows from Lemma 6 below. Thus, (a) is established.

(b) We write

$$
\begin{aligned}
& E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\left(\Delta^{d_1-1}e_{t-1}\right)^2 \\
= & \left|E\left(\sum_{i=1}^{t-2}\frac{1}{i}\Delta^{d_1-1}e_{t-1-i}\right)^2\left(\Delta^{d_1-1}e_{t-1}\right)^2\right| \\
= & \left|E\left(\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\left(\left(\Delta^{d_1-1}e_{t-1-i}\right)\left(\Delta^{d_1-1}e_{t-1-j}\right)\right)\right)\left(\Delta^{d_1-1}e_{t-1}\right)^2\right| \\
= & \left|\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}E\left(\begin{array}{c}\left(\sum_{k=0}^{t-2-i}\sum_{l=0}^{t-2-j}\pi_k(d_1-1)\pi_l(d_1-1)e_{t-1-i-k}e_{t-1-j-l}\right)\\ \left(\sum_{m=0}^{t-2}\sum_{n=0}^{t-2}\pi_m(d_1-1)\pi_n(d_1-1)e_{t-1-m}e_{t-1-n}\right)\end{array}\right)\right| \\
= & \left|\begin{array}{c}\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\sum_{k=0}^{t-2-i}\sum_{l=0}^{t-2-j}\sum_{m=0}^{t-2}\sum_{n=0}^{t-2}\pi_k(d_1-1)\pi_l(d_1-1)\pi_m(d_1-1)\pi_n(d_1-1)\\ E\left(e_{t-1-j-k}e_{t-1-i-l}e_{t-1-m}e_{t-1-n}\right)\end{array}\right| \\
= & \left|\begin{array}{c}\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\sum_{k=i}^{t-2}\sum_{l=j}^{t-2}\sum_{m=0}^{t-2}\sum_{n=0}^{t-2}\pi_{k-i}(d_1-1)\pi_{l-j}(d_1-1)\pi_m(d_1-1)\pi_n(d_1-1)\\ E\left(e_{t-1-k}e_{t-1-l}e_{t-1-m}e_{t-1-n}\right)\end{array}\right| \qquad (84)
\end{aligned}
$$

We show that the terms for which non-zero expectations arise are uniformly bounded in $2 \le t \le T$. First, consider the case in which the summation indices $k = l$ and $m = n$. Equation (84) thus becomes

$$
\begin{aligned}
& \left|\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\sum_{k=i\vee j}^{t-2}\sum_{m=0}^{t-2}\pi_{k-j}(d_1-1)\pi_{k-i}(d_1-1)\pi_m^2(d_1-1)E\left(e_{t-1-k}^2e_{t-1-m}^2\right)\right| \\
\le & \left|\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\sum_{k=i\vee j}^{t-2}\sum_{m=0}^{t-2}\pi_{k-j}(d_1-1)\pi_{k-i}(d_1-1)\pi_m^2(d_1-1)\sqrt{E\left(e_{t-1-k}^4\right)E\left(e_{t-1-m}^4\right)}\right| \\
\le & \sup_t E\left(e_t^4\right)\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\sum_{k=i\vee j}^{t-2}\left|\pi_{k-j}(d_1-1)\right|\left|\pi_{k-i}(d_1-1)\right|\sum_{m=0}^{t-2}\pi_m^2(d_1-1) \\
\le & \sup_t E\left(e_t^4\right)\left(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{i}\frac{1}{j}\sum_{k=i\vee j}^{\infty}\left|\pi_{k-j}(d_1-1)\right|\left|\pi_{k-i}(d_1-1)\right|\right)\left(\sum_{m=0}^{\infty}\pi_m^2(d_1-1)\right) < \infty
\end{aligned}
$$

uniformly in $2 \le t \le T$ by equation (3), Lemma 6 below and the fact that $\pi_m(d_1-1)$ is square summable. Second, when $m=k$ and $n=l$, equation (84) becomes

$$\begin{aligned}
& 2\left|\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\sum_{k=i}^{t-2}\sum_{l=j}^{t-2}\pi_{k-i}(d_1-1)\,\pi_{l-j}(d_1-1)\,\pi_k(d_1-1)\,\pi_l(d_1-1)\,E\left(e_{t-1-k}^2 e_{t-1-l}^2\right)\right| \\
\leq\ & 2\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\sum_{k=i}^{t-2}\sum_{l=j}^{t-2}|\pi_{k-i}(d_1-1)|\,|\pi_{l-j}(d_1-1)|\,|\pi_k(d_1-1)|\,|\pi_l(d_1-1)|\sqrt{E\left(e_{t-1-k}^4\right)E\left(e_{t-1-l}^4\right)} \\
\leq\ & \sup_t E\left(e_t^4\right)2\sum_{i=1}^{t-2}\sum_{j=1}^{t-2}\frac{1}{i}\frac{1}{j}\sum_{k=i}^{t-2}|\pi_{k-i}(d_1-1)|\,|\pi_k(d_1-1)|\sum_{l=j}^{t-2}|\pi_{l-j}(d_1-1)|\,|\pi_l(d_1-1)| \\
\leq\ & \sup_t E\left(e_t^4\right)2\left(\sum_{i=1}^{t-2}\frac{1}{i}\sum_{k=i}^{\infty}|\pi_{k-i}(d_1-1)|\,|\pi_k(d_1-1)|\right)\left(\sum_{j=1}^{t-2}\frac{1}{j}\sum_{l=j}^{\infty}|\pi_{l-j}(d_1-1)|\,|\pi_l(d_1-1)|\right) \\
\leq\ & \sup_t E\left(e_t^4\right)2\left(\sum_{i=1}^{\infty}\frac{1}{i}\gamma_i^*(1-d_1)\right)^2 < B_c < \infty
\end{aligned}$$

where $\gamma_i^*(1-d_1)$ denotes the autocovariance function defined in Lemma 6 below. The last inequality follows from the fact that $\gamma_i^*(1-d_1)\leq Bi^{2(1-d_1)}$ the sequence $\left\{i^{-1}\gamma_i^*(1-d_1)\right\}$ is summable since the asymptotic approximation of $\gamma_i^*(1-d_1)$ is $i^{2(1-d_1)-1}=i^{1-2d_1}$.

Finally, when the summation indices $m=l$ and $n=k$, equation (84) is uniformly bounded in $2\leq t\leq T$ by arguments similar to those above. Therefore, (b) is proved.

(c) Note that $\left\{e_t\Delta^{d_1-1}\log\Delta e_{t-1},\mathcal{F}_t\right\}$ is a martingale difference sequence as

$$E\left(e_t\Delta^{d_1-1}\log\Delta e_{t-1}|\mathcal{F}_{t-1}\right)=\Delta^{d_1-1}\log\Delta e_{t-1}E\left(e_t|\mathcal{F}_{t-1}\right)=0\ \ a.s.$$

By (a) and equation (3) and taking into account that $\{e_t\}$ are independent, we have

$$\begin{aligned}
& E\left(T^{-1/2}\sum_{t=2}^{T}e_t\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2 \\
\leq\ & T^{-1}\sum_{t=2}^{T}E\left(e_t^2\right)E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2 \\
\leq\ & \sup_t E\left(e_t^2\right)\sup_{2\leq t\leq T}E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2 < \infty
\end{aligned}$$

uniformly in $d_1\in[1-c-\varepsilon,1-c]$ for any $0<\varepsilon<1/2-c$. Then (c) follows directly from the Markov's inequality.

(d) Since $\left\{e_t\Delta^{d_1-1}\log\Delta e_{t-1}\right\}$ is a martingale difference sequence, we have by argument similar to (c)

$$\begin{aligned}
& E\left(T^{-1}\sum_{t=2}^{T}e_t\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2 \\
\leq\ & T^{-2}\sum_{t=2}^{T}\sup_t E\left(e_t^2\right)\sup_{2\leq t\leq T}E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2 \xrightarrow{p} 0.
\end{aligned}$$

(e) Note that $\left\{\left(\Delta^{d_1-1}e_{t-1}\right)^2 e_t\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right), \mathcal{F}_t\right\}$ is a martingale difference sequence. By Lemma 3(i), equation (3) and (a),

$$\begin{aligned}
& E\left(3T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^2 e_t\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)\right)^2 \\
= \; & 9T^{-2}\sum_{t=2}^{T}E\left(\Delta^{d_1-1}e_{t-1}\right)^4 E\left(e_t^2\right)E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2 \\
\leq \; & 9T^{-2}\sum_{t=1}^{T}\sup_{2\leq t\leq T}E\left(\Delta^{d_1-1}e_{t-1}\right)^4\sup_{t}E\left(e_t^2\right)\sup_{2\leq t\leq T}E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\xrightarrow{p}0
\end{aligned}$$

uniformly in $d_1\in[1-c-\varepsilon,1-c]$ for any $0<\varepsilon<1/2-c$. Thus, (e) follows from Markov inequality.

(f) By Lemma 3(i) and (a),

$$\begin{aligned}
& E\left|2T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)\right| \\
\leq \; & 2E\left(T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^2\right)^{1/2}\left(T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\right)^{1/2} \\
\leq \; & 2\sqrt{E\left(T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^2\right)E\left(T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\right)} \\
\leq \; & 2\sqrt{\left(\sup_{2\leq t\leq T}E\left(\Delta^{d_1-1}e_{t-1}\right)^2\right)\left(\sup_{2\leq t\leq T}E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\right)}<\infty
\end{aligned}$$

uniformly in $d_1\in[1-c-\varepsilon,1-c]$ for any $0<\varepsilon<1/2-c$. Thus (f) follows from Markov inequality.

(g) By Lemma 3(i) and (b),

$$\begin{aligned}
& E\left|4T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^3\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)\right| \\
= \; & E\left|4T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^2\left(\Delta^{d_1-1}e_{t-1}\right)\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)\right| \\
\leq \; & 4E\left(T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^4\right)^{1/2}\left(T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^2\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\right)^{1/2} \\
\leq \; & 4\sqrt{E\left(T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^4\right)E\left(T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}e_{t-1}\right)^2\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\right)} \\
= \; & 4\sqrt{\left(\sup_{2\leq t\leq T}E\left(\Delta^{d_1-1}e_{t-1}\right)^4\right)\left(\sup_{2\leq t\leq T}E\left(\Delta^{d_1-1}e_{t-1}\right)^2\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\right)}<\infty
\end{aligned}$$

uniformly in $d_1 \in [1-c-\varepsilon, 1-c]$ for any $0 < \varepsilon < 1/2 - c$. Thus (g) follows from Markov inequality.

(h) By argument similar to (f)

$$
\begin{aligned}
& E\left|2T^{-1}\sum_{t=2}^{T} e_t^2\left(\Delta^{d_1-1}e_{t-1}\right)\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)\right| \\
\leq\; & 2E\left(T^{-1}\sum_{t=2}^{T} e_t^4\left(\Delta^{d_1-1}e_{t-1}\right)^2\right)^{1/2}\left(T^{-1}\sum_{t=2}^{T}\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\right)^{1/2} \\
=\; & 2\sqrt{\left(T^{-1}\sum_{t=2}^{T} E\left(e_t^4\right)E\left(\Delta^{d_1-1}e_{t-1}\right)^2\right)\left(T^{-1}\sum_{t=2}^{T} E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\right)} \\
\leq\; & 2\sqrt{\left(\sup_t E\left(e_t^4\right)\sup_{2\leq t\leq T} E\left(\Delta^{d_1-1}e_{t-1}\right)^2\right)\left(\sup_{2\leq t\leq T} E\left(\Delta^{d_1-1}\log\Delta e_{t-1}\right)^2\right)} < \infty.
\end{aligned}
$$

Lemma 6 *Let u_t be a fractionally integrated process defined by $u_t = \Delta^{-\theta}v_t = \sum_{j=0}^{\infty}\pi_j(-\theta)v_{t-j}$, where $\theta \in [0, 1/2)$ and v_t is an i.i.d. process with $E(v_t) = 0$ and $E\left(v_t^2\right) = 1$. Let $\gamma_j^*(\theta) = E(u_t u_{t-j})$. Then*

$$
\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{i}\frac{1}{j}\sum_{k=i\vee j}^{\infty}\left|\pi_{k-i}(-\theta)\right|\left|\pi_{k-j}(-\theta)\right| \leq 2B\sum_{j=1}^{\infty}(\log j)\, j^{2\theta-2} + \gamma_0^*(\theta)\frac{\pi^2}{6} < \infty,
$$

where B is a positive constant.

Proof of Lemma 6

Notice that the upper bound of $\left|\gamma_{j-i}^*(\theta)\right|$ can be expressed as

$$
\begin{aligned}
\left|\gamma_{j-i}^*(\theta)\right| &= \left|E\left(u_{t-i}u_{t-j}\right)\right| \\
&= \left|\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\pi_k(-\theta)\pi_l(-\theta)E\left(v_{t-i-k}v_{t-j-l}\right)\right| \\
&\leq \sum_{k=i}^{\infty}\sum_{l=j}^{\infty}\left|\pi_{k-i}(-\theta)\right|\left|\pi_{l-j}(-\theta)\right|\left|E\left(v_{t-k}v_{t-l}\right)\right| \\
&= \sum_{k=i\vee j}^{\infty}\left|\pi_{k-i}(-\theta)\right|\left|\pi_{k-j}(-\theta)\right|. \qquad (85)
\end{aligned}
$$

The last equality follows from the fact that $E\left(v_t^2\right) = 1$.

To show Lemma 6, we use equation (85) and write

$$\begin{aligned}
& \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{i}\frac{1}{j}\sum_{k=i\vee j}^{\infty}\left|\pi_{k-i}\left(-\theta\right)\right|\left|\pi_{k-j}\left(-\theta\right)\right| \\
\leq & \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{i}\frac{1}{j}\left|\gamma_{j-i}^{*}\left(\theta\right)\right| \\
= & \sum_{i=1}^{\infty}\sum_{j=1-i}^{\infty}\frac{1}{i}\frac{1}{j+i}\left|\gamma_{j}^{*}\left(\theta\right)\right| \\
= & 2\sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty}\frac{1}{i}\frac{1}{j+i}\right)\left|\gamma_{j}^{*}\left(\theta\right)\right|+\sum_{k=1}^{\infty}\frac{1}{k^{2}}\gamma_{0}^{*}\left(\theta\right) \\
= & 2\sum_{j=1}^{\infty}\left(\frac{1}{j}\sum_{i=1}^{j}\frac{1}{i}\right)\left|\gamma_{j}^{*}\left(\theta\right)\right|+\gamma_{0}^{*}\left(\theta\right)\sum_{k=1}^{\infty}\frac{1}{k^{2}}.
\end{aligned}$$

The last equality is obtained by making use of the identity $\sum_{i=1}^{\infty}\frac{1}{i}\frac{1}{j+i}=\frac{1}{j}\sum_{i=1}^{j}\frac{1}{i}$. Now, since: (i) $\sum_{i=1}^{j}i^{-1}\leq\log j$; (ii) $\gamma_{j}^{*}\left(\theta\right)\leq Bj^{2\theta-1}$ for some constant B (Hosking (1981)); and (iii) $\sum_{k=1}^{\infty}k^{-2}=\pi^{2}/6$, it follows that

$$\begin{aligned}
\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{i}\frac{1}{j}\sum_{k=i\vee j}^{\infty}\left|\pi_{k-i}\left(-\theta\right)\right|\left|\pi_{k-j}\left(-\theta\right)\right| & \leq 2B\sum_{j=1}^{\infty}j^{-1}\left(\log j\right)j^{2\theta-1}+\gamma_{0}^{*}\left(\theta\right)\frac{\pi^{2}}{6} \\
& = 2B\sum_{j=1}^{\infty}\left(\log j\right)j^{2\theta-2}+\gamma_{0}^{*}\left(\theta\right)\frac{\pi^{2}}{6}<\infty.
\end{aligned}$$

The last inequality holds since the sequence $\left\{\left(\log j\right)j^{2\theta-2}\right\}_{j=1}^{\infty}$ is summable for $\theta<1/2$.Thus Lemma 6 is proved.

7 Appendix: Tables

Table 1: Single Variance Shift: Empirical Size

Volatility τ	δ	T	$d_1=0.6$ $t(d_1)$	$t_W(d_1)$	$d_1=0.7$ $t(d_1)$	$t_W(d_1)$	$d_1=0.8$ $t(d_1)$	$t_W(d_1)$	$d_1=0.9$ $t(d_1)$	$t_W(d_1)$	$d_1=0.95$ $t(d_1)$	$t_W(d_1)$
		250	5.68	5.98	5.66	5.99	5.52	5.62	4.84	5.06	4.84	5.21
	1	500	5.46	5.65	5.12	5.38	4.63	4.76	5.11	5.19	5.29	5.37
		1000	5.34	5.45	5.12	5.16	5.19	5.33	5.31	5.32	4.59	4.69
		250	6.45	5.89	6.33	5.85	6.08	5.46	5.85	5.34	5.90	5.30
0.1	5	500	6.15	5.54	6.17	5.46	5.59	5.03	5.86	5.19	5.78	5.12
		1000	6.13	5.24	5.96	5.04	6.19	5.45	6.26	5.43	5.53	4.77
		250	25.03	7.91	25.10	7.67	24.14	7.38	23.35	7.21	23.14	6.98
	0.2	500	24.64	6.70	24.34	6.61	24.26	6.21	24.27	6.58	24.63	6.47
		1000	25.28	6.32	25.24	5.93	24.41	5.76	24.35	5.86	23.81	5.69
		250	12.55	6.11	12.03	5.96	11.60	5.48	11.45	5.65	11.54	5.46
0.5	5	500	12.24	5.80	11.95	5.81	11.67	5.32	11.19	5.21	11.64	5.61
		1000	11.91	5.73	11.54	5.08	11.50	5.30	11.70	5.34	11.11	4.80
		250	12.55	6.25	11.64	5.86	11.81	5.90	11.48	5.50	11.30	5.50
	0.2	500	11.60	5.77	11.69	5.66	11.22	5.44	11.19	5.26	11.72	5.60
		1000	11.92	5.64	11.75	4.98	11.16	5.28	11.62	5.53	11.14	5.08
		250	24.92	7.95	25.46	7.58	24.26	7.58	22.77	7.02	23.29	6.90
0.9	5	500	25.55	6.60	25.46	6.86	25.14	6.34	24.05	6.37	23.79	6.08
		1000	24.81	5.99	25.11	6.20	24.74	5.86	25.00	5.68	22.96	5.01
		250	6.42	5.92	6.17	5.57	6.01	5.57	5.77	5.30	6.05	5.56
	0.2	500	6.31	5.65	6.00	5.33	5.35	4.74	5.98	5.14	6.08	5.38
		1000	6.23	5.44	5.92	5.24	5.86	5.11	5.99	5.31	5.78	4.96

Table 2: Double Variance Shifts: Empirical Size

Volatility			$d_1 = 0.6$		$d_1 = 0.7$		$d_1 = 0.8$		$d_1 = 0.9$		$d_1 = 0.95$	
τ	δ	T	$t(d_1)$	$t_W(d_1)$	$t(d_1)$	$t_W(d_1)$	$t(d_1)$	$t_W(d_1)$	$t(d_1)$	$t_W(d_1)$	$t(d_1)$	$t_W(d_1)$
		250	6.28	5.83	6.19	5.90	5.91	5.49	5.80	5.35	6.01	5.57
0.05	5	500	6.18	5.47	6.01	5.39	5.48	4.68	5.98	5.40	5.86	5.11
		1000	6.13	5.13	5.82	5.16	6.31	5.39	6.00	5.28	5.60	4.81
		250	24.14	7.51	24.66	7.55	23.44	7.37	23.12	7.16	23.55	7.52
	0.2	500	24.03	6.80	25.25	6.57	24.11	6.40	24.39	6.17	24.68	6.23
		1000	25.21	6.26	25.69	5.98	24.47	6.01	24.82	5.65	23.81	5.67
		250	24.13	7.95	23.57	7.18	23.55	7.34	23.37	7.32	23.53	7.27
0.45	5	500	24.54	6.68	24.32	6.41	24.45	6.05	23.86	6.21	24.49	6.46
		1000	24.45	6.08	24.73	5.84	24.81	5.86	24.59	5.48	23.61	5.60
		250	6.54	5.91	6.41	5.85	6.19	5.52	5.76	5.24	5.84	5.12
	0.2	500	6.46	5.88	6.20	5.45	5.55	4.87	6.00	5.30	6.09	5.40
		1000	5.97	5.29	6.26	5.30	5.90	5.20	6.00	5.31	5.80	5.04

Table 3: Trending Variances Model: Empirical Size

Volatility			$d_1 = 0.6$		$d_1 = 0.7$		$d_1 = 0.8$		$d_1 = 0.9$		$d_1 = 0.95$	
m	δ	T	$t(d_1)$	$t_W(d_1)$	$t(d_1)$	$t_W(d_1)$	$t(d_1)$	$t_W(d_1)$	$t(d_1)$	$t_W(d_1)$	$t(d_1)$	$t_W(d_1)$
		250	7.81	5.83	7.65	5.65	7.67	5.67	7.37	5.43	7.25	5.20
1	5	500	7.70	5.54	7.64	5.54	7.24	5.48	7.57	5.32	7.61	5.66
		1000	7.43	5.36	7.74	5.41	7.54	5.48	7.75	5.66	6.83	4.76
		250	8.04	6.39	7.30	5.59	7.55	5.88	7.29	5.45	7.40	5.50
	0.2	500	7.44	5.68	7.21	5.32	6.94	4.86	7.21	5.27	7.81	5.59
		1000	7.77	5.67	7.37	5.14	7.13	5.16	7.58	5.46	7.36	5.11
		250	10.56	5.98	10.22	5.85	10.12	6.08	9.64	5.48	9.61	5.43
2	5	500	10.14	5.63	10.80	5.73	9.74	5.67	10.14	5.32	10.59	5.52
		1000	10.40	5.28	10.29	5.49	10.04	5.36	10.65	5.55	9.39	4.60
		250	7.10	6.12	6.45	5.46	6.61	5.68	6.31	5.25	6.76	5.63
	0.2	500	6.63	5.68	6.49	5.37	6.05	4.79	6.42	5.26	6.73	5.53
		1000	6.88	5.51	6.65	5.25	6.41	5.19	6.78	5.34	6.60	5.19

Table 4: Constant Variance Model: Empirical Power

Test	$T \mid d_0$	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
	250	100.00	100.00	99.99	99.46	95.21	77.76	46.43	19.27
$t(d_1)$	500	100.00	100.00	100.00	100.00	99.85	96.63	73.14	29.91
	1000	100.00	100.00	100.00	100.00	100.00	99.96	94.38	47.40
	250	100.00	100.00	99.99	99.50	95.25	77.94	47.17	19.88
$t_W(d_1)$	500	100.00	100.00	100.00	100.00	99.84	96.57	73.41	30.16
	1000	100.00	100.00	100.00	100.00	100.00	99.96	94.53	47.75

Table 5: Single Variance Shift Model: Empirical Power of $t_W(d_1)$ test

τ	δ	$T \vert d_0$	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
		250	100.00	100.00	99.93	99.09	93.14	74.26	43.94	19.17
0.1	5	500	100.00	100.00	100.00	100.00	99.78	95.03	69.42	28.28
		1000	100.00	100.00	100.00	100.00	100.00	99.90	92.38	44.41
		250	91.95	83.61	71.83	59.03	44.78	31.46	20.83	12.40
	0.2	500	99.57	98.14	92.62	81.47	64.16	44.88	27.28	14.14
		1000	100.00	99.95	99.89	97.74	88.35	67.38	40.14	17.12
		250	100.00	99.80	98.24	91.84	77.22	54.64	31.43	14.84
0.5	5	500	100.00	100.00	99.99	99.73	96.17	80.40	50.62	20.69
		1000	100.00	100.00	100.00	100.00	99.93	97.54	76.18	30.59
		250	99.96	99.81	98.28	91.83	77.03	55.15	31.29	14.90
	0.2	500	100.00	100.00	100.00	99.68	96.18	81.40	49.81	20.61
		1000	100.00	100.00	100.00	100.00	99.99	97.43	77.14	31.40
		250	90.23	81.30	69.93	56.69	43.55	30.37	19.84	12.65
0.9	5	500	99.44	97.53	91.56	80.34	62.89	43.63	26.51	13.91
		1000	100.00	99.98	99.63	97.13	87.54	66.99	39.62	17.12
		250	100.00	100.00	99.97	99.10	93.22	75.15	44.34	19.04
	0.2	500	100.00	100.00	100.00	100.00	99.79	95.14	69.85	28.54
		1000	100.00	100.00	100.00	100.00	100.00	99.88	92.43	44.39

Table 6: Double Variance Shift Model: Empirical Power of $t_W(d_1)$ test

τ	δ	$T \vert d_0$	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
		250	100.00	100.00	99.94	99.02	93.37	75.05	44.24	19.36
0.05	5	500	100.00	100.00	100.00	100.00	99.80	95.18	69.58	28.58
		1000	100.00	100.00	100.00	100.00	100.00	99.91	92.46	44.76
		250	91.36	82.84	71.63	58.39	45.07	31.54	21.14	12.13
	0.2	500	99.47	97.67	92.32	81.22	63.78	44.19	27.39	13.79
		1000	100.00	99.98	99.78	97.23	88.80	65.78	39.72	17.00
		250	91.70	83.99	72.56	60.04	44.85	32.39	21.13	12.47
0.45	5	500	99.71	97.96	92.43	81.78	64.10	45.66	26.48	14.29
		1000	100.00	99.98	99.75	97.55	88.47	67.46	39.83	17.36
		250	100.00	100.00	99.90	99.02	93.20	74.36	43.75	19.00
	0.2	500	100.00	100.00	100.00	100.00	99.77	95.33	69.73	28.26
		1000	100.00	100.00	100.00	100.00	100.00	99.91	92.63	44.50

Table 7: Trending Variances Model: Empirical Power of of $t_W(d_1)$ test

m	δ	$T\|d_0$	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
		250	100.00	99.99	99.79	97.86	89.12	68.37	40.01	17.86
1	5	500	100.00	100.00	100.00	100.00	99.41	91.96	63.69	26.00
		1000	100.00	100.00	100.00	100.00	99.99	99.74	88.75	39.96
		250	100.00	100.00	99.76	97.99	89.37	69.64	40.37	17.52
	0.2	500	100.00	100.00	100.00	100.00	99.37	92.33	63.74	25.91
		1000	100.00	100.00	100.00	100.00	100.00	99.65	88.79	40.34
		250	100.00	99.92	98.96	94.76	81.73	59.86	34.72	15.98
2	5	500	100.00	100.00	100.00	99.84	97.70	84.96	54.76	22.67
		1000	100.00	100.00	100.00	100.00	99.97	98.71	81.06	34.29
		250	100.00	100.00	99.89	98.66	91.41	72.48	42.17	18.29
	0.2	500	100.00	100.00	100.00	100.00	99.62	94.18	66.82	27.30
		1000	100.00	100.00	100.00	100.00	100.00	99.80	90.85	42.54

Table 8: Constant Variance Model: Empirical Size and Power

Test	$T\|d_0$	1	0.95	0.9	0.85	0.8	0.75	0.7	0.65	0.6
	250	5.11	20.00	47.96	76.15	93.11	98.60	99.82	99.96	100.00
$t\left(\hat{d}_1\right)$	500	4.76	30.61	72.79	95.39	99.61	99.98	100.00	100.00	100.00
	1000	5.43	48.10	93.12	99.85	100.00	100.00	100.00	100.00	100.00
	250	5.39	20.37	48.54	76.47	93.08	98.61	99.81	99.96	100.00
$t_W\left(\hat{d}_1\right)$	500	4.90	30.94	72.80	95.45	99.62	99.98	100.00	100.00	100.00
	1000	5.46	48.38	93.13	99.86	100.00	100.00	100.00	100.00	100.00

Table 9: Single Variance Shift Model: Empirical Size and Power

			$t\left(\hat{d}_1\right)$	$t_W\left(\hat{d}_1\right)$								
τ	δ	$T\|d_0$	1	1	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
		250	6.06	5.63	19.57	46.03	73.40	91.07	97.99	99.65	99.97	100.00
0.1	5	500	5.86	5.18	29.19	69.25	93.81	99.50	99.96	100.00	100.00	100.00
		1000	6.39	5.64	45.58	91.18	99.79	100.00	100.00	100.00	100.00	100.00
		250	25.63	8.39	14.87	24.42	35.59	48.56	61.89	72.56	81.53	88.46
	0.2	500	25.42	6.39	16.10	30.52	48.51	65.15	80.40	90.25	95.67	98.39
		1000	24.46	6.01	19.44	42.09	68.30	86.94	96.02	99.14	99.81	99.99
		250	12.25	6.44	16.50	34.72	56.21	76.33	89.36	96.18	98.86	99.56
0.5	5	500	11.84	5.64	22.08	51.43	80.12	94.67	99.12	99.93	99.98	100.00
		1000	11.53	5.29	32.68	75.44	96.34	99.82	99.99	100.00	100.00	100.00
		250	12.00	6.32	16.14	34.51	56.95	76.59	89.47	96.35	98.64	99.59
	0.2	500	11.22	5.37	22.67	52.04	80.47	94.68	99.02	99.90	100.00	100.00
		1000	11.77	5.52	32.60	74.99	96.48	99.84	100.00	100.00	100.00	100.00
		250	25.64	8.43	14.60	23.50	34.56	46.99	60.75	71.85	80.47	87.21
0.9	5	500	24.92	6.75	16.24	29.37	47.37	64.94	79.45	89.64	94.79	97.83
		1000	24.70	6.35	18.96	41.31	66.70	86.26	95.80	98.94	99.79	99.96
		250	6.23	5.73	19.68	46.13	73.72	91.39	97.78	99.63	99.93	99.99
	0.2	500	5.59	4.95	28.86	69.89	93.89	99.36	99.96	100.00	100.00	100.00
		1000	6.10	5.40	45.48	91.23	99.70	100.00	100.00	100.00	100.00	100.00

Table 10: Double Variance Shifts Model: Empirical Size and Power

			$t\left(\hat{d}_1\right)$	$t_W\left(\hat{d}_1\right)$								
τ	δ	$T\|d_0$	1	1	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
		250	5.93	5.37	19.52	46.38	73.80	91.39	98.06	99.61	99.96	100.00
0.05	5	500	5.64	4.95	29.01	69.69	94.02	99.37	99.96	100.00	100.00	100.00
		1000	6.16	5.35	45.47	90.97	99.71	100.00	100.00	100.00	100.00	100.00
		250	25.10	8.04	14.91	23.45	34.64	48.30	61.27	72.01	81.02	88.32
	0.2	500	25.02	6.83	15.68	30.37	47.82	65.47	80.21	90.14	95.54	98.20
		1000	25.09	6.09	19.80	42.46	67.28	87.23	95.83	98.90	99.81	99.98
		250	25.24	8.34	14.45	24.04	36.09	49.24	62.07	72.75	82.17	89.12
0.45	5	500	25.01	6.58	15.64	30.64	47.70	66.10	80.30	90.35	95.54	98.26
		1000	25.73	5.79	19.26	42.06	67.94	86.61	95.95	98.87	99.83	99.99
		250	5.94	5.34	19.99	46.42	73.39	90.93	97.84	99.61	99.95	100.00
	0.2	500	6.09	5.50	28.99	69.77	93.80	99.45	99.95	100.00	100.00	100.00
		1000	6.10	5.39	45.43	91.06	99.77	100.00	100.00	100.00	100.00	100.00

Table 11: Trending Variances Model: Empirical Size and Power

			$t\left(\hat{d}_1\right)$	$t_W\left(\hat{d}_1\right)$								
m	δ	$T\|d_0$	1	1	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
		250	7.54	5.62	18.87	42.61	67.75	87.21	96.41	99.12	99.88	99.98
1	5	500	7.58	5.56	26.85	64.10	90.59	98.68	99.90	100.00	100.00	100.00
		1000	7.73	5.53	41.30	87.40	99.31	99.99	100.00	100.00	100.00	100.00
		250	8.04	6.02	18.78	42.61	69.02	87.45	96.06	99.26	99.77	99.99
	0.2	500	7.12	5.03	27.34	64.58	90.75	98.62	99.86	100.00	100.00	100.00
		1000	7.63	5.36	41.00	86.99	99.33	100.00	100.00	100.00	100.00	100.00
		250	10.45	5.84	17.63	37.36	60.31	80.45	92.60	97.58	99.36	99.81
2	5	500	10.57	5.83	23.59	55.88	84.03	96.47	99.56	99.95	99.99	100.00
		1000	10.39	5.57	35.68	79.98	97.85	99.92	99.99	100.00	100.00	100.00
		250	7.13	5.94	19.34	44.36	71.97	89.46	97.12	99.53	99.90	99.99
	0.2	500	6.27	4.74	28.47	67.35	92.73	99.17	99.90	100.00	100.00	100.00
		1000	6.70	5.54	43.60	89.40	99.56	100.00	100.00	100.00	100.00	100.00