DISCUSSION PAPER

2015/27

The effect of additive outliers on
a fractional unit root test

Hafner C. and A. Preminger

The effect of additive outliers on a fractional unit root test

Christian M. Hafner^{*1} and Arie Preminger²

¹*Université catholique de Louvain*

²*Ben Gurion University*

November 16, 2015

Abstract

It is well known that additive outliers that occur with a small probability have a bias effect on the asymptotic distribution of classical unit root statistics. This paper shows that such outliers do not affect the asymptotic distribution in the case where the error term is fractionally integrated of order d , where $0 < d < 1/2$, while there is a bias for the case $-1/2 < d < 0$. Convergence to the asymptotic distribution is slow, such that the bias effect of outliers may be important in finite samples, for which numerical evidence is provided. We then show that these results essentially do not change if the unknown d is replaced by a consistent estimator, which may have a slow rate of convergence. Such an estimator can be obtained by using first an outlier-correction procedure, and then estimate d for the outlier-corrected data. We finally apply our results to a realized volatility series of the S&P 500 for which we find evidence against the unit root hypothesis, as opposed to a procedure which neglects outliers.

Keywords: Fractional Brownian motion, additive outliers, long memory, unit root
JEL Classification Number: C12, C22.

^{*}Corresponding author, Institute of statistics, biostatistics and actuarial sciences, and CORE, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, email: christian.hafner@uclouvain.be

1 Introduction

The past two decades have witnessed an increasing interest in fractionally integrated processes as a convenient way of describing the long memory properties of many time series, see the original papers Granger and Joyeux (1980) and Hosking (1981), and the review papers Baillie (1996) and Robinson (2003). Fractionally integrated processes have numerous applications in financial econometrics, where, for example, the availability of high-frequency data allows to generate realized volatility series that are characterized by long memory, see e.g. Andersen et al. (2003).

Since the derivation of the asymptotic distribution of unit root statistics by Phillips (1987), many studies have analyzed extensions of the classical unit root framework. Sowell (1990) derives the asymptotic distribution of the unit root statistic for the case where the error term follows a fractionally integrated white noise process. He shows that for the long memory case where the fractional integration coefficient d is between 0 and $1/2$, the rate of convergence is N , where N is the sample size, while for $-1/2 < d < 0$, the rate is N^{1+2d} . His results were extended by Lubian (1999) and Lee and Shie (2004).

In this paper, we are interested in the behaviour of unit root statistics under long memory error terms if there happens to be additive outliers (AO's) in the levels which occur with some positive probability. These are events with large but temporary effects on the series. If they are sufficiently large or sufficiently frequent, they can dominate the remaining information contained in the series and bias unit root inference towards rejection of the unit root hypothesis, see Franses and Haldrup (1994), Shin et al. (1996), Vogelsang (1999), Perron and Rodriguez (2003), among others.

We show that additive outliers have no effect on the asymptotic distribution if the fractional integration coefficient is positive, $d > 0$, while they do enter the asymptotic distribution if $d < 0$. Our results generalize those of Franses and Haldrup (1994) and Sowell (1990). We provide numerical evidence for the convergence of statistics to the asymptotic distribution, which can be very slow. Also, the effect of outliers only disappears asymptotically (for $d > 0$), while in finite samples there may be a substantial bias in neglecting outliers when using classical unit root distribution theory.

In practice, a problem is that outliers can bias the estimate of the long memory parameter d and hence of the unit root statistic, see Tolvi (2003) and Haldrup and Nielsen (2007). Therefore, simply replacing the unknown parameter d by an estimate may lead to biased and inconsistent unit root statistics. We propose to deal with this problem by first detecting additive outliers using one of the tests that have been proposed in the

literature, and second, removing them from the observed series. Franses and Haldrup (1994) suggested applying Dickey–Fuller (1979) unit root tests by incorporating dummy variables in the autoregression chosen on the basis of the outlier detection procedure proposed by Chen and Liu (1993). We follow a similar approach using the outlier detection statistic of Perron and Rodriguez (2003), which is based on first differences of the data and has more power than the statistic of Vogelsang (1999) based on the levels. We show in simulations that the procedure provides asymptotically unbiased estimates of the long memory parameter and, hence, can be used consistently in the unit root statistics.

In an empirical illustration we apply the proposed procedure to a jump-robust realized volatility series of the S&P 500 index, for which there is evidence of long memory. Noting that the differenced series is anti-persistent ($d < 0$), our results imply that neglecting outliers such as the flash crash of May 6, 2010, may lead to biased unit root statistics. In our case, the unit root hypothesis is not rejected when outliers are ignored, while there is statistical evidence against the unit root after having corrected for outliers.

The remainder of the paper is organized as follows. The following section presents the model and the asymptotic theory. Section 3 provides a simulation study giving numerical evidence for the finite sample distributions, Section 4 presents the empirical illustration and Section 5 concludes. Proofs are delegated to an appendix.

2 The model

Consider the univariate process

$$z_t = z_{t-1} + u_t \tag{1}$$

where the error term u_t is a fractionally integrated $I(d)$ process. Thus, u_t can be written as

$$(1 - L)^d u_t = \varepsilon_t \tag{2}$$

where ε_t is an i.i.d. error term with mean zero and variance σ_ε^2 , d is the degree of differencing, and L denotes the lag operator. Since ε_t is assumed to be i.i.d., no nuisance parameter will enter the limiting distribution of our test statistics when there are no outliers. This assumption was also made by Sowell (1990). His results were extended by Lee and Shie (2004) who consider the case where ε_t can be a weakly dependent process.

If $d = 0$, then u_t is white noise, whereas if $d = 1$, u_t , is a random walk. However, as Granger and Joyeux (1980) and Hosking (1981) have shown, d need not be an integer.

From the binomial theorem, we have the relation

$$u_t = (1 - L)^{-d} \varepsilon_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \varepsilon_{t-j}$$

Stirling's approximation formula for the gamma function $\Gamma(\cdot)$ implies that $a_j \sim j^{d-1}$, hence the coefficient decays hyperbolically, at rate j^{d-1} , rather than at the exponential rate typical for short memory processes. A fractionally integrated series is covariance stationary and invertible if $-0.5 < d < 0.5$. In that case, the process z_t in (1) is difference stationary in the sense that $(1 - L)z_t = u_t$ is a stationary process.

Now suppose that the true process (1) is contaminated by additive outliers (AO) of magnitude $\pm\theta$, which may occur with a probability π . In this case, the observed time series is given by

$$y_t = z_t + \theta \delta_t \quad (3)$$

where δ_t is an i.i.d. random variable taking the value 1 or -1 with equal probability $\pi/2$, and is zero with probability $1 - \pi$. Note that an AO has a temporary effect on the series. However, if the AO are sufficiently large or sufficiently frequent, they can bias unit root inference towards rejection of the unit root hypothesis, i.e., the actual size is larger than the nominal size in the presence of AO.

In order to test for a unit root we estimate the following regression

$$\Delta y_t = \phi y_{t-1} + \eta_t,$$

where $\eta_t = u_t + \theta(1 - L)\delta_t$. The least squares estimate of ϕ is given by

$$\hat{\phi}_{AO} = \sum_{t=1}^N y_{t-1} \Delta y_t \bigg/ \sum_{t=2}^N y_{t-1}^2 \quad (4)$$

Under the null of unit root, $\phi = 0$, Franses and Haldrup (1994) show that for $d = 0$ the limiting distribution of $N\hat{\phi}$ shifts to the left in the presence of additive outliers. Thus, by using the usual critical values one tends to over-reject the null hypothesis in the presence of additive outliers. Their results were extended to models with a constant and linear trend by Yin and Maddala (1997). However, they obtain similar results in the sense that the limiting distributions of their test statistics are downward biased in the presence of outliers.

Before we continue, let us make the following notation and assumption. Define the variance of the partial sum of u_t by $\sigma_N^2 = E \left(\sum_{t=1}^N u_t \right)^2$. Let $W_d(r)$ be the normalized

fractional Brownian motion that is defined by the following stochastic integral,

$$W_d(r) = \frac{1}{\Gamma(1+d)\sqrt{K(d)}} \left(\int_0^r (r-x)^d dW(x) + \int_{-\infty}^0 [(r-d)^d - (-x)^d] dW(x) \right),$$

where $K(d) = \Gamma(1-2d)/[(1+2d)\Gamma(1+d)\Gamma(1-d)]$ and $W(\cdot)$ is the standard Brownian motion.

Assumption A. $u_t \sim I(d)$, $d \in (-0.5, 0.5)$, such that $\sigma_N^{-1} \sum_{t=1}^{[Tr]} u_t \Rightarrow W_d(r)$

Proposition 1 Under Assumption A, as $N \rightarrow \infty$,

i) If $d > 0$, then

$$N\hat{\phi}_{AO} \Rightarrow \frac{\frac{1}{2}W_d(1)^2}{\int_0^1 W_d(r)^2 dr}$$

ii) If $d < 0$, then

$$N^{1+2d}\hat{\phi}_{AO} \Rightarrow - \left[\frac{\pi\theta^2}{\sigma_\varepsilon^2 K(d)} + \frac{(0.5+d)\Gamma(1+d)}{\Gamma(1-d)} \right] \bigg/ \int_0^1 W_d(r)^2 dr$$

iii) If $d = 0$, then

$$N\hat{\phi}_{AO} \Rightarrow \frac{\frac{1}{2}[W(1)^2 - 1]}{\int_0^1 W(r)^2 dr} - \frac{\theta^2\pi}{\sigma_\varepsilon^2 \int_0^1 W(r)^2 dr},$$

where $\Rightarrow$ signifies weak convergence.

Remark 1. If $\theta = 0$, i.e. there are no outliers, our results are equivalent to those of Sowell (1990). In the presence of additive outliers, ϕ is still estimated consistently, where the estimator is super-consistent for $d \geq 0$.

Remark 2. If $d \leq 0$ and $\theta \neq 0$, we observe that the limiting distribution is downward biased. Hence the calculated test statistic will have an actual size in excess of the nominal size, and we will tend to reject the unit root hypothesis too often. Hence, the results of Franses and Haldrup (1994) for the case $d = 0$ generalize to those of $d \leq 0$. Moreover, note that only the product of π and $\left(\theta/\sqrt{K(d)}\sigma_\varepsilon\right)^2$ matters. Thus, large shocks with small probability have the same effect as that of small shocks with high chance of occurrence. Further note that a positive AO and a negative AO with the same magnitude have the same effect on the limiting distribution.

Remark 3. If $d > 0$ and $\theta \neq 0$, our results are still equivalent to those of Sowell(1990) and the limiting distribution does not depend on the nuisance parameters θ and π . Hence, we

obtain the important conclusion that in contrast to the short memory case, the fractional unit root test suggested by Sowell (1990) is robust to outliers when the long memory errors are persistent (i.e. $d > 0$). However, in small samples the test statistics will be biased, since the critical values are derived via simulations under the assumption that the data is not contaminated with outliers. As we show in Section 3, the finite sample bias converges to zero at a slow rate.

Remark 4. An alternative approach to testing unit roots is to consider the t statistic. Unfortunately, this statistic cannot be considered in our setup, since it only converges when $d = 0$ by Theorem 4 of Sowell (1990), see also Lubian (1999).

An intuitive explanation for the result of Proposition 1 is not obvious, but it may be related to the fact that AO is comparable to generating a negative MA(1) component, and having an $I(d)$ process with positive d dominates the MA process in terms of unit root statistics, while this is not the case for negative d .

Many economic time series appear to have deterministic components such as a constant or a linear trend. Hence, we generalize the above setting to the following models:

$$\bar{y}_t = \beta_0 + y_t, \quad \text{or} \quad (5)$$

$$\bar{y}_t = \beta_0 + \beta_1 t + y_t \quad (6)$$

where y_t is the stochastic trend model generated by (3). By eliminating it from the equations above, we obtain the following regression models

$$\Delta \bar{y}_t = \beta_0^* + \varphi^1 y_{t-1} + \eta_t \quad (7)$$

where $\beta_0^* = -\varphi^1 \beta_0$ or,

$$\Delta \bar{y}_t = \beta_0^* + \beta_1^* t + \varphi^2 y_{t-1} + \eta_t \quad (8)$$

where $\beta_0^* = -\varphi^2 \beta_0 + (1 + \varphi^2) \beta_1$, $\beta_1^* = -\varphi^2 \beta_1$ and $\eta_t = u_t + \theta(1 - L)\delta_t$ in both models.

Let $\hat{\varphi}_{AO}^1$ and $\hat{\varphi}_{AO}^2$ be the OLS estimators of φ^1 and φ^2 , respectively. In order to check for the presence of a unit root in both cases, we test the null hypothesis that $\varphi^1 = 0$ or $\varphi^2 = 0$ for model (5) or (6), respectively. The asymptotic distribution for the normalized coefficient estimators $\hat{\varphi}_{AO}^1$ or $\hat{\varphi}_{AO}^2$ involve functionals of fractional Brownian motion. Hence, to simplify the presentation of our results, we define for $d \in (-0.5, 0.5)$,

the following terms:

$$A(d) = 2 \left(3 \int_0^1 r W_d(r) dr - \int_0^1 W_d(r) dr \right) W_d(1) + 6 \left(\left(\int_0^1 W_d(r) dr \right) \left(\int_0^1 r W_d(r) dr \right) - 2 \left(\int_0^1 r W_d(r) dr \right)^2 \right) \quad (9)$$

$$B(d) = \int_0^1 W_d(r)^2 dr - 4 \left(\int_0^1 W_d(r) dr \right)^2 + 12 \left(\int_0^1 W_d(r) dr \int_0^1 r W_d(r) dr \right) - 12 \left(\int_0^1 r W_d(r) dr \right)^2 \quad (10)$$

$$C(d) = \int_0^1 W_d(r)^2 dr - \left(\int_0^1 W_d(r) dr \right)^2 \quad (11)$$

Note that if $d = 0$, then $W_d(r) = W(r)$, i.e. the standard Brownian motion.

Proposition 2 *Under Assumption A, as $N \rightarrow \infty$ we have for model (5):*

- i) If $d > 0$, then $N \hat{\varphi}_{AO}^1 \Rightarrow \frac{0.5 W_d(1)^2 - W_d(1) \int_0^1 W_d(r) dr}{C(d)}$
- ii) If $d < 0$, then $N^{1+2d} \hat{\varphi}_{AO}^1 \Rightarrow -C(d)^{-1} \left[\frac{\pi \theta^2}{\sigma_\varepsilon^2 K(d)} + \frac{(0.5+d)\Gamma(1+d)}{\Gamma(1-d)} \right]$
- iii) If $d = 0$, then $N \hat{\varphi}_{AO}^1 \Rightarrow \frac{0.5(W(1)^2 - 1) - W(1) \int_0^1 W(r) dr}{C(0)} - \frac{\pi \theta^2}{\sigma_\varepsilon^2 C(0)}$

and for model (6):

- iv) If $d > 0$, then $N \hat{\varphi}_{AO}^2 \Rightarrow \frac{0.5 W_d(1)^2 - A(d)}{B(d)}$
- v) If $d < 0$, then $N^{1+2d} \hat{\varphi}_{AO}^2 \Rightarrow -B(d)^{-1} \left[\frac{\pi \theta^2}{\sigma_\varepsilon^2 K(d)} + \frac{(0.5+d)\Gamma(1+d)}{\Gamma(1-d)} \right]$
- vi) If $d = 0$, then $N \hat{\varphi}_{AO}^2 \Rightarrow \frac{0.5(W(1)^2 - 1 - 2A(0))}{B(0)} - \frac{\pi \theta^2}{\sigma_\varepsilon^2 B(0)}$

Remark 5. The asymptotic distributions obtained under model (5) and (6) are more complicated compared to the ones appearing in the previous proposition. However, our basic conclusions remain unchanged. Thus, we find that in the presence of additive outliers, $\hat{\varphi}_{AO}^1$ and $\hat{\varphi}_{AO}^2$ are super-consistent for $d \geq 0$. Furthermore, for both models, when $d = 0$, our statistics are equivalent to the ones obtained by Yin and Maddala (1997) and their asymptotic distributions shift to the left in the presence of AO. Thus, by using the usual critical values, one would tend to over-reject the null hypothesis of a unit root in the presence of AO. Similar results hold also for the case that $d < 0$ since the statistic depends

on the nuisance parameters θ and π . For $d > 0$, however, the asymptotic distributions do not depend on these nuisance parameters. Thus, in the presence of constant or linear components, we continue to obtain the important conclusion that fractional unit root testing is robust to outliers when the long memory errors are persistent.

The following proposition establishes, under weak conditions, that the asymptotic distribution of the unit root statistic is the same whether the parameter d is estimated or known.

Proposition 3 *For the model in (3), let $\hat{d}_N$ be an estimate for d such that $\hat{d}_N = d + o_P((\log N)^{-1})$. Then $N^{1+2\hat{d}_N} \hat{\phi}_{AO} = N^{1+2d} \hat{\phi}_{AO} + o_P(1)$.*

An extension of Proposition 3 to allow for deterministic trends as in (5) or (6) is straightforward. Note that there exist several $\sqrt{N}$ -consistent estimators of d . For example, the maximum likelihood estimator due to Sowell (1992) and Chung and Baillie (1993), or the semiparametric estimator of Geweke and Porter-Hudak (1983). By using the result of Proposition 3 we can obtain consistent unit root tests when the long memory parameter is unknown. It was shown in the literature that the presence of outliers can bias estimates of d , see Haldrup and Nielsen (2007). A possible solution would be to perform an outlier detection procedure which removes the additive outliers from the data. We are going to use such a procedure in the simulation study of the following section.

3 Numerical evidence

We present numerical evidence for the convergence of $N\hat{\phi}$ (for $d \geq 0$) and $N^{1+2d}\hat{\phi}$ (for $d < 0$) in Table 1. The process y_t in (3) is simulated using standard Gaussian white noise ε_t and outliers $\theta \in \{0, 1, 2\}$. Hence, the outliers correspond to, respectively, one and two standard deviations of the error term. We use a sample size N ranging from 100 to 100,000, and the statistic $\hat{\phi}$ is calculated. This is repeated K times, so that we obtain a simulated distribution of $N\hat{\phi}$ and $N^{1+2d}\hat{\phi}$, respectively, and empirical quantile q_α with $\alpha = 0.05$. For computational reasons, K is chosen as 100,000 for $N = 100, 500$ and 2000, and $K = 1000$ for $N = 10,000$ and $N = 100,000$. To obtain additional information about the uncertainty of these estimates, we repeat this exercise M times, where $M = 100$ for $N \in \{100, 500\}$, $M = 10$ for $N \in \{2000, 10000\}$, and $M = 5$ for $N = 100,000$, and then obtain quantile estimates $q_{\alpha,m}$, $m = 1, \dots, M$. Then, the final quantile estimator

is $q_\alpha = M^{-1} \sum_{m=1}^M q_{\alpha,m}$, and standard errors are calculated as $S/\sqrt{M}$, where S is the standard deviation of $q_{\alpha,m}$.

N	$\theta = 0$	$\theta = 1$	$\theta = 2$		
$d = 0$					
	$q_{0.05}$	$q_{0.05}$	5% size	$q_{0.05}$	5% size
100	-7.94 (0.005)	-8.83 (0.040)	0.063	-10.99 (0.130)	0.104
500	-8.02 (0.005)	-8.89 (0.020)	0.063	-11.40 (0.060)	0.102
2000	-8.04 (0.004)	-8.91 (0.010)	0.063	-11.51 (0.030)	0.103
$d = 0.3$					
100	-1.32 (0.001)	-1.46 (0.007)	0.057	-1.97 (0.030)	0.081
500	-0.41 (0.001)	-0.46 (0.005)	0.058	-0.62 (0.010)	0.076
2000	-0.24 (0.001)	-0.27 (0.001)	0.056	-0.35 (0.004)	0.077
10000	-0.24 (0.001)	-0.25 (0.001)	0.052	-0.31 (0.004)	0.071
100000	-0.23 (0.002)	-0.24 (0.002)	0.051	-0.24 (0.003)	0.051
$d = -0.3$					
100	-1.86 (0.001)	-1.98 (0.005)	0.067	-2.34 (0.020)	0.117
500	-1.70 (0.001)	-1.83 (0.006)	0.070	-2.25 (0.032)	0.158
2000	-1.58 (0.001)	-1.71 (0.005)	0.073	-2.05 (0.007)	0.143
10000	-1.12 (0.002)	-1.22 (0.003)	0.060	-1.52 (0.006)	0.110
100000	-0.37 (0.005)	-0.40 (0.007)	0.061	-0.46 (0.011)	0.114

Table 1: 5% quantiles of the simulated distribution of $N\hat{\phi}$ (for $d \geq 0$) and $N^{1+2d}\hat{\phi}$ (for $d < 0$), $\pi = 0.05$ and alternative outlier sizes θ . Standard errors are in parentheses. The columns "5% size" give the actual size of a test at nominal level 5% assuming absence of outliers.

The results for $d = 0$ confirm those of Franses and Haldrup (1994). For $d \neq 0$, convergence to the limiting distribution is very slow. This can be seen by the fact that if, for example, $d = 0.3$, then the limiting distribution has non-negative support, while the 5% quantile of the distribution with $N = 100,000$ is still negative. However, one clearly observes a convergence of the distribution with outliers to the one without outliers, as predicted by the theory. The actual size of the tests assuming absence of outliers converges to the nominal size.

On the other hand, if $d = -0.3$, there is no convergence: For example, the estimated

actual size for $\theta = 2$ and $N = 100,000$ is 11.4 % at a nominal size of 5%, confirming our results.

In practice, the fractional integration coefficient d is not known. Estimation of d in the presence of various types of outliers has been shown to be biased by Haldrup and Nielsen (2007). However, adopting an outlier detection and correction strategy, it is possible to obtain unbiased estimates of d and an outlier corrected series, following the approach of Perron and Rodriguez (2003). The idea is then to apply the above testing procedure to the outlier-corrected series.

To be precise, we use the τ_d statistic of Perron and Rodriguez (2003), based on first differences of y_t , to detect the presence of outliers. In order to perform statistical inference, they assumed that Δy_t is normally distributed and require that the variance of their test statistic can be consistently estimated from the data. These assumptions are common practice for statistical inference in dummy variable regressions, see Hendry and Santos (2005). In our setting, in order to apply their test procedure, we first need to assume the normality of Δy_t . Second, we note that the first difference of y_t is stationary with finite long-run variance. Hence, by the ergodic theorem the variance of the test statistic can be consistently estimated via its sample counterpart. Let $1_{(\cdot)}$ denote the indicator function and consider the regression

$$\Delta y_t = \delta \{1_{(t=T)} - 1_{(t-1=T)}\} + v_t$$

with least squares estimator $\hat{\delta} = (\Delta y_t - \Delta y_{t+1})/2$. The statistic for a fixed time point $1 < T < N$ is given by $\tau_d = \sup_T |t_{\hat{\delta}}(T)|$, where

$$t_{\hat{\delta}}(T) = \frac{\hat{\delta}}{\sqrt{(R_v(0) - R_v(1))/2}}$$

and $R_v(j)$ is the auto-covariance of v_t of order j . Since $v_t = \Delta y_t$ for all $t \neq T-1, T$ and under our assumptions, the first difference of y_t is stationary with finite long-run variance. Hence, by the ergodic theorem, a consistent estimator of $R_v(j)$ is obtained by $\hat{R}_v(j) = 1/N \sum_{t=1}^{N-j} \Delta y_t \Delta y_{t+j}$. In order to obtain critical values for their test statistic Perron and Rodriguez (2003) assumed that Δy_t is normally distributed, which also has to be assumed in this case.

We choose the critical values based on simulated data for $T = 100, 200, 500$ and 2000 , corresponding to a 5% significance level under normality, see Table 2. For $T = 100$ and 200 these values are the same as in Table 4 of Perron and Rodriguez (2003). The table shows the size and power estimates of the test for simulated $I(d)$ processes with

T	crit	$d = 0.3$			$d = -0.3$		
		$\theta = 0$	$\theta = 1$	$\theta = 2$	$\theta = 0$	$\theta = 1$	$\theta = 2$
100	3.65	0.208	0.241	0.485	0.054	0.102	0.233
200	3.75	0.170	0.237	0.619	0.052	0.068	0.249
500	3.92	0.103	0.204	0.669	0.054	0.078	0.453
2000	4.20	0.064	0.220	0.861	0.052	0.107	0.606

Table 2: *Simulated critical values (crit), size ($\theta = 0$) and power ($\theta = 1$ and $\theta = 2$) estimates of the Perron-Rodriguez outlier test using 100,000 Monte Carlo simulated $I(d)$ processes of length T at nominal level 5%.*

$d = 0.3$ and $d = -0.3$. Size estimates are obtained for $\theta = 0$, i.e. absence of outliers, and power estimates for two outlier sizes, $\theta = 1$ and $\theta = 2$, occurring with probability $\pi = 0.05$. In small samples, the test is size-distorted for persistent processes, but not for anti-persistent ones, which is analogous to the results of Perron and Rodriguez (2003) for MA(1) processes with positive and negative coefficients. The empirical size converges to the nominal size as the sample size increases. The test appears to have good power also for fractionally integrated processes.

Next, detected and estimated outliers are removed from the observed value of y_t , and the unit root statistic is applied to the corrected series. The fractional difference parameter d is estimated using conditional maximum likelihood as in Haldrup and Nielsen (2007), which among the estimators of their study is the most robust with respect to the presence of outliers. Table 3 shows the estimation results for d both for the original and the outlier-corrected series. Clearly, estimates of d are negatively biased for the series containing outliers, whereas biases are much smaller and insignificant for the outlier-corrected series, at least in larger samples.

We finally apply the above unit root testing procedure to the outlier-corrected series, again using $K = 10000$ replications but $M = 1$, so that the quantile estimates are reported without standard deviation. Results are shown in Table 4. Apart from some small deviations from the nominal level for $N = 100$, the empirical level converges quickly to the nominal level. Deviations are slightly larger for negative than for positive d . Comparing the results of Table 4 with those of Table 1, it is obvious that the outlier removal has delivered unit root statistics that are again reliable when judged by critical values of a distribution assuming absence of outliers.

N	$\theta = 0$		$\theta = 1$		$\theta = 2$	
$d = 0$						
	NC	C	NC	C	NC	C
100	-0.01 (0.08)	0.00 (0.08)	-0.06 (0.08)	-0.04 (0.08)	-0.08 (0.08)	-0.04 (0.08)
500	0.00 (0.03)	0.00 (0.03)	-0.04 (0.05)	-0.02 (0.05)	-0.07 (0.03)	-0.04 (0.03)
2000	0.00 (0.01)	0.00 (0.01)	-0.03 (0.04)	-0.01 (0.04)	-0.10 (0.02)	-0.03 (0.02)
$d = 0.3$						
100	0.29 (0.08)	0.29 (0.08)	0.25 (0.09)	0.28 (0.09)	0.17 (0.08)	0.22 (0.07)
500	0.30 (0.03)	0.30 (0.03)	0.24 (0.03)	0.29 (0.03)	0.17 (0.03)	0.25 (0.03)
2000	0.30 (0.01)	0.30 (0.01)	0.25 (0.01)	0.29 (0.01)	0.16 (0.02)	0.27 (0.02)
$d = -0.3$						
100	-0.28 (0.08)	-0.29 (0.08)	-0.34 (0.08)	-0.32 (0.08)	-0.38 (0.08)	-0.33 (0.08)
500	-0.29 (0.03)	-0.30 (0.03)	-0.35 (0.03)	-0.31 (0.03)	-0.39 (0.03)	-0.32 (0.03)
2000	-0.30 (0.01)	-0.30 (0.01)	-0.34 (0.02)	-0.30 (0.01)	-0.37 (0.02)	-0.32 (0.02)

Table 3: *Mean of estimated d over $K = 10000$ replications, alternative outlier sizes θ and sample sizes N . Standard deviations are in parentheses. Columns with "C" correspond to estimates with a prior outlier correction using the procedure of Perron and Rodriguez (2003). Columns with "NC" do not have an outlier correction.*

4 Empirical application

The statistical analysis of realized volatility series has led to an important branch of financial econometrics, see the seminal papers of Andersen et al. (2001, 2003) and the recent handbook of Bauwens et al. (2012). A common finding is the presence of long memory in daily realized volatility series, and modelling and forecasting using fractionally integrated models. The effect of outliers on models for realized volatility does not seem to have attracted a lot of attention. There is, however, related work on the effect of measurement errors in the sense of realized volatility series affected by microstructure noise. For example, Hansen and Lunde (2014) investigate unit root tests in this framework, and Rossi and Santucci de Magistris (2014) find evidence for non-stationary fractionally integrated realized volatility after correcting for the bias due to measurement error.

To illustrate the procedure suggested in the preceding sections, we apply the test to a median truncated realized volatility sequence of the S&P 500 from January 3, 2000 to

N	$\theta = 0$	$\theta = 1$	$\theta = 2$		
$d = 0$					
	$q_{0.05}$	$q_{0.05}$	5% size	$q_{0.05}$	5% size
100	-7.94	-8.52	0.060	-9.15	0.071
500	-8.02	-8.44	0.058	-8.78	0.066
2000	-8.04	-8.37	0.055	-8.45	0.057
$d = 0.3$					
100	-1.32	-1.40	0.055	-1.77	0.065
500	-0.41	-0.42	0.053	-0.50	0.062
2000	-0.24	-0.25	0.052	-0.27	0.056
$d = -0.3$					
100	-1.86	-1.89	0.061	-2.01	0.078
500	-1.70	-1.75	0.057	-1.88	0.071
2000	-1.58	-1.61	0.052	-1.69	0.058

Table 4: 5% quantiles of the simulated distribution of $N\hat{\phi}$ (for $d = 0.3$) and $N^{1+2\hat{d}}\hat{\phi}$ (for $d = -0.3$), applied to the outlier corrected series. Other comments as in Table 1.

February 4, 2014¹. Similar to bi-power variations, the median truncated realized volatility is defined by Andersen, Dobrev and Schaumburg (2010) as a realized volatility measure robust to jumps. The total number of observations is $N = 3515$. Figure 1 depicts the logarithm of the series, while Figure 2 plots the ACF of the levels and first differences.

Without outlier correction, the estimator of d is -0.44 and that of ϕ is -0.13, leading to a non-corrected statistic $N^{1+2\hat{d}}\hat{\phi}$ of -0.346. For the estimated values of d and ϕ , the number of observations $N = 3515$, and assuming absence of outliers ($\theta = 0$), we conducted an additional simulation exercise to find the critical values of the distribution of $N^{1+2\hat{d}}\hat{\phi}$, using 10,000 replications. The 5% and 10 % critical values were calculated as -0.625 and -0.532, respectively. Hence, assuming absence of outliers, one would not reject the null hypothesis of a unit root. This appears to match with the empirical results of Shephard and Sheppard (2010), whose "integrated HEAVY models" for various realized volatility type measures seems to work very well in forecasting. Note however that the data generating process of an integrated HEAVY process, similar to integrated GARCH, may be strictly stationary.

¹The data were downloaded from the website of the Oxford-Man Institute of Quantitative Finance: <http://realized.oxford-man.ox.ac.uk/>.

Applying the outlier detection procedure outlined above at a five percent level, seven outliers are identified: 2002/01/28, 2003/11/28, 2007/11/23, 2008/10/13, 2010/05/06, 2012/12/24, and 2013/09/17. There are both positive and negative outliers. The positive May 6, 2010 outlier is the so-called flash crash that had a very strong impact on realized volatility measures. Another positive outlier is the strong 11.6% recovery in the S&P 500 of October 13, 2008, after the Lehman crisis of September 2008. November 23, 2007, was a negative outlier with unusually low volatility.

After outlier correction, the estimated d drops to -0.40 and the estimated ϕ decreases slightly to -0.12. We use the d -differencing test of Shimotsu (2006) to test whether the outlier-corrected series is not affected by structural breaks. After taking d -differences, the residual series should be $I(0)$ if long memory was genuine, but not necessarily so if it was spurious. We then apply the KPSS test to the d -differenced residuals, which has power against structural breaks. The statistic takes the value 0.427, which is smaller than the critical 5% value of 0.463, see Table 1 of Kwiatkowski et al (1992). Hence, we do not reject the null hypothesis of $I(0)$ and we do not have evidence in favor of structural breaks.

The statistic $N^{1+2\hat{d}}\hat{\phi}$ now takes the value -0.61, very close to the 5% critical value. At a 10 % level, one would reject the null hypothesis of a unit root for the outlier corrected series.

We conclude that, for the analyzed realized volatility series, after having accounted and corrected for outliers, we have evidence against the null hypothesis of a unit root and, hence, this volatility sequence would be mean-reverting and stationary, which is compatible with the broad view in the literature on financial time series.

To investigate the robustness of our results with respect to the particular realized volatility measure, we did the same analysis for two other measures: the 5-minute realized variance and the 5-minute bipower variation, both using 1-minute subsampling. We obtain as estimates for d : -0.470 and -0.458, and for ϕ : -0.151 and -0.143, giving statistics $N^{1+2\hat{d}}\hat{\phi}$ of -0.246 and -0.285. After outlier correction as above, we obtain statistics of -0.362 and -0.450. In both cases, the outlier corrected statistics are beyond the 10% critical value of the corresponding simulated distributions. In the case of bipower variation, it is also beyond the 5% critical value. Hence, we have similar evidence against the null hypothesis of a unit root as for the median truncated realized volatility.

5 Conclusions

We have shown that additive outliers in a model with fractionally integrated error terms do not affect the asymptotic distribution of classical unit root statistics in the case of persistence ($d > 0$), but that there is a bias in the case of anti-persistence ($d < 0$). Convergence to the asymptotic distribution is typically very slow, so that there is likely to be a bias in realistic sample sizes even in the persistence case. Our simulation study demonstrates this effect. The empirical example of a realized volatility series has shown that correcting for outliers may lead to statistical evidence against the unit root hypothesis, where otherwise there would not be such evidence.

The results could be extended further in various ways. For example, the additive outlier model can be considered as a special case of a more general class of outlier models with temporary effect, see Chen and Liu (1993). In this case the contamination part can be written as $\theta/(1 - \alpha L)$ with $\alpha \in [0, 1)$. We focus on the situation where $\alpha = 0$. For other values of α the outlier effects will gradually die out. It can be shown that our results can be adjusted to this type of outliers, replacing the term $\theta^2\pi$ in Proposition 1 by $\theta^2\pi/(1 - \alpha^2)$.

Appendix

We assume that $y_0 = O_p(1)$ and $\omega_t = \theta\delta_t$. Let $S_N(r) = \sigma_N^{-1} \sum_{t=1}^{\lfloor rN \rfloor} u_t$, where $\lfloor rN \rfloor$ denotes the largest integer part that is less than or equal to rN , where $r \in [0, 1]$. Define $x_t = [1, t,]'$, and let $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1]$ and $e = [e_1, e_2, \dots, e_N]$ be the OLS coefficient estimates and the residuals, respectively, from running the regression of y_t on x_t . Further, let $A(d)$, $B(d)$ and $C(d)$ for $d \in (-0.5, 0.5)$ be defined as in equations (9), (10) and (11), respectively.

Lemma 1 *Under Assumption A, we have the following results:*

- a) $\sigma_N^2/N^{1+2d} = \sigma_\varepsilon^2 K(d)$
- b) $\sigma_N^{-1} y_N \Rightarrow W_d(1)$, $\sigma_N^{-2} y_N^2 \Rightarrow W_d(1)^2$
- c) $\frac{1}{N} \sum_{t=1}^N \omega_t^2 \rightarrow_{a.s.} \pi\theta^2$, $\frac{1}{N} \sum_{t=1}^N u_t^2 \rightarrow_{a.s.} \sigma_u^2$
- d) $\sigma_N^{-2} N^{-1} \sum_{t=1}^N y_t^2 \Rightarrow \int_0^1 W_d(r)^2 dr$, $\sigma_N^{-2} N^{-1} \sum_{t=1}^N t y_t \Rightarrow \int_0^1 r W_d(r) dr$, and $\sigma_N^{-1} \bar{y}_N \Rightarrow \int_0^1 W_d(r) dr$, where $\bar{y}_N = N^{-1} \sum_{t=1}^N y_t$

- e) $\sigma_N^{-2} \sum_{t=1}^N y_{t-1} \Delta y_t \Rightarrow \frac{1}{2} W_d(1)^2$, when $d > 0$
- f) $N^{2d} \sigma_N^{-2} \sum_{t=1}^N y_{t-1} \Delta y_t \rightarrow_p -\frac{\pi \theta^2}{\sigma_\varepsilon^2 K(d)} - \frac{(0.5+d)\Gamma(1+d)}{\Gamma(1-d)}$, when $d < 0$
- g) $\sum_{t=1}^N \sigma_\varepsilon^{-2} y_{t-1} \Delta y_t \Rightarrow 0.5(W(1)^2 - 1) - \sigma_\varepsilon^{-2} \pi \theta^2$, when $d = 0$

Proof of Lemma 1:

- a) The result follows from Theorem 1 of Sowell (1990).
- b) Note that, $\Delta y_t = u_t + \theta(1 - L)\delta_t = u_t + \omega_t - \omega_{t-1}$. Summing over t we get by Assumption A, the continuous mapping theorem and using that $\omega_t = O_p(1)$ and $y_0 = O_p(1)$,

$$\sigma_N^{-1} y_N = \sigma_N^{-1} \sum_{t=1}^N u_t + \sigma_N^{-1} (\eta_N + y_0) = S_N(1) + o_p(1) \Rightarrow W_d(1)$$

A similar argument can be used to show the second result as well.

- c) By assumption on the outlier δ_t , $E(\omega_t^2) = \pi \theta^2$ and the first result follows by the ergodic theorem. Further, since $-0.5 < d < 0.5$, $\{u_t\}$ is a stationary sequence with finite variance for which we can apply the ergodic theorem to obtain the second result.
- d) Assumption A and the continuous mapping theorem imply that

$$\begin{aligned} \sigma_N^{-2} N^{-1} \sum_{t=1}^N y_{t-1}^2 &= \sum_{t=1}^N \int_{(t-1)/N}^{t/N} (S_N(r))^2 dr = \int_0^1 S_N^2(r) dr \Rightarrow \int_0^1 W_d(r)^2 dr \\ \sigma_N^{-1} N^{-2} \sum_{t=1}^N t y_t &= \sum_{t=1}^N \int_{(t-1)/N}^{t/N} \left(\frac{t}{N}\right) S_N(r) dr = \int_0^1 r S_N(r) dr \Rightarrow \int_0^1 r W_d(r) dr \\ \sigma_N^{-1} \bar{y}_N &= N^{-1} \sum_{t=1}^N \sigma_N^{-1} y_t = \sum_{t=1}^N \int_{(t-1)/N}^{t/N} S_N(r) dr \Rightarrow \int_0^1 W_d(r) dr \end{aligned}$$

- e) Note further that

$$y_t^2 = (y_{t-1} + \Delta y_t)^2 = (y_{t-1} + \eta_t)^2 = y_{t-1}^2 + 2y_{t-1}\eta_t + \eta_t^2 = y_{t-1}^2 + 2y_{t-1}\eta_t + (u_t + \omega_t - \omega_{t-1})^2$$

implying that

$$y_{t-1}\eta_t = \frac{1}{2}(y_t^2 - y_{t-1}^2) - \frac{1}{2}(u_t + \omega_t - \omega_{t-1})^2.$$

Hence,

$$\begin{aligned}
\sigma_N^{-2} \sum_{t=1}^N y_{t-1} \eta_t &= \frac{1}{2} \sigma_N^{-2} (y_N^2 - y_0^2) - \frac{1}{2} N \sigma_N^{-2} \left\{ \frac{1}{N} \sum_{t=1}^N (u_t^2 + \omega_t^2) + \frac{N-1}{N} \right. \\
&\quad \times \left[\frac{1}{N-1} \sum_{t=1}^N \omega_{t-1}^2 - \frac{2}{N-1} \sum_{t=1}^N \omega_t \omega_{t-1} + \frac{2}{N-1} \sum_{t=1}^N u_t (\omega_t - \omega_{t-1}) \right] \Big\} \\
&= \frac{1}{2} \sigma_N^{-2} y_N^2 - N \sigma_N^{-2} A_N + o_p(1)
\end{aligned}$$

Since $E(u_t(\omega_t - \omega_{t-1})) = E(u_t)E(\omega_t - \omega_{t-1}) = 0$ and $E(\omega_t \omega_{t-1}) = 0$, we can apply the ergodic theorem to the terms inside the curly parentheses, to get by Lemma 1(a) and 1(c) that

$$A_N \rightarrow_{a.s.} \pi \theta^2 + \frac{1}{2} \sigma_u^2$$

and $N \sigma_N^{-2} = o(1)$ if $d > 0$, so that $\sigma_N^{-2} \sum_{t=1}^N y_{t-1} \eta_t = \frac{1}{2} \sigma_N^{-2} y_N^2 + o_p(1)$, and the desired result follows from Lemma 1(b).

- f) We have $\sigma_u^2 = \Gamma(1 - 2d) \sigma_\varepsilon^2 / \Gamma(1 - d)^2$ see Sowell (1990, p.496) and by Lemmas 1(a) and 1(d) we get for $d < 0$

$$\begin{aligned}
\frac{N^{2d}}{\sigma_N^2} \sum_{t=1}^N y_{t-1} \eta_t &= \frac{1}{2} N^{2d} (\sigma_N^{-2} y_N^2) - N^{1+2d} \sigma_N^{-2} A_N + o_p(N^{2d}) \\
&= -\sigma_\varepsilon^{-2} [\pi \theta^2 + \frac{1}{2} \sigma_u^2] / K(d) + o_p(1) \\
&= \frac{-\sigma_\varepsilon^{-2} (1 + 2d) \Gamma(1 + d) \Gamma(1 - d)}{\Gamma(1 - 2d)} \left[\pi \theta^2 + \frac{\Gamma(1 - 2d) \sigma_\varepsilon^2}{2 \Gamma(1 - d)^2} \right] + o_p(1) \\
&= -\frac{\pi \theta^2}{\sigma_\varepsilon^2 K(d)} - \frac{(0.5 + d) \Gamma(1 + d)}{\Gamma(1 - d)} + o_p(1)
\end{aligned}$$

- g) The result for $d = 0$ follows directly from Franses and Haldrup (1994, p.472)

Lemma 2 *Under Assumption A the following holds:*

- a) $N^{-1} \sigma_N^{-2} \sum_{t=1}^N y_{t-1} (y_{t-1} - \bar{y}_N) \Rightarrow C(d)$
- b) $\sigma_N^{-1} \begin{bmatrix} \hat{\beta}_0 \\ N \hat{\beta}_1 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 \int_0^1 W_d(r) dr - 6 \int_0^1 r W_d(r) dr \\ -6 \int_0^1 W_d(r) dr + 12 \int_0^1 r W_d(r) dr \end{bmatrix}$
- c) $\sigma_N^{-2} B_N \Rightarrow A(d)$, where $B_N = \hat{\beta}_0 \sum_{t=1}^N \Delta y_t + N \hat{\beta}_1 \sum_{t=1}^N \left(\frac{t}{N}\right) \Delta y_t$
- d) $N^{-1} \sigma_N^{-2} \sum_{t=1}^N e_{t-1}^2 \Rightarrow B(d)$

Proof of Lemma 2:

We use the notation $X_N = [x_1, x_2, \dots, x_N]'$ and $Y_N = [y_1, y_2, \dots, y_N]'$.

a) By Lemma 1(b) and 1(d), we get

$$\sigma_N^{-2} N^{-1} \sum_{t=1}^N y_{t-1} (y_{t-1} - \bar{y}_N) = N^{-1} \sigma_N^{-2} \sum_{t=1}^N y_{t-1}^2 - (\sigma_N^{-1} \bar{y}_N)^2 \Rightarrow C(d)$$

b) Let $\Upsilon_N = \text{diag}(N^{-1}, N^{-2})$. First note that

$$\lim_{N \rightarrow \infty} \Upsilon_N X_N' X_N \Upsilon_N N = \lim_{N \rightarrow \infty} N \Upsilon_N \begin{bmatrix} N & \sum_{t=1}^N t \\ \sum_{t=1}^N t & \sum_{t=1}^N t^2 \end{bmatrix} \Upsilon_N = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}$$

Second, by Lemma 1(d) ,

$$\sigma_N^{-1} \Upsilon_N X_N Y_N = \begin{bmatrix} N^{-1} \sigma_N^{-1} \sum_{t=1}^N y_t \\ N^{-2} \sigma_N^{-1} \sum_{t=1}^N t y_t \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^N \int_{(t-1)/N}^N S_N(r) \\ \sum_{t=1}^N \int_{(t-1)/N}^N r S_N(r) \end{bmatrix} \Rightarrow \begin{bmatrix} \int_0^1 W_d(r) dr \\ \int_0^1 r W_d(r) dr \end{bmatrix}$$

Third, by definition of $\hat{\beta}$, we have that $(\Upsilon_N X_N' X_N \Upsilon_N N)(N^{-1} \Upsilon_N^{-1} \sigma_N^{-1} \hat{\beta}) = \sigma_N^{-1} \Upsilon_N X_N Y_N$ which implies that

$$\begin{aligned} \sigma_N^{-1} \begin{bmatrix} \hat{\beta}_0 \\ N \hat{\beta}_1 \end{bmatrix} &\Rightarrow 12 \begin{bmatrix} 1/3 & -1/2 \\ -1/2 & 1 \end{bmatrix} \begin{bmatrix} \int_0^1 W_d(r) dr \\ \int_0^1 r W_d(r) dr \end{bmatrix} \\ &= \begin{bmatrix} 4 \int_0^1 W_d(r) dr - 6 \int_0^1 r W_d(r) dr \\ -6 \int_0^1 W_d(r) dr + 12 \int_0^1 r W_d(r) dr \end{bmatrix} \end{aligned}$$

c) Note that

$$\begin{aligned} &\frac{1}{N} \sum_{t=1}^N \sigma_N^{-1} y_t \left(\frac{t}{N} \right) \\ &= N^{-1} \sigma_N^{-1} \sum_{t=1}^N ((\Delta y_1) + \dots + (\Delta y_1 + \dots + \Delta y_t) + \dots + (\Delta y_1 + \Delta y_2 + \dots + \Delta y_N)) \\ &= \sum_{t=1}^N \sigma_N^{-1} \Delta y_t \left(1 - \frac{t}{N} \right) = \sum_{t=1}^N \sigma_N^{-1} \Delta y_t - \sum_{t=1}^N \sigma_N^{-1} \Delta y_t \frac{t}{N} \end{aligned}$$

Hence, by Lemmas 1(b) and 1(d)

$$\sum_{t=1}^N \sigma_N^{-1} \Delta y_t \frac{t}{N} \Rightarrow W_d(1) - \int_0^1 r W_d(r) dr$$

Next, by using this result, Lemmas 1(b), 2(b) and some direct calculations, we obtain,

$$\sigma_N^{-2} \left(\hat{\beta}_0 \sum_{t=1}^N \Delta y_t + N \hat{\beta}_1 \sum_{t=1}^N \left(\frac{t}{N} \right) \Delta y_t \right) \Rightarrow A(d)$$

d) Since $e'X_N = 0$, we have that

$$\begin{aligned} N^{-1} \sigma_N^{-2} \sum_{t=1}^N e_{t-1}^2 &= N^{-1} \sigma_N^{-2} \sum_{t=1}^N e_{t-1} y_{t-1} \\ &= N^{-1} \sum_{t=1}^N \sigma_N^{-2} y_{t-1}^2 - \sigma_N^{-1} \hat{\beta}_0 N^{-1} \sum_{t=1}^N \sigma_N^{-1} y_{t-1} - \sigma_N^{-1} N \hat{\beta}_1 \left(N^{-1} \sum_{t=1}^N \left(\frac{t}{N} \right) \sigma_N^{-1} y_{t-1} \right) \\ &\Rightarrow \int_0^1 W_d(r)^2 dr - 4 \left(\int_0^1 W_d(r) dr \right)^2 \\ &\quad + 12 \left(\int_0^1 W_d(r) dr \int_0^1 r W_d(r) dr \right) - 12 \left(\int_0^1 r W_d(r) dr \right)^2 = B(d) \end{aligned}$$

and the convergence result follows by Lemmas 1(b), 1(d) and 2(b).

Proof of Proposition 1:

i) If $d > 0$, we get from (4) and Lemmas 1(d) and 1(e) that

$$N \hat{\phi}_{AO} = \frac{\sigma_N^{-2} \sum_{t=1}^N y_{t-1} \Delta y_t}{N^{-1} \sigma_N^{-2} \sum_{t=1}^N y_{t-1}^2} \Rightarrow \frac{\frac{1}{2} W_d(1)^2}{\int_0^1 W_d(r)^2 dr}$$

ii) If $d < 0$ we get from (4) and Lemmas 1(d) and 1(f) that

$$N^{1+2d} \hat{\phi}_{AO} = \frac{N^{2d} \sigma_N^{-2} \sum_{t=1}^N y_{t-1} \Delta y_t}{N^{-1} \sigma_N^{-2} \sum_{t=1}^N y_{t-1}^2} \Rightarrow - \left[\frac{\pi \theta^2}{\sigma_\varepsilon^2 K(d)} + \frac{(0.5 + d) \Gamma(1 + d)}{\Gamma(1 - d)} \right] / \int_0^1 W_d(r)^2 dr$$

iii) The results follow directly from Franses and Haldrup (1994, p.477).

Proof of Proposition 2:

i) For $d > 0$, Lemmas 1(b), 1(d)-(e) and 2(a) imply that

$$\begin{aligned} N \hat{\varphi}_{AO}^1 &= \frac{\sigma_N^{-2} \sum_{t=1}^N (y_{t-1} - \bar{y}_N) \Delta y_t}{N^{-1} \sigma_N^{-2} \sum_{t=1}^N y_{t-1} (y_{t-1} - \bar{y}_N)} = \frac{\sum_{t=1}^N \sigma_N^{-2} y_{t-1} \Delta y_t - \sigma_N^{-1} \bar{y}_N \sum_{t=1}^N \sigma_N^{-1} \Delta y_t}{N^{-1} \sigma_N^{-2} \sum_{t=1}^N y_{t-1}^2 - (\sigma_N^{-1} \bar{y}_N)^2} \\ &\Rightarrow \frac{0.5 W_d(1)^2 - W_d(1) \int_0^1 W_d(r) dr}{C(d)} \end{aligned}$$

ii) For $d < 0$, Lemmas 1(b), 1(d), 1(f) and 2(a) imply that,

$$\begin{aligned} N^{1+2d}\hat{\varphi}_{AO}^1 &= \frac{N^{2d} \sum_{t=1}^N \sigma_N^{-2} y_{t-1} \Delta y_t + o_p(N^{2d})}{N^{-1} \sigma_N^{-2} \sum_{t=1}^N y_{t-1} (y_{t-1} - \bar{y}_N)} \\ &\Rightarrow -C(d)^{-1} \left(\frac{\pi \theta^2}{\sigma_\varepsilon^2 K(d)} + \frac{(0.5 + d) \Gamma(1 + d)}{\Gamma(1 - d)} \right) \end{aligned}$$

iii) For $d = 0$ note that $u_t = \varepsilon_t$ and by using Lemmas 1(d), 1(g) and 2(a) we get

$$\begin{aligned} N\hat{\varphi}_{AO}^1 &= \frac{\sum_{t=1}^N \sigma_\varepsilon^{-2} (y_{t-1} - \bar{y}_N) \Delta y_t}{N^{-1} \sigma_\varepsilon^{-2} \sum_{t=1}^N y_{t-1} (y_{t-1} - \bar{y}_N)} \\ &\Rightarrow \frac{0.5 (W(1)^2 - 1) - W(1) \int_0^1 W(r) dr}{C(0)} - \frac{\pi \theta^2}{\sigma_\varepsilon^2 C(0)} \end{aligned}$$

iv) Let $\tilde{e}_t$ be the residual from the regression of $\bar{y}_t$ on x_t . The Frisch-Waugh theorem (Hayashi (2000, p.73)) and some direct calculations imply that $\hat{\varphi}_{AO}^2$, is numerically equal to the OLS regression estimate of $\tilde{e}_{t-1}$ on $\bar{y}_t$, thus,

$$\hat{\varphi}_{AO}^2 = \left(\sum_{t=1}^N \tilde{e}_{t-1}^2 \right)^{-1} \sum_{t=1}^N \tilde{e}_{t-1} \Delta \bar{y}_t.$$

Because y_t differs from $\bar{y}_t$ only by a linear trend $\beta_0 + \beta_1 t$, we also have that $\tilde{e}_t = e_t$ and $\Delta \bar{y}_t = \beta_1 + \Delta y_t$. Hence, if $d > 0$, by Lemmas 1(e), 2(c) and 2(d),

$$\begin{aligned} N\hat{\varphi}_{AO}^2 &= \frac{\sigma_N^{-2} \sum_{t=1}^N e_{t-1} \Delta y_t}{N^{-1} \sigma_N^{-2} \sum_{t=1}^N e_{t-1}^2} \\ &= \frac{\sum_{t=1}^N \sigma_N^{-2} y_{t-1} \Delta y_t - \sigma_N^{-1} \hat{\beta}_0 \sum_{t=1}^N \sigma_N^{-1} \Delta y_t - N \sigma_N^{-1} \hat{\beta}_1 \sum_{t=1}^N \left(\frac{t}{N}\right) \sigma_N^{-1} \Delta y_t}{N^{-1} \sigma_N^{-2} \sum_{t=1}^N e_{t-1} y_{t-1}} \\ &= \frac{\sigma_N^{-2} \sum_{t=1}^N y_{t-1} \Delta y_t - \sigma_N^{-2} B_N}{N^{-1} \sigma_N^{-2} \sum_{t=1}^N e_{t-1} y_{t-1}} \Rightarrow \frac{0.5 W_d(1)^2 - A(d)}{B(d)} \end{aligned}$$

v) Lemmas 1(e), 1(f), 2(c) and 2(d) imply that

$$\begin{aligned} N^{1+2d}\hat{\varphi}_{AO}^2 &= \frac{\sigma_N^{-2} \sum_{t=1}^N e_{t-1} \Delta y_t}{N^{-1} \sigma_N^{-2} \sum_{t=1}^N e_{t-1}^2} = \frac{\sum_{t=1}^N N^{2d} \sigma_N^{-2} y_{t-1} \Delta y_t - N^{2d} (\sigma_N^{-2} B_N)}{N^{-1} \sigma_N^{-2} \sum_{t=1}^N e_{t-1} y_{t-1}} \\ &\Rightarrow -B(d)^{-1} \left[\frac{\pi \theta^2}{\sigma_\varepsilon^2 K(d)} + \frac{(0.5 + d) \Gamma(1 + d)}{\Gamma(1 - d)} \right] \end{aligned}$$

vi) Lemmas 1(g), 2(c) and 2(d) imply that

$$\begin{aligned}
N\hat{\varphi}_{AO}^2 &= \frac{\sigma_\varepsilon^{-2} \sum_{t=1}^N e_{t-1} \Delta y_t}{N^{-1} \sigma_\varepsilon^{-2} \sum_{t=1}^N e_{t-1}^2} \\
&= \frac{\sigma_\varepsilon^{-2} \sum_{t=1}^N y_{t-1} \Delta y_t - \sigma_N^{-2} B_N}{N^{-1} \sigma_\varepsilon^{-2} \sum_{t=1}^N e_{t-1} y_{t-1}} \Rightarrow \frac{0.5(W(1)^2 - 1 - 2A(0))}{B(0)} - \frac{\pi\theta^2}{\sigma_\varepsilon^2 B(0)}
\end{aligned}$$

Proof of Proposition 3:

Note that since $\hat{d}_N = d + o_P((\log N)^{-1})$ it follows that $N^{1+2\hat{d}_N} / N^{1+2d} \rightarrow 1$ in probability. So, $N^{1+2d} \hat{\phi}_{AO} = N^{1+2\hat{d}_N} (1 + o_P(1)) \hat{\phi}_{AO} = N^{1+2\hat{d}_N} \hat{\phi}_{AO} + o_P(N^{1+2\hat{d}_N} \hat{\phi}_{AO}) = N^{1+2\hat{d}_N} \hat{\phi}_{AO} + o_P(1)$, since $N^{1+2d} \hat{\phi}_{AO} = O_P(1)$ by Proposition 1.

References

- Andersen, T., Bollerslev, T., Diebold, F.X. and Ebens, P. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61, 43-76.
- Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P.(2003). Modelling and forecasting realized volatility. *Econometrica*, 71, 579–625.
- Andersen, T.G., Dobrev, D. and Schaumburg, E. (2010). Jump-Robust Volatility Estimation using Nearest Neighbor Truncation, Federal Reserve Bank Staff Reports 465.
- Baillie, R.T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73, 6–59.
- Bauwens L., Hafner, C.M. and Laurent, S. (2012). *Handbook of Volatility Models and Their Applications*, Hoboken NJ, John Wiley and Sons.
- Chung, C-F. and Baillie, R.T.(1993). Small sample bias in conditional sum of squares estimators of fractionally integrated ARMA models. *Empirical Economics*, 18, 791–806.
- Chen, Ch. and Liu, L.-M. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*, 88, 284-297.
- Franses, P.H. and Haldrup, N. (1994). The effects of additive outliers on tests for unit roots and cointegration. *Journal of Business and Economics Statistics*, 12(4) 471-478.
- Geweke J. and Porter-Hudak, S.(1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4, 221–238.
- Granger, C. W. J. and Joyeux, R. (1980). An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*, 1, 14-29.
- Haldrup N. and Nielsen MO (2007). Estimation of fractional integration in the presence of data noise. *Computational Statistics and Data Analysis*, 51, 3100 – 3114
- Hansen, P. and Lunde, A. (2014), Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory*, 30, 60-93.

- Hayashi, F. (2000). *Econometrics*, Princeton NJ, Princeton University Press.
- Hendry, D.F. and Santos, C. (2005). Regression Models with Data-based Indicator Variables. *Oxford Bulletin of Economics and Statistics*, 67, 571-595.
- Hosking, J.R.M. (1981). Fractional Differencing. *Biometrika*, 68, 165-176.
- Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54, 159-178.
- Lee, C. and Shie, F.-S. (2004). Fractional integration and Phillips-Perron test. *Academia Economic papers*, 32, 273-312.
- Lubian, D. (1999). Long memory errors in time series regression with a unit root. *Journal of time series analysis*, 20, 565-577.
- Lucas, A. (1995). Unit roots tests based on M estimators. *Econometric Theory*, 11, 331-346.
- Robinson, P.M. (2003). Long-memory time series. In: Robinson, P.M. (Ed.), *Time Series With Long Memory*. Oxford, Oxford University Press.
- Perron, P. and Rodriguez, G. (2003). Searching for additive outliers in nonstationary time series. *Journal of Time Series Analysis*, 24, 193-220.
- Perron, P. and Vogelsang, T.J. (1992). Testing for a unit root in a time series with a changing mean: corrections and extensions. *Journal of Business and Economic Statistics*, 10, 467-470.
- Rossi, E. and Santucci de Magistris, P. (2014), Estimation of long memory in integrated variance. *Econometric Reviews*, 33, 785-817.
- Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25, 197-231.
- Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks, working paper, Kingston, Canada, Queen's University.

- Shin, D.W., Sarkar, S. and Lee, J.H. (1996). Unit root tests for time series with outliers. *Statistics and Probability Letters*, 30, 189–197.
- Sowell, F. (1990). The fractional unit root distribution. *Econometrica*, 58(2), 495–505.
- Tolvi, J. (2003). Long memory and outliers in stock market Returns. *Applied Financial Economics*, 13, 495–502.
- Vogelsang, T.J. (1999). Two simple procedures for testing for a unit root when there are additive outliers. *Journal of Time Series Analysis*, 20, 237–252.
- Yin, Y. and Maddala, G.S. (1997). The effects of different types of outliers on unit roots tests. In: T.Fomby and R.C.Hill (eds.) *Advances in Econometrics*, 13, 269–305, JAI Press.

Figure 1: Log median truncated realized volatility (MEDRV) measure for the S&P 500.

(a) ACF of levels of log MEDRV

(b) ACF of first differences of log MEDRV

Figure 2: ACF of levels and first differences of the log median truncated realized volatility (MEDRV) measure for the S&P 500.