Different Strokes for Different Folks: Long Memory and Roughness*

Shuping Shi
Macquarie University

Jun Yu
Singapore Management University

August 3, 2021

Abstract

The log realized volatility of financial assets is often modeled as an autoregressive fractionally integrated moving average model (ARFIMA) process, denoted by ARFIMA(p,d,q), with $p = 1$ and $q = 0$. Two conflicting results have been found in the literature regarding the dynamics. One stream shows that the data series has a long memory (i.e., the fractional parameter $d > 0$) with strong mean reversion (i.e., the autoregressive coefficient $|\alpha_1| \approx 0$). The other stream suggests that the volatility is rough (i.e., $d < 0$) with highly persistent dynamic (i.e., $\alpha_1 \to 1$). To consolidate the findings, this paper first examines the finite sample properties of alternative estimation methods employed in the literature for the ARFIMA$(1,d,0)$ model and then applies the outperforming techniques to a wide range of financial assets. The candidate methods include two parametric maximum likelihood (ML) methods (the maximum time-domain modified profile likelihood (MPL) and maximum frequency-domain likelihood) and two semiparametric methods (the local Whittle method and log periodogram estimation method). The two parametric methods work well across all parameter settings, with the MPL method outperforming. In contrast, the two semiparametric methods have a very large upward bias for d and an equally large downward bias for α_1 when α_1 is close to unity. The poor performance of the semiparametric methods in the presence of a highly persistent dynamic might lead to a false conclusion of long memory. In the empirical applications, we find that the log realized volatilities of exchange rate futures over the past decade have a long memory, where the point estimate of d is between 0.4 and 0.5 and the estimate of α_1 is near zero. For other financial assets considered (including stock indices and industry indices), we find that they have rough volatility, with the point estimate of d being negative and the point estimates of α_1 close to unity.

JEL classification: C15, C22, C32
Keywords: Long memory; fractional integration; roughness; short-run dynamics; realized volatility

*We are grateful to Dacheng Xiu for providing us his realized volatility data, Peter Phillips and Jia Li for comments. Shi acknowledges research support from the Australian Research Council under project No. DE190100840. Yu acknowledges the support from the Lee Foundation. Shuping Shi, Department of Economics, Macquarie University; E-mail: shuping.shi@mq.edu.au. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Rd, Singapore 178903. Email: yujun@smu.edu.sg.

1 Introduction

The availability of intraday prices of financial assets fosters the development of high-frequency financial econometrics, making an accurate measurement of daily 'realized' volatility possible. The estimated daily realized volatility (RV) has been shown helpful for various purposes, including forecasting macroeconomic fundamentals (Andersen, Bollerslev, Diebold, and Wu, 2005), making investment decisions (Fleming, Kirby, and Ostdiek, 2003), pricing options (Christoffersen, Feunou, Jacobs, and Meddahi, 2014), managing financial risk (Christoffersen and Diebold, 2000), and estimating model parameters (Tao, Phillips, and Yu, 2019).

It is now well known that the log realized volatility of financial assets exhibits long-range dependence, with its autocorrelation function (ACF) delaying slowly. A class of autoregressive fractionally integrated moving average (ARFIMA) models, particularly ARFIMA(p, d, q) with $p = 1$ and $q = 0$, has gained much prominence in modeling daily log realized volatility. Suppose the first-order autoregressive coefficient $\alpha_1 = 0$ (i.e., ARFIMA$(0, d, 0)$). When $d > 0$, its ACF decays hyperbolically and is not absolutely summable. On the other hand, when $d < 0$, the process is anti-persistent, and sample paths generated from the model are rougher than that when $d > 0$. The value of the fractional parameter has important implications for both theoretical and empirical analysis of realized volatility. As such, the main focus of the literature has been on the estimation of d.

Several estimation techniques for the fractional parameter have been proposed, including the local Whittle method (Künsch, 1987; Robinson, 1995a) and the log periodogram estimation method (Geweke and Porter-Hudak, 1983; Robinson, 1995b). These two methods rely on the asymptotic behavior of the spectral density at frequencies near zero (ignoring short-run dynamics) and hence are often referred to as semiparametric methods. When the semiparametric methods are applied to log realized volatilities, it is often found that the point estimate of d is around 0.4. See, for example, Andersen and Bollerslev (1997), Andersen, Bollerslev, Diebold, and Labys (2001), Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2003). Such an estimate implies that the log realized volatility has a long memory. Further, one could compute the short-run parameter from the prefiltered data (based on the estimated d). The estimated short-run parameter typically suggests strong mean reversion or weak short-run behavior. For example, if an AR(1) model is fitted to the filtered data, the estimated AR(1) parameter is often close to zero.

One advantage of the semiparametric methods is their asymptotic robustness to short-run dynamics, as short-run behavior does not change the asymptotic spectral density at near-zero frequencies. This insensitive relationship, however, does not necessarily hold in finite samples. In particular, when the autoregressive coefficient α_1 is close to unity, the spectral density that ignores the near-unity behavior is expected to approximate the actual spectral density poorly, even with a large sample size and at frequencies near zero. This concern might have important empirical implications.

The ARFIMA$(0, d, 0)$ model is asymptotically equivalent to the fractional Gaussian noise (fGn) with the Hurst parameter H. In particular, when $H = d + 0.5$, the ACF of the two models has the same order as the lag length goes to infinity. It is well known that the fGn is the increment of the fractional Brownian motion (fBm) whose sample path is (locally) Hölder continuous up to order H. The sample

path of fBm is rough when $H < 0.5$ and smooth when $H > 0.5$. Interestingly, Gatheral, Jaisson, and Rosenbaum (2018) have found strong evidence of $H < 0.5$ (i.e., roughness) in the implied volatility surface obtained from options prices. In the same paper, the fBm with $H = 0.14$ is used to forecast RVs and log RVs. It is found that the rough fBm can generate very accurate forecasts out-of-sample. Evidence of roughness in volatility is also found in Bayer, Friz, and Gatheral (2016) using variance swaps.

Wang, Xiao, and Yu (2019) consider a fractional Ornstein-Uhlenbeck (fOU) process. Under the infill asymptotic scheme, the exact discrete-time representation of the fOU process is a local-to-unity (Phillips, 1987) process with fGn.[1] A two-stage estimation method is proposed to estimate parameters in the fOU. The estimated H from several well known daily realized volatility time series is similar to what Gatheral, Jaisson, and Rosenbaum (2018) find in the risk-neutral measure, i.e., $H < 0.5$. The (pre-imposed) local-to-unity dynamic generates strong persistency that is attenuated by the anti-persistent error. The fOU process can generate rough sample paths and produce ACF that decays slowly at small and moderate lags.

Clearly, the empirical evidence for RV found in the discrete-time literature by semiparametric approaches is at odds with that in the risk-neutral measure with the fBm model and the physical measure with the fOU model. While the semiparametric methods suggest a process with weak short-run behavior and long memory errors (referred to as Model 1) for log realized volatilities, the empirical evidence obtained from the fOU model reveals near-unity short-run behavior and anti-persistent errors (referred to as Model 2).

The main goal of this paper is to consolidate these conflicting empirical findings. To achieve this goal, we first examine the finite sample properties of different estimation techniques for the ARFIMA$(1, d, 0)$ model under a wide range of parameter settings. The methods include two parametric maximum likelihood (ML) methods (modified profile time-domain likelihood (MPL) and frequency domain or Whittle likelihood) and the two aforementioned semiparametric methods. Importantly, when there is a near-unity behavior in the short-run dynamics (i.e., α_1 is close to 1), the two semiparametric methods estimate d with a severe upward bias. In particular, if d is negative, the two semiparametric methods tend to find a positive estimate for d and a near-zero estimate for α_1. On the contrary, the two ML methods can always estimate d and α_1 accurately with no noticeable bias and small standard error, regardless of the value of α_1. We provide the theoretical reasons why the two semiparametric methods estimate d and α_1 with severe bias when α_1 takes a value close to 1.

The alternative methods are employed to estimate the ARFIMA$(1, d, 0)$ model for the log realized volatilities of thirteen financial assets for the maximum available sample period in the past decade. The realized volatilities are computed using the novel quasi-ML (QML) technique proposed by Da and Xiu (2021), accounting for market microstructure noise and using the highest sampling frequency. The application reveals very interesting empirical results. The ML methods suggest near-unity short-run behavior and anti-persistent errors for the log RVs of the five market index futures or ETFs and five

[1] Similar models have been considered in other papers. For example, Magdalinos (2012) proposes a mildly explosive autoregressive process with a long memory error (i.e., $d \in (0, 0.5)$). Yu (2021) considers a latent local-to-unity model with fractionally integrated process.

industry index ETFs. Specifically, the point estimates of the fractional parameter d are negative (between -0.482 and -0.262), and the point estimates of the autoregressive coefficient α_1 are close to one (between 0.988 and 0.998). In sharp contrast, from the ML methods, the log RVs of the three exchange rate futures have a long memory with strong mean-reversion. The point estimates of d are positive, in the range of $(0.438, 0.473)$, and the estimates of α_1 are close to 0, in the range of $(-0.075, -0.057)$. Regardless of the log RV series, the two semiparametric methods always lead to similar estimates of d (greater than zero) and α_1 (close to zero). These estimation results are consistent with our findings in the Monte Carlo studies.

Our paper contributes to the literature in two aspects. First, our results help to understand the finite sample properties of alternative methods in a wide range of parameter values for d and α_1. Existing Monte Carlo studies in the context of ARFIMA$(1, d, 0)$ model set the maximum value of α to 0.8 or 0.9 and the sample size (T) to 96, 256, 512, or 576; see Smith, Taylor, and Yadav (1997), Nielsen and Frederiksen (2005), Nadarajah, Martin, and Poskitt (2021). The existing studies found that when $\alpha_1 = 0.8$ or 0.9, there is a substantial upward bias in d with the semiparametric methods. An empirically more relevant case is when α_1 is much closer to unity and the sample size is much larger. We consider the case of $\alpha_1 = 0.99$ and $T = 1024, 2048$, in addition to the existing parameter settings. Our simulation results show that a substantial upward bias in d continues to exist in the semiparametric methods in these more empirically relevant cases. For example, when $\alpha_1 = 0.99$, the semiparametric methods yield a positive estimate for d, even when the true value of d is negative. Moreover, we show that the second-stage estimate of α_1 is substantially downward biased. For example, when the true value of α_1 is near unity, the semiparametric methods yield a near-zero estimate for α_1. As in Smith, Taylor, and Yadav (1997) and Nielsen and Frederiksen (2005), we find that the two ML methods can estimate all parameters accurately with the small standard error and negligible bias. This result holds even when α_1 is closer to unity. Unlike Nielsen and Frederiksen (2005) where the Whittle method is shown to produce the most accurate estimate for d, we find the MPL works slightly better than the Whittle method. This difference is likely due to how we calculate the variance-covariance matrix of the model when implementing the MPL method.

Second, our empirical results help to make sense of the conflicting empirical findings in the RV literature. Whether the log RV should be captured by Model 1 (i.e., $d > 0$ and $\alpha_1 \approx 0$) or Model 2 (i.e., $d < 0$ and $\alpha_1 \to 1$) depends on the asset class at hand. These two models represent the two local maxima of the likelihood function of the ARFIMA$(1, d, 0)$ model. We find that Model 2 provides a better fit for the log RVs of market index futures and industry index ETF. This finding is consistent with what Gatheral, Jaisson, and Rosenbaum (2018) find in the risk-neutral measure and what Wang, Xiao, and Yu (2019) find in the physical measure. Meanwhile, Model 1 captures the dynamics of the log RV of exchange rate futures better than Model 2. This result is consistent with what Andersen and Bollerslev (1997), Andersen, Bollerslev, Diebold, and Labys (2001), Andersen, Bollerslev, Diebold, and Labys (2003) find for the log RVs of exchange rates.

The paper is organized as follows. Section 2 introduces the realized volatility estimators. Section 3 presents the model specification and reviews some statistical properties of the model. Section 4

introduces alternative estimation methods. Section 5 presents the simulation designs and reports the finite sample properties of alternative estimation approaches. Section 6 reports empirical results when the ARFIMA$(1,d,0)$ model is fitted to the log RVs of a wide range of financial assets over the past decade. Section 7 concludes. The Appendix reviews a technique, known as tapering, for the Whittle and the local Whittle methods and examines their finite sample performance.

2 Realized Volatility

Suppose the observed log asset prices consist of two components:

$$X_t^o = X_t + U_t,$$

where X_t is the underlying log efficient price and U_t is the noise component. The underlying price is assumed to be an Itó-semimartingale process defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and satisfies

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \left(\delta 1_{\{|\delta|\leq 1\}} * (\eta - \upsilon)\right)_t + \left(\delta 1_{\{|\delta|>1\}}\right)\mu_t, \tag{1}$$

where μ_t and σ_t are adapted and locally bounded, W is a standard Brownian motion, η is a Poisson random measure on $\mathbb{R}_+ \times E$ with a non-random intensity measure $\upsilon(dt, ds) = dt \otimes \lambda(ds)$, and λ is a σ-finite measure on (E, ξ) which is a Polish space. The last two components of (1) capture the dynamics of jumps. The noise process U_t is assumed to have flexible serial correlations, modeled as an MA(∞) process. See, for example, Jacod, Li, and Zheng (2017) or Da and Xiu (2021) for more details of the assumptions.

Assume data are observed at a regular frequency. Let $t = 1, \cdots, T$ and $n = T/\delta$ be the total number of intra-day observations available within sample period, where δ is the distance between two consecutive observations. The traditional realized volatility is constructed as

$$RV_t = \sum_{i=2}^{1/\delta} \left(\Delta X_{t,i}^o\right)^2, \text{ with } \Delta X_{t,i}^o = X_{t,i}^o - X_{t,i-1}^o, \tag{2}$$

where $X_{t,i}^o$ is the observed i^{th} log prices at period t. In the absence of noise U_t and jumps, the realized volatility is shown to be a consistent estimator of the integrated volatility and converges to a normal distribution such that

$$\delta^{-1/2}\left(RV_t - \int_{t-1}^t \sigma_s^2 ds\right) \to_d N\left(0, \int_{t-1}^t \sigma_s^4 ds\right). \tag{3}$$

It is well known that the realized volatility estimator is inconsistent when the data is contaminated by market microstructure noises and jumps. The literature on noise-robust volatility estimators is enormous. See Da and Xiu (2021) for a brief review.

One of the most recent contribution in this literature is made by Da and Xiu (2021), who develop a

quasi-ML (QML) approach providing uniform valid inference on volatility under the extremely general model setting (1). The likelihood function is taken from a much simplified process, assuming the efficient price follows a Brownian motion with constant volatility and a Gaussian MA(q) noise component. The QML estimator of the volatility, denoted by $\hat{\sigma}^2(\hat{q})$ with $\hat{q}$ obtained from the Akaike information criterion, is shown to converges to the following quantity

$$C_t = \delta \left[\int_{t-1}^{t} \sigma_s^2 ds + \sum_{i=2}^{1/\delta} (\Delta X_{t,i})^2 \right],$$

which comprises both continuous (integrated variance) and discontinuous (jump) components. The QML realized volatility estimator has an asymptotically normal distribution. The convergence rate depends on the magnitude of the noise component. Further, the QML estimator is shown to provide more accurate estimation results than other noise-robust volatility estimators (including the traditional RV obtained from returns sampled at the 5-minute frequency, the pre-averaging method of Jacod, Li, and Zheng (2019), and the flat-top realized kernel estimator of Varneskov (2017)) in finite samples.

The aim of this paper is to investigate the dynamics of the QML realized volatility estimates of various financial assets. For convenience, we refer to the QML volatility estimator as realized volatility subsequently and, with a slight abuse of notation, we denote the log QML volatility estimator $\log \hat{\sigma}^2(\hat{q})$ by y_t.

3 Model Specification

Consider the ARFIMA(p, d, q) model

$$\alpha(L)(y_t - \mu) = \sigma_u \beta(L) u_t, \tag{4}$$

where $\alpha(L) = 1 - \alpha_1 L - \ldots - \alpha_p L^p$, $\beta(L) = 1 - \beta_1 L - \ldots - \beta_q L^q$, L is the lag operator, and u_t is the error term. We assume the roots of $\alpha(L) = 0$ and $\beta(L) = 0$ lie outside of the unit circle. We are particularly interested in the case where one root of $\alpha(L)$ is very close to the unit circle. The error term is a fractionally integrated process (Granger and Joyeux, 1980) such that

$$u_t = (1 - L)^{-d} \varepsilon_t \text{ with } \varepsilon_t \sim_{iid} N(0, 1), \tag{5}$$

where $d \in (-1/2, 1/2)$ is the memory parameter. The fractional integrated process can be rewritten as

$$u_t = \sum_{k=0}^{\infty} \frac{\Gamma(k + d)}{\Gamma(d)\Gamma(k + 1)} \varepsilon_{t-k},$$

where $\Gamma(\cdot)$ is the gamma function. The process $\{u_t\}$ is stationary and invertible (Bloomfield, 1985). The instantaneous variance of u_t is $E(u_t^2) = \frac{\Gamma(1-2d)}{(\Gamma(1-d))^2}$. The long-run variance of u_t is 1 when $d = 0$, ∞

when $d > 1/2$, and 0 when $d < 1/2$.

3.1 Spectral density

Under the model assumption, the process is covariance stationary and hence we can write $Cov(y_t, y_s) := \gamma_y(k)$ with $k = |t-s|$. The spectral density of y_t is the Fourier transformation of the autocovariance such that

$$f_y(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} \gamma_y(k),$$

where $-\pi \leq \lambda \leq \pi$. Under the specification of (4) and (5), the spectral density of y_t is

$$f_y(\lambda) = \frac{\sigma_u^2}{2\pi} |1 - \exp(-i\lambda)|^{-2d} \frac{|\beta(\exp(-i\lambda))|^2}{|\alpha(\exp(-i\lambda))|^2}.$$

See, for example, Robinson (2003). Using Euler's formula and trigonometric identities, the spectral density can be rewritten as

$$f_y(\lambda) = \frac{\sigma_u^2}{2\pi} \left(\sqrt{2 - 2\cos(\lambda)}\right)^{-2d} \frac{\left(1 + \sum_{j=1}^{q} \beta_j \cos(j\lambda)\right)^2 + \left(\sum_{j=1}^{q} \sin(j\lambda)\right)^2}{\left(1 - \sum_{j=1}^{p} \alpha_j \cos(j\lambda)\right)^2 + \left(\sum_{j=1}^{p} \alpha_j \sin(j\lambda)\right)^2}. \tag{6}$$

For the special case of $p = 0$ and $q = 0$, the spectral density becomes

$$f_y(\lambda) = f_u(\lambda) = \frac{\sigma_u^2}{2\pi} \left(\sqrt{2 - 2\cos(\lambda)}\right)^{-2d}.$$

The spectral density is infinite at the origin ($\lambda = 0$) for $d > 0$ and zero for $d < 0$.

3.2 Variance-covariance matrix of u_t

Let $\gamma_u(k) := Cov(u_t, u_{t-k})$ be the k^{th} order autocovariance of u_t. Under the specification of (5), the autocovariance function of u_t is (Hosking, 1981)

$$\gamma_u(k) = \frac{(-1)^k (-2d)!}{(k-d)!\,(-k-d)!} = \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\,\Gamma(1-k-d)}, \tag{7}$$

where $(\cdot)!$ is the factorial of the argument.

When $k \to \infty$, Hosking (1981) obtains the k^{th} order ACF of u_t as

$$\rho_u(k) = \frac{(-d)!\,(k+d-1)!}{(d-1)!\,(k-d)!} \sim \frac{(-d)!}{(d-1)!} k^{2d-1}.$$

The correlation coefficient $\rho_u(k)$ decays at a hyperbolic rate as k goes to infinity. This is in contrast to the exponential decaying rate of an ARMA(p, q) model.

4 Estimation Methods

In this section, we review the time-domain ML, the approximate frequency domain ML, the local Whittle estimation (LWE) method, and the log periodogram estimation (LPE) method.

4.1 Time-domain ML estimation

Let $y = (y_1, y_2, \cdots, y_T)'$ and $\theta = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, d)$. Under the model specification of (4) with u_t specified as (5), $y_t - \mu$ follows a normal distribution with mean zero and variance-covariance matrix, denoted by Σ_y. The objective function of the ML estimator is given by

$$(\hat{\theta}, \hat{\sigma}_u) = \arg\max_{\theta, \sigma_u} \log L_N(\mu, \sigma_u, \theta),$$

where

$$\log L_N(\mu, \sigma_u, \theta) = \frac{1}{2T} \log|\Sigma_y| + \frac{1}{2T} (y - \mu l)' \Sigma_y^{-1} (y - \mu l), \tag{8}$$

and $l = (1, \ldots, 1)'$.

For the case of known mean value μ, the limiting properties of $\hat{\theta}$ was derived by Hannan (1973) for short memory processes and Yajima (1985) for long memory processes. That is, under some mild regularity conditions,

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \to_d N\left(0, \Xi_{\theta_0}^{-1}\right),$$

where θ_0 is the true parameter vector and Ξ_{θ_0} is the Fisher information matrix.

4.1.1 Modified profile likelihood

Dahlhaus (1989) extends the results of Yajima (1985) to the case with unknown mean. In case of unknown μ, a plug-in method is required. The plug-in method substitutes μ by a consistent estimator of the mean (e.g., the sample mean). Although the method provides a $\sqrt{T}$ consistent and asymptotically normal estimator, it is contaminated by an additional second order negative bias (Lieberman, 2005) due to the need of estimating μ.

An alternative solution is the modified profile likelihood (MPL) estimator proposed by Cox and Reid (1987). The idea of the MPL estimator is to use a linear transformation of parameters of interest to make them orthogonal to nuisance parameters (μ and σ_u). The modified profile likelihood is given by

$$\log L_M(y, \hat{\mu}, \theta) = \left(\frac{1}{T} - \frac{1}{2}\right) \log|R| - \frac{1}{2} \log\left(l' R^{-1} l\right) + \frac{3 - T}{2} \log\left[T^{-1} (y - \hat{\mu} l)' R^{-1} (y - \hat{\mu} l)\right], \tag{9}$$

where $R = \Sigma_y / \sigma_u^2$ and $\hat{\mu} = \left(l' R^{-1} l\right)^{-1} l' R^{-1} Y$. The asymptotic distribution of the MPL estimator is unchanged compared with the exact ML but eliminates some degree of bias in the exact ML (An and Bloomfield, 1993; Hauser, 1999).

4.1.2 Variance-covariance matrix Σ_y

The variance-covariance matrix Σ_y is of dimension $T \times T$ and its $(t,s)^{th}$ element is $\gamma_y(k)$, where $t, s = 1, \cdots, T$ and $k = |t-s|$. The covariance function of the ARFIMA(p,d,q) process was derived by Hosking (1981, Lemma 1(c)) and Sowell (1992, eq. (8)-(9)) and approximated to improve computational speed by Chung (1994). In the special case of $p = 1$ and $q = 0$, the covariance function is

$$\gamma_y(k) = \frac{\sigma_u^2}{\alpha_1\left(1-\alpha_1^2\right)}\gamma_u(k)\, A(k,\alpha_1). \tag{10}$$

where $A(k,\alpha_1) = C(k,\alpha_1) + C(-k,\alpha_1) - 1$, $C(k,\alpha_1) = F(d+k, 1; 1-d+k; \alpha_1)$, and $F(\cdot)$ is the hypergeometric function.

As noted in Liu, Shi, and Yu (2020), the hypergeometric function is computational costly and extremely large when diverges k is large and α_1 is far from unity. As such, we employ the splitting method (Bertelli and Caporin, 2002) which is based on the following property of the covariance function for stationary processes (Brockwell and Davis, 2009):

$$\gamma_y(k) = \sum_{s=-\infty}^{\infty} \tilde{\gamma}(s)\,\gamma_u(k-s),$$

where $\tilde{\gamma}(s)$ is the autocovariance of the pure ARMA component. The summand is truncated at m, which takes a larger value when α_1 is close to unity.[2]

4.2 Approximate frequency-domain ML: Whittle estimator

To avoid inverting Σ_y that is required in calculating the time-domain likelihood function, following Whittle (1953, 1954), one can approximate Σ_y^{-1} by $(2\pi)^{-2}\int_{-\pi}^{\pi} f_y(\lambda)^{-1}\cos((i-j)\lambda)\,d\lambda$ and $\log|\Sigma_y|$ by $T(2\pi)^{-1}\int_{-\pi}^{\pi}\log f_y(\lambda)\,d\lambda$. The discrete-time version of the Whittle likelihood function (up to a scale multiplication) is

$$\log L_W\left(\theta,\sigma_u^2\right) = -\sum_{j=1}^{m}\log f\left(\lambda_j|\theta,\sigma_u^2\right) - \sum_{j=1}^{m}\frac{I(\lambda_j)}{f\left(\lambda_j|\theta,\sigma_u^2\right)}, \tag{11}$$

where $I(\lambda_j)$ denotes the periodogram at the j^{th} Fourier frequency $\lambda_j = 2\pi j/T$ with $j = 1, 2, \ldots, m$ and m being the largest integer contained in $(T-1)/2$. Specifically, we have

$$I(\lambda_j) = \frac{1}{2\pi T}\left|\sum_{t=0}^{T} y_t \exp(-it\lambda_j)\right|^2 \text{ with } \lambda_j = 2\pi j/T, \tag{12}$$

which is a nonparameteric estimate of the density.

The Whittle likelihood function was presented in Künsch (1987) and Dahlhaus (1988). Fox and Taqqu (1986) show that the asymptotic properties of the estimators remain the same if we simplify the

[2]Specifically, we set $m = 1,000$ for $\alpha_1 < 0.9$, $m = 2,000$ for $0.9 \le \alpha_1 < 0.99$, $m = 4,000$ for $0.99 \le \alpha_1 < 0.995$, and $m = 7,000$ for $\alpha_1 \ge 0.995$.

objective function to the following:

$$\log L_W\left(\theta,\sigma_u^2\right)' = -\sum_{j=1}^{m}\frac{I\left(\lambda_j\right)}{f\left(\lambda_j|\theta,\sigma_u^2\right)}, \tag{13}$$

where the distance between the spectrum density $f\left(\lambda_j|\theta,\sigma_u^2\right)$ and $I\left(\lambda_j\right)$ is minimized. We employ the simplified objective function for the estimation.[3] Like MPL, the parameter μ does not enter the objective function of the Whittle method as the zero frequency is not included. The spectral density of the ARFIMA(p,d,q) model is given in (6). The Whittle estimation method yields $\sqrt{T}$-consistent, asymptotically normal and efficient parameter estimates (Hannan, 1973; Fox and Taqqu, 1986; Giraitis and Surgailis, 1990) when $d\in(0,1/2)$.

4.3 Local Whittle estimator

Instead of considering a parametric model, Künsch (1987) and Robinson (1995a) investigate a class of models whose spectral densities satisfy the following property:

$$f_y\left(\lambda\right)\sim C\lambda^{-2d} \text{ as } \lambda\to 0^+ \tag{14}$$

with C being a positive constant and $d\in(-1/2,1/2)$. The property concerns only frequencies approaching zero. When $d\geq 1/2$, a function behaving like λ^{-2d} as $\lambda\to 0^+$ is not integrable so that covariance stationarity cannot be obtained. $d>-1/2$ corresponds to an invertibility condition in parametric models with the above property. The model given by (4) and (5) is a special case in the class with C being a function of θ and σ_u.

The local Whittle likelihood estimator of Künsch (1987) and Robinson (1995a) is defined as

$$\begin{aligned}(\hat{C},\hat{d}) &= \arg\max_{C,d}\frac{1}{m}\sum_{j=1}^{m}\left[-\log f\left(\lambda_j|\theta,\sigma_u^2\right)-\frac{I\left(\lambda_j\right)}{f\left(\lambda_j|\theta;\sigma_u^2\right)}\right] &(15)\\ &= \arg\max_{C,d}\frac{1}{m}\sum_{j=1}^{m}\left[-\log C+2d\log\lambda_j-\frac{1}{C}\lambda^{2d}I\left(\lambda_j\right)\right]. &(16)\end{aligned}$$

with the solution

$$\hat{d}=\arg\max_{d}\left[-\log\hat{C}\left(d\right)+2d\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j\right] \text{ with } \hat{C}\left(d\right)=\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d}I\left(\lambda_j\right), \tag{17}$$

where m satisfies the condition $m\leq(T-1)/2$ and diverges to infinity at a rate that is slower than T as $T\to\infty$. We set $m=T^{0.65}$ in applications.

Robinson (1995a) shows that the local Whittle estimator is consistent at the $\sqrt{m}$ rate and asymp-

[3] Coursol and Dacunha-Castelle (1982) study the approximation error $\log L_N-\log L_W$.

totically normal with variance $1/(4m)$, that is,

$$\sqrt{m}\left(\hat{d}-d\right) \to_d N(0,1/4),$$

when $d \in (-1/2, 1/2)$.[4]

As the local Whittle method only uses information near zero frequency, when a model is parametrically correctly specified, it is expected to be less efficient than the MPL and Whittle methods. However, it is robust against model misspecification asymptotically as long as the misspecification does not violate the assumption in (14).

4.4 Log periodogram estimation

Another alternative is the log periodogram estimation method proposed by Geweke and Porter-Hudak (1983). The proposed regression model is

$$\log I(\lambda_j) = \alpha + \beta z_j + v_j,$$

where $I(\lambda_j)$ is the periodogram of y_t as defined in (12), $z_j = -\log\left(4\sin^2(\lambda_j/2)\right)$, and v_j is the error term. The fractional parameter is estimated as

$$\hat{d} = \hat{\beta} = \frac{\sum_{j=l+1}^{m}(z_j - \bar{z})\log I(\lambda_j)}{\sum_{j=l+1}^{m}(z_j-\bar{z})^2},$$

where $\bar{z} = (m-l)^{-1}\sum_{j=l+1}^{m} z_j$ and $0 \le l < m < T$. The weakest possible upper bound for m is $m/T^{0.9} \to 0$ (Robinson, 1995b). We set $l = 1$ and $m = T^{0.65}$ in applications.

The rationale is as follows. The spectral density in (14) can be approximated by $C\left(4\sin^2(\lambda/2)\right)^{-d}$ (Geweke and Porter-Hudak, 1983). The log spectral dentisy of y_t is therefore

$$\log f(\lambda) \sim \log(C) - d\log\left\{4\sin^2\left(\frac{\lambda}{2}\right)\right\},$$

[4]Velasco (1999) investigates the possibility of using the LW estimator for some non-stationary situations (i.e., $1/2 \le d < 3/2$), showing that the consistency of the LW estimator holds for $d \in (-1/2, 1)$ and the asymptotic normality holds for $d < 3/4$ with the same variance as in the stationary situation. Shimotsu and Phillips (2006) propose an exact local Whittle estimation method, which can be applied to both stationary and non-stationary variables. Unlike the conventional local Whittle estimator, which approximates $I_u(\lambda_j)$ by $\lambda_j^{2d} I_y(\lambda_j)$, the exact local Whittle method is based on the relationship that $I_u(\lambda_j) = I_{\Delta^d y}(\lambda_j)$, where $I_{\Delta^d y}(\lambda_j)$ is the periodogram of $\Delta^d y = (1-L)^d y_t$. While the approximation $I_u(\lambda_j) \sim \lambda_j^{2d} I_y(\lambda_j)$ holds for $|d| < 1/2$, it becomes less accurate when d becomes larger. In contrast, the equivalence between $I_u(\lambda_j)$ and $I_{\Delta^d y}(\lambda_j)$ is valid for all value of d. Further, Shimotsu (2010) proposes a two-stage approach, which uses a tapered Local Whittle estimator (Velasco, 1999) in the first stage and a modified ELW objective function in the second stage. The 2-stage ELW method is designed to improve the performance of ELW when the mean (initial value) of the process is unknown. Unreported simulations show that both the ELW and the 2-stage ELW perform similar to the LWE method under our model setting.

and the log periodogram can be written as

$$\begin{aligned}\log I(\lambda_j) &= \log f(\lambda_j) + \log\left(\frac{I(\lambda_j)}{f(\lambda_j)}\right) \\ &\sim \log C - d\log\left\{4\sin^2\left(\frac{\lambda_j}{2}\right)\right\} + \log\left(\frac{I(\lambda_j)}{C\left(4\sin^2(\lambda/2)\right)^{-d}}\right).\end{aligned}$$

In the proposed regression, $\alpha = \log C - \eta$ and $v_j = \log\left(\frac{I(\lambda_j)}{C\left(4\sin^2(\lambda/2)\right)^{-d}}\right) + \eta$, where η is the Euler constant ($= 0.5772...$). The error term v_j becomes negligible when the attention is confined to frequencies near zero. The log periodogram estimator is shown to be $\sqrt{m}$-consistent and asymptotically normal (Geweke and Porter-Hudak, 1983; Robinson, 1995b), that is,

$$\sqrt{m}\left(\hat{d} - d\right) \to_d N\left(0, s_d^2\right),$$

with s_d being the usual OLS standard error of $\hat{d}$.

5 Monte Carlo Simulations

We examine the estimation accuracy of the fractional parameter d (long-run dynamic) and the autoregressive parameter α_1 (short-run dynamic) with various techniques. The data generating process is the ARFIMA$(1, d, 0)$ model.[5] We assume α_1 takes one of the values in $\{-0.2, 0, 0.3, 0.7, 0.9, 0.99\}$ and d takes one of the values in $\{-0.4, -0.2, 0, 0.2, 0.4\}$. We set $\sigma_u = 1$ and the intercept μ to zero. The initial value of each simulated sample path is set to the long-run mean (i.e., $\mu/(1-\alpha_1)$), which is zero under this setting. The first $5,000$ observations are discarded from each simulated sample path to minimize the impact of initial value. The number of replications is $1,000$.

The log likelihoods of the two parametric ML methods are optimized using the *fmincon* function in MATLAB with the sequential quadratic programming algorithm. The lower bound of both d and α_1 is -1 and the upper bound is 1. The objective function of the local Whittle method is optimized with the command *fminbnd* in MATLAB, as there is only one model parameter. The initial values of the optimization algorithms are the true model parameters.

Similar simulation studies have been conducted in Smith, Taylor, and Yadav (1997), Nielsen and Frederiksen (2005), and Nadarajah, Martin, and Poskitt (2021). See Table 1 for a brief summary of their Monte Carlo designs. Our Monte Carlo design extends those in the existing studies by considering more empirically relevant parameter settings. We (1) allow maximum value of α_1 to be much closer to the unit circle (i.e. 0.99 versus 0.8 in Smith, Taylor, and Yadav (1997) and Nielsen and Frederiksen (2005) and 0.9 in Nadarajah, Martin, and Poskitt (2021)), (2) choose larger sample sizes (i.e. $T = 1024$ as well as $T = 512$ and 2048 in the case of $\alpha_1 = 0.99$). We also examine the finite sample performance of $\hat{\alpha}_1$.

[5] The fractionally integrated process in (5) is simulated with the *fracdiff* function provided by Katsumi Shimotsu.

5.1 Long-run dynamic parameter d

Table 2 shows the biases and standard errors of $\hat{d}$ of the four alternative methods from $1,000$ replications. We first set the sample size to $T = 1,024$.[6] We highlight (in bold) the two cases that are most relevant to our empirical applications. There are several interesting observations. First, when α_1 takes a value close to zero, both the local Whittle and the log periodogram estimators can estimate d accurately. The biases are close to zero and the standard errors are small. Together with their asymptotic robustness property against short-run dynamics, the good finite sample property may be the reason why they have been popular in estimating d for log RVs.

However, strikingly, the two semiparametric methods have significant upward biases when the process is persistent (i.e., $\alpha_1 \geq 0.7$). The substantial upward bias in d of the semiparametric methods when $\alpha_1 = 0.8$ or 0.9 and $T = \{96, 256, 512, 576\}$ has been documented in Smith, Taylor, and Yadav (1997), Nielsen and Frederiksen (2005), and Nadarajah, Martin, and Poskitt (2021). Our results indicate that this upward bias problem continues to hold when $\alpha = 0.99$ and $T = 1,028$. The bias magnitudes of the local Whittle estimator are similar to that of the log periodogram estimator. The bias increases towards one as α_1 gets closer to unity. When $\alpha_1 = 0.9$ and $d = -0.4$, the two semiparametric methods tend to yield a positive estimate of d (with the average value of $\hat{d}$ being 0.18), implying spurious long memory. When $\alpha_1 = 0.99$, these estimators cannot distinguish between the non-stationary behavior from the long memory feature. The bias of $\hat{d}$ is close to one across all value of d, and the standard error is much larger when d increases. Further, we report simulation results for the case of $\alpha_1 = 0.99$ with $T = \{512, 1024, 2048\}$ in Table 3. The results suggest that the bias becomes only slightly smaller as the sample size increases from 512 to 2048. Our studies suggest that one should be cautious with using the two semiparametric methods, especially when there is a possibility of having persistent dynamics.

On the contrary, the two ML methods work well across all parameter settings. The performance of the Whittle method follows closely of MPL. One exception is when $\alpha_1 = 0.99$ and $d = 0.4$ where the bias of the Whittle estimation is suddenly much larger with a much greater standard error. In general, the two ML methods have smaller bias than the two semiparametric methods, especially for large values of α_1 (i.e., $\alpha_1 \geq 0.7$). Moreover, the standard error of the two ML methods varies slightly with the true value of α_1. This feature is also observed by Smith, Taylor, and Yadav (1997) and Nielsen and Frederiksen (2005). In addition, one can see from Table 3 that the standard error changes approximately by the factor of $\sqrt{T}$, as predicted by the asymptotic theory.

Tapering has been shown capable of removing deterministic time trends (e.g., Žurbenko (1979); Robinson (1986); Dahlhaus (1988); Hurvich and Ray (1995); Velasco (1999); Hurvich and Chen (2000)). Tables 4 investigates the effect of tapering on the Whittle and local Whittle estimators. The tapering methods are detailed in the Appendix. One can see from Table 4 that tapering improves the estimation accuracy of the Whittle estimator when the process is highly persistent with a strong pattern of long memory. Specifically, with the Tukey-Hanning tapering, the bias of the Whittle estimator $\hat{d} - d$ reduces from -0.206 to 0.004 when $\alpha_1 = 0.99$ and $d = 0.4$, while the performance of the Whittle estimator under other parameter settings remains roughly unchanged. In contrast, tapering does not improve

[6] We set the sample size to be the power of two to ensure the accuracy of Fourier transformation.

the estimation accuracy of the local Whittle estimator. Indeed, both biases and standard errors of the estimator increase when the parameter is estimated from tapered data.

Our Monte Carlo results are different from those in Smith, Taylor, and Yadav (1997) and Nielsen and Frederiksen (2005) in two aspects. First, we find the finite sample performance of MPL is almost always slightly better than that of the Whittle method, while Nielsen and Frederiksen (2005) find the opposite. Since the Whittle method approximates $\log L_N$ by $\log L_W$, it is reasonable to believe that the exact ML method outperforms the approximate method. Second, Smith, Taylor, and Yadav (1997) and Nielsen and Frederiksen (2005) report negative bias in $\hat{d}$ for the ML estimators in all cases they considered. While we observe negative bias when $\alpha_1 < 0.7$, we find the bias can be either negative or positive when $\alpha_1 > 0.7$.

In conclusion, as far as parameter d is concerned, when the model is correctly specified, the preferred estimation method is MPL according to our simulation studies. While the two semiparametric methods can estimate d accurately when α_1 is close to zero, they can lead to a very substantial upward bias in $\hat{d}$ when α_1 takes a value greater than 0.7. The closer α_1 to the unit circle is, the bigger the bias in $\hat{d}$.

5.2 Short-run dynamic parameter α_1

The two ML methods can not only estimate the long-run parameter d but also the short-run dynamic parameter α_1. Unlike the parametric ML methods, the semiparametric methods only estimate d. To obtain an estimate of α_1, we fit an AR(1) model to prefiltered data series using $\hat{d}$ obtained from the semiparametric methods. This two-stage approach has been used in the literature; see, for example, Andersen, Bollerslev, Diebold, and Labys (2003).

Table 5 reports the biases and standard errors of $\hat{\alpha}_1$ of the four alternative methods from $1,000$ replications. Evidently, the two parametric ML methods provide very accurate estimation results for α_1 across all parameter settings. For the semiparametric methods, when $\alpha_1 \geq 0.7$, the substantially upward biases in $\hat{d}$ lead to equally significant downward biases in $\hat{\alpha}_1$. For data simulated from a near-unity model with anti-persistent errors (for example, when $\alpha_1 = 0.9$ and $d = -0.4$), the two semiparametric methods tend to conclude that the model is ARFIMA$(1, d, 0)$ with α_1 close to 0 and $d > 0$. From Table 6, we see that when using the parametric ML methods, the estimation accuracy of α_1 improves further as the sample size expands. This is, however, not the case for the two semiparametric methods. From Table 7, tapering does not improve the estimation accuracy of α_1 for both the Whittle and local Whittle methods.

In conclusion, the two ML methods can estimate not only d but also α_1 accurately. The two semiparametric can lead to a very substantial downward bias in $\hat{\alpha}_1$ when α_1 takes a value greater than 0.7. The closer α_1 is to the unit circle, the bigger the bias in $\hat{\alpha}_1$.

5.3 Discussions

To understand further the difference between the parametric ML methods and the semiparametric methods, we show the gaps between the theoretical spectral densities of y_t and the approximate spectral densities $C\lambda^{-2d}$ (used by the semiparametric methods) under various parameter settings. The larger the distance between $f(\lambda)$ and $C\lambda^{-2d}$ is, the less accurate estimated results are expected from the semiparametric methods. Figure 1 plots the quantity $\log(f(\lambda)) - \log(C\lambda^{-2d})$ against the frequency λ. We choose the value of C such that the quantity takes value zero at frequency zero. It is obvious that the distances at frequencies close to zero are affected substantially by the autoregressive coefficient but not so much by d. This is consistent with our findings in Table 2 that the semiparametric estimators have significant biases when α_1 is close to unity, while the biases are similar across various d for a given value of α_1.

When the autoregressive coefficient α_1 is close to one, the ARFIMA$(1,d,0)$ model can be written as a local-to-unity AR(1) model with fractionally integrated errors, that is,

$$x_t = \left(1 - \frac{c}{T}\right) x_{t-1} + \upsilon_t, \text{, with } c > 0, \upsilon_t = \sigma (1-L)^{-d} \varepsilon_t \text{ and } \varepsilon_t \sim_{iid} N(0,1). \tag{18}$$

It can be rewritten as

$$\begin{aligned} (1-L) x_t &= -\frac{c}{T} x_{t-1} + \sigma (1-L)^{-d} \varepsilon_t, \\ x_t &= -\frac{c}{T} (1-L)^{-1} x_{t-1} + \sigma (1-L)^{-1-d} \varepsilon_t. \end{aligned} \tag{19}$$

Denote the second quantity of (19) by $z_t = \sigma (1-L)^{-1-d} \varepsilon_t$, which is a fractionally integrated process with fractional parameter $1+d$. The first quantity, which is a function of c and x_{t-1}, represents deviations of x_t from z_t. When $c = 0$, it disappears and $x_t = z_t$. It is, therefore, not surprising that in finite samples the estimated memory parameter is close to $1+d$ (the fractional parameter of z_t) instead of d and the bias is close to one when $c \to 0$ or when α_1 of the ARFIMA$(1,d,0)$ model is close to unity.

Next, we analyze the limiting properties of the two quantities on the right-hand side of (19). From Tanaka (2013), as $T \to \infty$, we have

$$\frac{\delta_H \Gamma(H+0.5)}{T^H \sigma} x_{\lfloor Tr \rfloor} \Rightarrow J_c^H(r), \tag{20}$$

where $H = d + 0.5$, $\delta_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+0.5)\Gamma(2-2H)}}$, $J_c^H(r) := \exp(cr) \int_0^r \exp(-cs)\, dB_H(s)$ is an fOU process, $B_H(s)$ is the fBm with Hurst parameter H, $\lfloor . \rfloor$ denotes the integer part of the argument, and $r \in [0,1]$. The z_t process can be rewritten as,

$$z_t = z_{t-1} + \sigma (1-L)^{-d} \varepsilon_t,$$

which can be viewed as a special case of (18) with $c = 0$. Therefore,

$$\frac{\delta_H \Gamma(H+1/2)}{T^H \sigma} z_{\lfloor Tr \rfloor} \Rightarrow B_H(r). \tag{21}$$

Combining the above two results, we know that the first quantity of (19) has the following limiting property:

$$-\frac{c\delta_H \Gamma(H+1/2)}{T^{1+H}\sigma}(1-L)^{-1} x_{t-1} \Rightarrow J_c^H(r) - B_H(r).$$

That is, $-\frac{c}{T}(1-L)^{-1} x_{t-1} = O_p\left(T^H\right)$ which is of the same order of magnitude as the second quantity. This result implies that the impact of the first quantity will not disappear as the sample size increases and one cannot obtain a consistent estimate of d from Δx_t (by taking first difference) when $c \neq 0$.

5.4 Distributions of $\hat{\alpha}_1 - \alpha_1$

It is well known that the least squares estimator of the autoregressive parameter of an AR(1) model does not have the standard asymptotics (i.e., it does not converge to a normal distribution) when the process is local-to-unity; see for example Phillips (1987). Wang, Xiao, and Yu (2020) and Yu (2021) extend the asymptotic theory from the case of the weakly dependent errors to the cases of the strongly dependent errors with and without normality, respectively. To get some ideas on the finite sample distributions of the ML estimators of α_1 when the data are simulated from ARFIMA$(1,d,0)$, we plot in Figure 2 the kernel density of $\hat{\alpha}_1 - \alpha_1$ for the MPL and Whittle methods obtained from $1,000$ replications. When simulating data, we assume α_1 takes one of the values in $\{-0.2, 0, 0.3, 0.7, 0.9, 0.99\}$ and d takes one of the values in $\{-0.4, 0, 0.4\}$. The sample size is $1,024$. Note that the ARFIMA$(1,d,0)$ model reduces to a local-to-unity AR(1) model when $d = 0$.

It is obvious from the graphs that the distribution of $\hat{\alpha}_1 - \alpha_1$ deviates substantially from the normal distribution when α_1 is close to unity. In this case, the distribution of $\hat{\alpha}_1$ becomes quite asymmetric and is not centered at 0. This observation applies to both parametric ML methods and is consistent across different values of d. To the best of our knowledge, the asymptotic of $\hat{\alpha}_1$ for any of the two ML methods in the context of local-to-unity ARFIMA$(1,d,0)$ model has not been developed in the literature. Our conjecture is that the asymptotic theory of $\hat{\alpha}_1$ is nonstandard.

6 Empirical Applications to RV

In this section, we investigate the dynamics of log RVs across a broad set of financial assets, including five market index futures or ETFs, five industry index ETFs, and three exchange rate futures, over the past decade. We consider the maximum available sample period after 5 January 2010. The end date is 25 May 2021.

6.1 Data

The QML realized volatility data are provided by the Risk Lab[7] and computed using transaction prices sampled at the highest available frequency. In total, thirteen assets are considered. The assets under consideration are listed in Table 8, along with the starting date, number of observations, and summary statistics (such as the sample mean, sample standard deviation, sample skewness, and sample kurtosis) of RV and log RV of each data series. Most of the data series start from 5 January 2010. The sample size ranges from 1305 to 2757. Figure 3 displays the dynamics of each log RV series. It is obvious from both Table 8 and Figure 3 that exchange rate futures are in generally less volatile. While the sample skewness and sample kurtosis suggest that the distributions of RV are very far away from the normal distribution, the distributions of log RV are much closer. This finding is well known in the literature; see Andersen, Bollerslev, Diebold, and Labys (2001), Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2003).

Figure 5 presents the ACFs of the log RVs. As expected, the log RV series are highly persistent. The ACF remains sizable at the 100^{th} lag in all cases. Interestingly, the ACFs of the log RV of the exchange rate futures (Figure 5(k)-(m)) seem to decay more slowly than the other financial assets.

6.2 Estimation results

We apply the parametric and semiparametric methods to estimate an ARFIMA$(1,d,0)$ model for the log RVs. The setting of initial values is crucial for the parametric methods as the likelihood surfaces are not concave. Between the two ML methods, we first implement the Whittle method, where we use a grid searching method to choose the 'optimal' initial values of d and α_1. For d the grids range from -0.499 to 0.499 with an increment of 0.005. For α_1 the grids range from -0.999 to 0.999 with an increment of 0.001. We evaluate the Whittle log-likelihood for all possible combinations of d and α_1. The pair that produces the highest log-likelihood value is taken as our initial values for the Whittle method. The estimated parameter values from the Whittle method are taken as the initial values of the MPL method. The optimization algorithms and settings are the same as those in the simulations.

Figure 4 shows two examples of the Whittle log-likelihood surface. We remove log-likelihood values that are smaller than certain thresholds to obtain better visualization of the surface at the peak. Interestingly, the log-likelihood surface is bi-modal for all the log RV series examined in this paper, with one at the left corner ($\alpha_1 \to 1$ and $d < 0$) and one at the right corner ($|\alpha_1| \to 0$ and $d > 0.35$). One would conclude that the data series is persistent and rough if the left modal is higher than the right one; otherwise, it has a long memory with strong mean-reversion. We have seen both conclusions in the RV literature. For example, with the LPE method, Andersen, Bollerslev, Diebold, and Labys (2001) finds the log RV of 30 Dow Jones Industrial Average firms over the period from 1993 to 1998 have a long memory, with the maximum estimate of d being 0.416 and minimum being 0.263. Using LPE again, Andersen, Bollerslev, Diebold, and Labys (2003) concludes that the estimated d for the log RVs of three exchange rates (US dollar, Deutsche Mark, and Japanese Yen) over the period 1986-1996 are between

[7] https://dachxiu.chicagobooth.edu/#risklab.

0.38 and 0.43. On the other hand, Wang, Xiao, and Yu (2019), Fukasawa, Takabatake, and Westphal (2021), and Bolko, Christensen, Pakkanen, and Veliyev (2021) examine the log RVs of stock indices over more recent periods using, respectively, a change of frequency approach, a QML estimation method, and GMM. All three studies conclude that RVs are rough (i.e., $d < 0$) instead of long memory.

Here, we consider the most recent sample period spanning over a decade. The ARFIMA$(1, d, 0)$ is fitted to each log RV series using the two parametric methods and two semiparametric methods. While simulation studies conclude that the ML methods are preferred, we also report estimation results from the semiparametric methods for comparison. The estimated parameters are reported in Table 9. Interestingly, the parameter methods suggest that log RVs of the market index futures or EFTs and industry EFTs are highly persistent and rough, with a close-to-unity $\widehat{\alpha}_1$ and a negative $\widehat{d}$. This result is consistent with the findings of Liu, Shi, and Yu (2020); Fukasawa, Takabatake, and Westphal (2021); Bolko, Christensen, Pakkanen, and Veliyev (2021) where volatility is treated as latent but assumed to follow an AR process with fractionally integrated errors or fGns.

On the contrary, the semiparametric methods suggest that the memory parameter is between 0.52 and 0.64, implying that the log RV series has a long memory. The contradicted results between the parametric and semiparametric methods are consistent with our findings from the simulations. The semiparametric methods tend to significantly overestimate the memory parameter when the autoregressive coefficient is close to unity.

Interestingly, unlike the other assets, the parametric methods suggest that the log RVs of exchange rate futures have long memory with $\widehat{d}$ between 0.437 and 0.472 and $\widehat{\alpha}_1$ being a small negative number. This result is consistent with the finding of Andersen, Bollerslev, Diebold, and Labys (2003) for the log RV of three exchange rates. The estimated fractional parameters from the semiparametric methods are around 0.6.

7 Conclusion

In this paper, we first examine the finite sample properties of four alternative methods in estimating the ARFIMA$(1, d, 0)$ model, including the two parametric ML (MPL and Whittle) methods and two semiparametric (local Whittle and log periodogram regression) methods. Special attention is paid to the part of the parameter space where the fractional differencing parameter d is negative and the short-run parameter α_1 is close to the unit circle. This choice of parameter setting is motivated by the empirical finding in the recent literature that documents evidence of rough volatility; for example, see Gatheral, Jaisson, and Rosenbaum (2018) and Bayer, Friz, and Gatheral (2016) for evidence in the risk-neutral measure, and Wang, Xiao, and Yu (2019), Liu, Shi, and Yu (2020), Fukasawa, Takabatake, and Westphal (2021), Bolko, Christensen, Pakkanen, and Veliyev (2021) for evidence in the physical measure. Via simulations, we find that the two ML methods have better finite sample performances than the two semiparametric methods. This is not surprising as when the model is correctly specified, the ML methods use more information than the two semiparametric methods. Moreover, when data are simulated from the ARFIMA$(1, d, 0)$ model with a negative d and a near-unity α_1, the two semiparamet-

ric methods lead to a false conclusion of a long memory error term.

When we apply the two ML methods to the log RVs of thirteen financial assets, interesting empirical results are revealed. For the log RVs of the market indices (futures or EFTs) and industry indices (EFTs), the two ML methods find evidence of a near-unity short-run dynamic and an anti-persistent error term, consistent with the recent literature on rough volatility. In contrast, the two semiparametric methods always lead to the conclusion of long memory for the log volatilities. The estimated d is positive, and the autoregressive coefficient is near zero. When we examine the log-likelihood surface of the Whittle estimator, we find the log-likelihood value at the left corner ($\alpha_1 \to 1$ and $d < 0$) is higher than at the right corner ($|\alpha_1| \to 0$ and $d > 0.35$) for those assets.

For the log RVs of the exchange rate futures, the two ML methods find evidence of a weak short-run behavior and a long memory error term, consistent with the finding of Andersen, Bollerslev, Diebold, and Labys (2003). The log-likelihood value at the left corner ($\alpha_1 \to 1$ and $d < 0$) of the Whittle log-likelihood surface is lower than that at the right corner ($|\alpha_1| \to 0$ and $d > 0.35$). In sum, the log volatilities of exchange rate futures display a different dynamic from the other financial assets considered.

Should one use a long memory model or a rough model for log realized volatility? The answer to this important question is found in the title of our article. That is which candidate model should be used depends on assets at hand and a careful analysis is needed before a choice is made.

References

An, S. and P. Bloomfield (1993). Cox and reid's modification in regression models with correlated errors. *Department of Statistics, North Carolina State University, Raleigh*.

Andersen, T. G. and T. Bollerslev (1997). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance 52*(3), 975–1005.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*(1), 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association 96*(453), 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and J. Wu (2005). A framework for exploring the macroeconomic determinants of systematic risk. *American Economic Review 95*(2), 398–404.

Bayer, C., P. Friz, and J. Gatheral (2016). Pricing under rough volatility. *Quantitative Finance 16*(6), 887–904.

Bertelli, S. and M. Caporin (2002). A note on calculating autocovariances of long-memory processes. *Journal of Time Series Analysis 23*(5), 503–508.

Bolko, A. E., K. Christensen, M. S. Pakkanen, and B. Veliyev (2021). Roughness in spot variance? A GMM approach for estimation of fractional log-normal stochastic volatility models using realized measures. *arXiv preprint arXiv:2010.04610*.

Brockwell, P. J. and R. A. Davis (2009). *Time Series: Theory and Methods*. Springer Science & Business Media.

Christoffersen, P., B. Feunou, K. Jacobs, and N. Meddahi (2014). The economic value of realized volatility: Using high-frequency returns for option valuation. *Journal of Financial and Quantitative Analysis 49*(3), 663–697.

Christoffersen, P. F. and F. X. Diebold (2000). How relevant is volatility forecasting for financial risk management? *Review of Economics and Statistics 82*(1), 12–22.

Chung, C.-F. (1994). A note on calculating the autocovariances of the fractionally integrated arma models. *Economics Letters 45*(3), 293–297.

Coursol, J. and D. Dacunha-Castelle (1982). Remarks on the approximation of the likelihood function of a stationary gaussian process. *Theory of Probability & Its Applications 27*(1), 162–167.

Cox, D. R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society: Series B (Methodological) 49*(1), 1–18.

Da, R. and D. Xiu (2021). When moving-average models meet high-frequency data: Uniform inference on volatility. *Econometrica, forthcoming*.

Dahlhaus, R. (1988). Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *The Annals of Statistics 16*(2), 808–841.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics 17*(4), 1749–1766.

Fleming, J., C. Kirby, and B. Ostdiek (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics 67*(3), 473–509.

Fox, R. and M. S. Taqqu (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics 14*(2), 517–532.

Fukasawa, M., T. Takabatake, and R. Westphal (2021). Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Mathematical Finance, forthcoming*.

Gatheral, J., T. Jaisson, and M. Rosenbaum (2018). Volatility is rough. *Quantitative Finance 18*(6), 933–949.

Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis 4*(4), 221–238.

Giraitis, L. and D. Surgailis (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of whittle's estimate. *Probability Theory and Related Fields 86*(1), 87–104.

Granger, C. W. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis 1*(1), 15–29.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *Journal of Applied Probability 10*(1), 130–145.

Hauser, M. A. (1999). Maximum likelihood estimators for arma and arfima models: A monte carlo study. *Journal of Statistical Planning and Inference 80*(1-2), 229–255.

Hosking, J. R. (1981). Fractional differencing. *Biometrika 68*(1), 165–76.

Hurvich, C. M. and W. W. Chen (2000). An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis 21*(2), 155–180.

Hurvich, C. M. and B. K. Ray (1995). Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis 16*(1), 17–41.

Jacod, J., Y. Li, and X. Zheng (2017). Statistical properties of microstructure noise. *Econometrica 85*(4), 1133–1174.

Jacod, J., Y. Li, and X. Zheng (2019). Estimating the integrated volatility with tick observations. *Journal of Econometrics 208*(1), 80–100.

Künsch, H. (1987). Statistical aspects of self-similar processes. In *Proceedings of the First Congress of the Bernoulli Society, 1987*.

Lieberman, O. (2005). On plug-in estimation of long memory models. *Econometric Theory 21*(2), 431–454.

Liu, X., S. Shi, and J. Yu (2020). Persistent and rough volatility. *Available at SSRN 3724733*.

Magdalinos, T. (2012). Mildly explosive autoregression under weak and strong dependence. *Journal of Econometrics 169*(2), 179–187.

Nadarajah, K., G. M. Martin, and D. Poskitt (2021). Optimal bias correction of the log-periodogram estimator of the fractional parameter: A jackknife approach. *Journal of Statistical Planning and Inference 211*, 41–79.

Nielsen, M. Ø. and P. H. Frederiksen (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews 24*(4), 405–443.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika 74*(3), 535–547.

Robinson, P. M. (1986). On the errors-in-variables problem for time series. *Journal of Multivariate Analysis 19*(2), 240–250.

Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics 23*(5), 1630–1661.

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics 23*(4), 1048–1072.

Robinson, P. M. (2003). *Time Series with Long Memory*. Oxford University Press, UK.

Shimotsu, K. (2010). Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory 26*(2), 501–540.

Shimotsu, K. and P. C. B. Phillips (2006). Local whittle estimation of fractional integration and some of its variants. *Journal of Econometrics 130*(2), 209–233.

Smith, J., N. Taylor, and S. Yadav (1997). Comparing the bias and misspecification in arfima models. *Journal of Time Series Analysis 18*(5), 507–527.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics 53*(1-3), 165–188.

Tanaka, K. (2013). Distributions of the maximum likelihood and minimum contrast estimators associated with the fractional Ornstein–Uhlenbeck process. *Statistical Inference for Stochastic Processes 16*(3), 173–192.

Tao, Y., P. C. B. Phillips, and J. Yu (2019). Random coefficient continuous systems: Testing for extreme sample path behavior. *Journal of Econometrics 209*(2), 208–237.

Tukey, J. W. (1967). An intoroduction to the calculation of numerical spectrum analysis. *Spectra Analysis of Time Series*, 25–46.

Varneskov, R. T. (2017). Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels. *Econometric Theory 33*(6), 1457.

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis 20*(1), 87–127.

Wang, X., W. Xiao, and J. Yu (2019). Estimation and inference of the fractional continuous-time model with discrete-sampled data. *SMU Economics and Statistics Working Paper Series, Paper No. 17-2019*.

Wang, X., W. Xiao, and J. Yu (2020). Asymptotic properties of least squares estimator in local to unity processes with fractional gaussian noises. *SMU Economics and Statistics Working Paper Series, Paper No. 27-2020*.

Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv för matematik 2*(5), 423–434.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika 41*(3-4), 434–449.

Yajima, Y. (1985). On estimation of long-memory time series models. *Australian Journal of Statistics 27*(3), 303–320.

Yu, J. (2021). Latent local-to-unity models. *SMU Economics and Statistics Working Paper Series, Paper No. 04-2021*.

Žurbenko, I. G. (1979). On the efficiency of estimates of a spectral density. *Scandinavian Journal of Statistics 6*(2), 49–56.

A Tapering

One pitfall of the periodogram $I(\lambda_j)$ is that there is leakage effect. In finite samples, when there are high peaks in the spectrum, the nonparametric estimator $I(\lambda_j)$ might significantly overestimate the spectrum at other frequencies and fail to discover spectrums with low peaks.

A.1 Whittle Estimator with Tapering

Dahlhaus (1988) proposes using tappering adapted from nonparametric spectral density estimation (Tukey, 1967) for the Whittle estimator. A tapered series is define as

$$y_t^T = h_t y_t,$$

where h_t is the data taper satisfying certain time series properties (Dahlhaus, 1988). The taperred periodogram is

$$I_T(\lambda_j) = \frac{1}{2\pi \sum_{t=0}^{T-1} h_t^2} \left| \sum_{t=0}^{T} h_t y_t \exp(-it\lambda_j) \right|^2.$$

Replacing $I(\lambda_j)$ in the Whittle estimator (13) by $I_T(\lambda_j)$ yields the tappered Whittle estimator. Dahlhaus (1988) show that the tappered Whittle estimator is $\sqrt{T}$-consistent and asymptotically normal.

There are many tapers satisfying the conditions outlined in Dahlhaus (1988). One example is the Tukey-Hanning taper specified as

$$h_\rho(x) = \begin{cases} \frac{1}{2}[1 - \cos(2\pi x/\rho)] & x \in [0, \rho/2) \\ 1 & x \in [\rho/2, 1/2] \\ h_\rho(1-x) & x \in (1/2, 1] \end{cases}$$

and $h_t = h_\rho(t/T)$. For practical implementation, one could set $\rho = T^{-\kappa/3}$ with $\kappa \in [0, 1/2)$. Here, we set $\kappa = 1/4$.

A.2 Local Whittle Estimator with Tapering

One popular tapering method in the local Whittle content is proposed by Velasco (1999). For each positive integer p, there is a Kolmogorov taper which is of order p in the sense of Velasco (1999). A taper with order p, if applied to the raw data, yields a tapered periodogram that is invariant to polynomial trends of order $p-1$, provided that the periodogram is evaluated on the grid λ_{ip} with $i = 1, 2, \ldots, \lfloor m/p \rfloor$. The objective function of the tapered LW estimator becomes

$$(\hat{C}_p, \hat{d}_p) = \arg\max_{C,d} \frac{p}{m} \sum_{i=1}^{\lfloor m/p \rfloor} \left[-\log C + 2d \log \lambda_{ip} - \frac{1}{C} \lambda_{ip}^{2d} I^T(\lambda_{ip}) \right], \tag{22}$$

where

$$\begin{aligned}\hat{d}_p &= \arg\max\left\{-\log\hat{C}_p(d)+2d\frac{p}{m}\sum_{i=1}^{\lfloor m/p\rfloor}\log\lambda_{ip}\right\},\\ \hat{C}_p(d) &= \frac{p}{m}\sum_{i=1}^{\lfloor m/p\rfloor}\lambda_{ip}^{2d}I_T(\lambda_{ip}).\end{aligned}$$

The discrete sums include only frequencies λ_{ip} with $i=1,2,\ldots,\lfloor m/p\rfloor$.

The tapered LW estimator is asymptotic normal with a variance of $p\Phi/(4m)$, where

$$\Phi=\lim_{T\to\infty}\left(\sum_{t=1}^{T}h_t^2\right)^{-2}\sum_{k=0,p,2p,\ldots}^{n-p}\left\{\sum_{t=1}^{n}h_t^2\cos(t\lambda_k)\right\}^2.$$

Suppose we employ the full cosine bell taper (Tukey, 1967)

$$h_t=0.5\left[1-\cos\left(\frac{2\pi t}{T}\right)\right]$$

and regard this taper as of order $p=3$, the tapered LW estimator is asymptotic normal with variance $pm\Phi/4$ with $\Phi=1$, when $\mu=0$ and $d<1.5$, . However, if we use all the Fourier frequencies from λ_2 to λ_m (i.e., $p=1$), then $\Phi=35/18$. In the simulation studies, we use the cosine bell taper with $p=3$. While the tapered local Whittle methods are invariant to trends and asymptotically normal, they lead to inflated asymptotic variance of the estimator.

B Tables and Figures

Table 1: Existing Monte Carlo Studies

Paper	Relevant Tables	Relevant Estimation Methods	Sample Size
Smith, Taylor, and Yadav (1997)	Tables I and VI	ML and LPE ($m=T^{0.5}$, $T^{0.6}$,$T^{0.7}$)	256
Nielsen and Frederiksen (2005)	Tables 8 and 9	Exact ML, MPL, Whittle, Conditional ML LWE and LPE ($m=T^{0.5}$, $T^{0.65}$)	$128, 256, 512$
Nadarajah, Martin, and Poskitt (2021)	Tables 6 and 7	ML and LPE ($m=T^{0.65}$)	$96, 576$

Table 2: Biases and standard errors (in brackets) of $\hat{d}$ when $T = 1,024$.

	MPL	Whittle	LWE	LPE
		$\alpha_1 = -0.2$		
d=-0.4	-0.002 (0.04)	-0.003 (0.04)	0.001 (0.06)	0.006 (0.08)
d=-0.2	-0.004 (0.04)	-0.010 (0.04)	-0.010 (0.06)	-0.006 (0.07)
d=0	-0.005 (0.05)	-0.012 (0.05)	-0.012 (0.06)	-0.008 (0.07)
d=0.2	-0.003 (0.04)	-0.022 (0.13)	-0.007 (0.06)	0.000 (0.07)
d=0.4	-0.004 (0.04)	-0.011 (0.08)	-0.008 (0.06)	0.000 (0.08)
		$\alpha_1 = 0$		
d=-0.4	-0.003 (0.04)	-0.007 (0.04)	0.005 (0.06)	0.011 (0.08)
d=-0.2	-0.005 (0.04)	-0.013 (0.04)	-0.004 (0.06)	-0.001 (0.07)
d=0	-0.005 (0.04)	-0.015 (0.04)	-0.006 (0.06)	-0.002 (0.07)
d=0.2	-0.004 (0.04)	-0.013 (0.04)	-0.002 (0.06)	0.006 (0.07)
d=0.4	**-0.006** (0.06)	**-0.010** (0.04)	**-0.003** (0.06)	**0.005** (0.08)
		$\alpha_1 = 0.3$		
d=-0.4	-0.015 (0.09)	-0.025 (0.09)	0.026 (0.06)	0.032 (0.08)
d=-0.2	-0.012 (0.07)	-0.036 (0.10)	0.018 (0.06)	0.022 (0.07)
d=0	-0.014 (0.07)	-0.037 (0.09)	0.017 (0.06)	0.020 (0.07)
d=0.2	-0.010 (0.07)	-0.029 (0.08)	0.021 (0.06)	0.028 (0.07)
d=0.4	-0.014 (0.07)	-0.033 (0.10)	0.020 (0.06)	0.028 (0.07)
		$\alpha_1 = 0.7$		
d=-0.4	0.005 (0.08)	-0.029 (0.09)	0.189 (0.06)	0.185 (0.07)
d=-0.2	-0.002 (0.09)	-0.035 (0.09)	0.185 (0.06)	0.180 (0.07)
d=0	0.001 (0.09)	-0.033 (0.09)	0.183 (0.06)	0.179 (0.07)
d=0.2	-0.003 (0.10)	-0.038 (0.09)	0.182 (0.06)	0.180 (0.07)
d=0.4	-0.001 (0.09)	-0.033 (0.09)	0.187 (0.06)	0.184 (0.07)
		$\alpha_1 = 0.9$		
d=-0.4	0.008 (0.06)	0.000 (0.05)	0.577 (0.07)	0.528 (0.07)
d=-0.2	0.007 (0.06)	0.000 (0.05)	0.578 (0.07)	0.529 (0.07)
d=0	0.006 (0.06)	-0.002 (0.05)	0.571 (0.07)	0.523 (0.07)
d=0.2	0.004 (0.05)	-0.007 (0.05)	0.576 (0.07)	0.526 (0.07)
d=0.4	0.005 (0.05)	-0.037 (0.06)	0.567 (0.07)	0.520 (0.08)
		$\alpha_1 = 0.99$		
d=-0.4	**0.003** (0.03)	**0.006** (0.03)	**0.950** (0.06)	**0.946** (0.08)
d=-0.2	0.001 (0.03)	0.006 (0.03)	0.950 (0.06)	0.944 (0.07)
d=0	0.002 (0.03)	0.004 (0.03)	0.948 (0.06)	0.940 (0.07)
d=0.2	0.003 (0.03)	-0.045 (0.05)	0.918 (0.07)	0.911 (0.08)
d=0.4	0.000 (0.03)	-0.206 (0.12)	0.810 (0.11)	0.785 (0.13)

Table 3: Biases and standard errors of the estimated memory parameter $\hat{d}$: $T = \{512, 1024, 2048\}$ and $\alpha = 0.99$

	Parametric Methods		Semi-parametric Methods	
	MPL	Whittle	LWE	LPE
		$T = 512$		
d=-0.4	**0.009** (0.06)	**0.014** (0.04)	**0.967** (0.08)	**0.971** (0.09)
d=-0.2	0.004 (0.04)	0.015 (0.04)	0.970 (0.08)	0.980 (0.10)
d=0	0.004 (0.04)	0.007 (0.04)	0.959 (0.08)	0.964 (0.10)
d=0.2	0.003 (0.04)	-0.063 (0.06)	0.912 (0.08)	0.910 (0.11)
d=0.4	0.007 (0.05)	-0.229 (0.12)	0.787 (0.12)	0.765 (0.15)
		$T = 1024$		
d=-0.4	**0.003** (0.03)	**0.006** (0.03)	**0.950** (0.06)	**0.946** (0.08)
d=-0.2	0.001 (0.03)	0.006 (0.03)	0.950 (0.06)	0.944 (0.07)
d=0	0.002 (0.03)	0.004 (0.03)	0.948 (0.06)	0.940 (0.07)
d=0.2	0.003 (0.03)	-0.045 (0.05)	0.918 (0.07)	0.911 (0.08)
d=0.4	0.000 (0.03)	-0.206 (0.12)	0.810 (0.11)	0.785 (0.13)
		$T = 2048$		
d=-0.4	**0.001** (0.02)	**0.002** (0.02)	**0.933** (0.05)	**0.915** (0.06)
d=-0.2	0.001 (0.02)	0.003 (0.02)	0.934 (0.05)	0.913 (0.06)
d=0	0.001 (0.02)	0.001 (0.02)	0.932 (0.05)	0.912 (0.06)
d=0.2	0.001 (0.02)	-0.031 (0.04)	0.913 (0.05)	0.893 (0.06)
d=0.4	0.002 (0.02)	-0.171 (0.11)	0.840 (0.09)	0.809 (0.11)

Table 4: Biases and standard errors of the estimated memory parameter $\hat{d} - d$: the effect of tapering $T = 1024$ and $\alpha = \{-0.2, 0, 0.3, 0.7, 0, 0.9, 0.99\}$

$\hat{d} - d$	Whittle	Whittle (taper)	LWE	LWE (taper)
		$\alpha = -0.2$		
d=-0.4	-0.003 (0.04)	-0.012 (0.04)	0.001 (0.06)	-0.014 (0.12)
d=-0.2	-0.010 (0.04)	-0.013 (0.04)	-0.010 (0.06)	-0.019 (0.12)
d=0	-0.012 (0.05)	-0.015 (0.06)	-0.012 (0.06)	-0.018 (0.12)
d=0.2	-0.022 (0.13)	-0.023 (0.13)	-0.007 (0.06)	-0.021 (0.12)
d=0.4	-0.011 (0.08)	-0.019 (0.12)	-0.008 (0.06)	-0.023 (0.11)
		$\alpha = 0$		
d=-0.4	-0.007 (0.04)	-0.016 (0.05)	0.005 (0.06)	-0.007 (0.12)
d=-0.2	-0.013 (0.04)	-0.017 (0.05)	-0.004 (0.06)	-0.013 (0.12)
d=0	-0.015 (0.04)	-0.021 (0.08)	-0.006 (0.06)	-0.011 (0.12)
d=0.2	-0.013 (0.04)	-0.013 (0.05)	-0.002 (0.06)	-0.015 (0.12)
d=0.4	-0.010 (0.04)	-0.009 (0.05)	-0.003 (0.06)	-0.017 (0.11)
		$\alpha = 0.3$		
d=-0.4	-0.025 (0.09)	-0.045 (0.12)	0.026 (0.06)	0.019 (0.12)
d=-0.2	-0.036 (0.10)	-0.047 (0.12)	0.018 (0.06)	0.014 (0.12)
d=0	-0.037 (0.09)	-0.045 (0.11)	0.017 (0.06)	0.015 (0.12)
d=0.2	-0.029 (0.08)	-0.033 (0.10)	0.021 (0.06)	0.011 (0.12)
d=0.4	-0.033 (0.10)	-0.024 (0.10)	0.020 (0.06)	0.009 (0.11)
		$\alpha = 0.7$		
d=-0.4	-0.029 (0.09)	-0.036 (0.10)	0.189 (0.06)	0.203 (0.12)
d=-0.2	-0.035 (0.09)	-0.037 (0.10)	0.185 (0.06)	0.197 (0.11)
d=0	-0.033 (0.09)	-0.035 (0.10)	0.183 (0.06)	0.197 (0.12)
d=0.2	-0.038 (0.09)	-0.032 (0.10)	0.182 (0.06)	0.194 (0.12)
d=0.4	-0.033 (0.09)	-0.020 (0.11)	0.187 (0.06)	0.200 (0.12)
		$\alpha = 0.9$		
d=-0.4	0.000 (0.05)	0.002 (0.06)	0.577 (0.07)	0.615 (0.12)
d=-0.2	0.000 (0.05)	0.004 (0.07)	0.578 (0.07)	0.609 (0.12)
d=0	-0.002 (0.05)	0.000 (0.06)	0.571 (0.07)	0.605 (0.13)
d=0.2	-0.007 (0.05)	-0.003 (0.06)	0.576 (0.07)	0.616 (0.13)
d=0.4	-0.037 (0.06)	0.003 (0.06)	0.567 (0.07)	0.605 (0.13)
		$\alpha = 0.99$		
d=-0.4	0.006 (0.03)	0.005 (0.03)	0.950 (0.06)	0.971 (0.12)
d=-0.2	0.006 (0.03)	0.004 (0.03)	0.950 (0.06)	0.968 (0.12)
d=0	0.004 (0.03)	0.006 (0.03)	0.948 (0.06)	0.971 (0.12)
d=0.2	-0.045 (0.05)	0.006 (0.03)	0.918 (0.07)	0.969 (0.12)
d=0.4	-0.206 (0.12)	0.004 (0.03)	0.810 (0.11)	0.971 (0.11)

Table 5: Biases and standard errors (in brackets) of $\hat{\alpha}_1$ when $T = 1,024$.

	MPL	Whittle	LWE	LPE
		$\alpha = -0.2$		
d=-0.4	0.002 (0.04)	0.002 (0.05)	0.002 (0.06)	-0.001 (0.07)
d=-0.2	0.002 (0.05)	0.006 (0.05)	0.008 (0.06)	0.007 (0.07)
d=0	0.004 (0.06)	0.009 (0.06)	0.011 (0.06)	0.010 (0.07)
d=0.2	0.004 (0.04)	0.022 (0.14)	0.009 (0.06)	0.005 (0.07)
d=0.4	0.003 (0.04)	0.010 (0.09)	0.008 (0.06)	0.003 (0.07)
		$\alpha = 0$		
d=-0.4	0.003 (0.05)	0.006 (0.05)	-0.003 (0.07)	-0.007 (0.08)
d=-0.2	0.003 (0.05)	0.010 (0.06)	0.003 (0.07)	0.001 (0.08)
d=0	0.004 (0.05)	0.012 (0.05)	0.006 (0.06)	0.004 (0.08)
d=0.2	0.005 (0.05)	0.012 (0.05)	0.004 (0.06)	-0.002 (0.08)
d=0.4	**0.005** (0.06)	**0.009** (0.05)	**0.003** (0.07)	**-0.003** (0.08)
		$\alpha = 0.3$		
d=-0.4	0.014 (0.09)	0.023 (0.10)	-0.028 (0.07)	-0.033 (0.08)
d=-0.2	0.009 (0.08)	0.032 (0.10)	-0.022 (0.07)	-0.025 (0.08)
d=0	0.012 (0.08)	0.034 (0.10)	-0.019 (0.07)	-0.022 (0.08)
d=0.2	0.010 (0.07)	0.028 (0.08)	-0.022 (0.07)	-0.028 (0.08)
d=0.4	0.012 (0.08)	0.031 (0.11)	-0.022 (0.07)	-0.029 (0.08)
		$\alpha = 0.7$		
d=-0.4	-0.010 (0.08)	0.017 (0.08)	-0.180 (0.06)	-0.176 (0.07)
d=-0.2	-0.005 (0.09)	0.022 (0.08)	-0.175 (0.06)	-0.171 (0.07)
d=0	-0.005 (0.08)	0.021 (0.07)	-0.171 (0.06)	-0.168 (0.07)
d=0.2	-0.004 (0.09)	0.023 (0.08)	-0.172 (0.06)	-0.171 (0.07)
d=0.4	-0.004 (0.08)	0.018 (0.08)	-0.174 (0.06)	-0.172 (0.07)
		$\alpha = 0.9$		
d=-0.4	-0.008 (0.04)	-0.005 (0.04)	-0.536 (0.08)	-0.479 (0.09)
d=-0.2	-0.007 (0.04)	-0.005 (0.03)	-0.535 (0.08)	-0.479 (0.09)
d=0	-0.007 (0.04)	-0.005 (0.03)	-0.528 (0.08)	-0.474 (0.09)
d=0.2	-0.005 (0.03)	-0.003 (0.03)	-0.530 (0.08)	-0.472 (0.09)
d=0.4	-0.006 (0.03)	0.005 (0.03)	-0.509 (0.08)	-0.455 (0.09)
		$\alpha = 0.99$		
d=-0.4	**-0.003** (0.01)	**-0.006** (0.01)	**-0.943** (0.07)	**-0.937** (0.09)
d=-0.2	-0.003 (0.01)	-0.006 (0.01)	-0.943 (0.07)	-0.935 (0.09)
d=0	-0.003 (0.01)	-0.006 (0.01)	-0.937 (0.08)	-0.926 (0.09)
d=0.2	-0.002 (0.01)	-0.004 (0.01)	-0.931 (0.19)	-0.914 (0.21)
d=0.4	-0.002 (0.01)	0.001 (0.01)	-1.011 (0.46)	-0.912 (0.48)

Table 6: Biases and standard errors of the estimated autoregressive parameter $\hat{\alpha}_1$: $T = \{512, 1024, 2048\}$ and $\alpha = 0.99$

	Parametric Methods		Semi-parametric Methods	
	MPL	Whittle	LWE	LPE
		$T = 512$		
d=-0.4	**-0.007** (0.04)	**-0.013** (0.01)	**-0.958** (0.09)	**-0.960** (0.10)
d=-0.2	-0.005 (0.01)	-0.014 (0.01)	-0.961 (0.09)	-0.968 (0.11)
d=0	-0.005 (0.01)	-0.013 (0.01)	-0.949 (0.11)	-0.952 (0.13)
d=0.2	-0.005 (0.01)	-0.010 (0.01)	-0.938 (0.27)	-0.920 (0.29)
d=0.4	-0.005 (0.01)	-0.005 (0.01)	-1.010 (0.51)	-0.896 (0.52)
		$T = 1024$		
d=-0.4	**-0.003** (0.01)	**-0.006** (0.01)	**-0.943** (0.07)	**-0.937** (0.09)
d=-0.2	-0.003 (0.01)	-0.006 (0.01)	-0.943 (0.07)	-0.935 (0.09)
d=0	-0.003 (0.01)	-0.006 (0.01)	-0.937 (0.08)	-0.926 (0.09)
d=0.2	-0.002 (0.01)	-0.004 (0.01)	-0.931 (0.19)	-0.914 (0.21)
d=0.4	-0.002 (0.01)	0.001 (0.01)	-1.011 (0.46)	-0.912 (0.48)
		$T = 2048$		
d=-0.4	**-0.001** (0.00)	**-0.003** (0.00)	**-0.926** (0.06)	**-0.906** (0.07)
d=-0.2	-0.001 (0.00)	-0.003 (0.00)	-0.926 (0.06)	-0.902 (0.07)
d=0	-0.001 (0.00)	-0.003 (0.00)	-0.920 (0.06)	-0.898 (0.07)
d=0.2	-0.001 (0.00)	-0.002 (0.00)	-0.906 (0.10)	-0.872 (0.11)
d=0.4	-0.001 (0.00)	0.003 (0.00)	-1.031 (0.40)	-0.935 (0.42)

Table 7: Biases and standard errors of the estimated memory parameter $\hat{\alpha}_1 - \alpha_1$: the effect of tapering $T = 1024$ and $\alpha = \{-0.2, 0, 0.3, 0.7, 0, 0.9, 0.99\}$

	Whittle	Whittle (taper)	LWE	LWE (taper)
		$\alpha = -0.2$		
d=-0.4	0.002 (0.05)	0.008 (0.05)	0.002 (0.06)	0.025 (0.12)
d=-0.2	0.006 (0.05)	0.009 (0.05)	0.008 (0.06)	0.027 (0.12)
d=0	0.009 (0.06)	0.012 (0.08)	0.011 (0.06)	0.027 (0.11)
d=0.2	0.022 (0.14)	0.023 (0.14)	0.009 (0.06)	0.033 (0.12)
d=0.4	0.010 (0.09)	0.017 (0.14)	0.008 (0.06)	0.032 (0.11)
		$\alpha = 0$		
d=-0.4	0.006 (0.05)	0.013 (0.07)	-0.003 (0.07)	0.018 (0.13)
d=-0.2	0.010 (0.06)	0.014 (0.06)	0.003 (0.07)	0.020 (0.12)
d=0	0.012 (0.05)	0.019 (0.09)	0.006 (0.06)	0.020 (0.12)
d=0.2	0.012 (0.05)	0.013 (0.06)	0.004 (0.06)	0.026 (0.13)
d=0.4	0.009 (0.05)	0.006 (0.06)	0.003 (0.07)	0.026 (0.12)
		$\alpha = 0.3$		
d=-0.4	0.023 (0.10)	0.042 (0.13)	-0.028 (0.07)	-0.015 (0.13)
d=-0.2	0.032 (0.10)	0.043 (0.13)	-0.022 (0.07)	-0.013 (0.12)
d=0	0.034 (0.10)	0.042 (0.12)	-0.019 (0.07)	-0.013 (0.12)
d=0.2	0.028 (0.08)	0.032 (0.11)	-0.022 (0.07)	-0.008 (0.13)
d=0.4	0.031 (0.11)	0.021 (0.10)	-0.022 (0.07)	-0.006 (0.12)
		$\alpha = 0.7$		
d=-0.4	0.017 (0.08)	0.021 (0.09)	-0.180 (0.06)	-0.195 (0.11)
d=-0.2	0.022 (0.08)	0.021 (0.09)	-0.175 (0.06)	-0.188 (0.11)
d=0	0.021 (0.07)	0.022 (0.09)	-0.171 (0.06)	-0.186 (0.11)
d=0.2	0.023 (0.08)	0.017 (0.09)	-0.172 (0.06)	-0.186 (0.11)
d=0.4	0.018 (0.08)	0.009 (0.10)	-0.174 (0.06)	-0.188 (0.11)
		$\alpha = 0.9$		
d=-0.4	-0.005 (0.04)	-0.008 (0.04)	-0.536 (0.08)	-0.576 (0.14)
d=-0.2	-0.005 (0.03)	-0.010 (0.04)	-0.535 (0.08)	-0.568 (0.14)
d=0	-0.005 (0.03)	-0.007 (0.04)	-0.528 (0.08)	-0.564 (0.14)
d=0.2	-0.003 (0.03)	-0.004 (0.04)	-0.530 (0.08)	-0.573 (0.14)
d=0.4	0.005 (0.03)	-0.006 (0.04)	-0.509 (0.08)	-0.553 (0.14)
		$\alpha = 0.99$		
d=-0.4	-0.006 (0.01)	-0.007 (0.01)	-0.943 (0.07)	-0.957 (0.13)
d=-0.2	-0.006 (0.01)	-0.006 (0.01)	-0.943 (0.07)	-0.954 (0.12)
d=0	-0.006 (0.01)	-0.006 (0.01)	-0.937 (0.08)	-0.956 (0.13)
d=0.2	-0.004 (0.01)	-0.004 (0.01)	-0.931 (0.19)	-1.019 (0.24)
d=0.4	0.001 (0.01)	-0.002 (0.01)	-1.011 (0.46)	-1.350 (0.35)

Table 8: Summary statistics of the log RVs of various financial assets

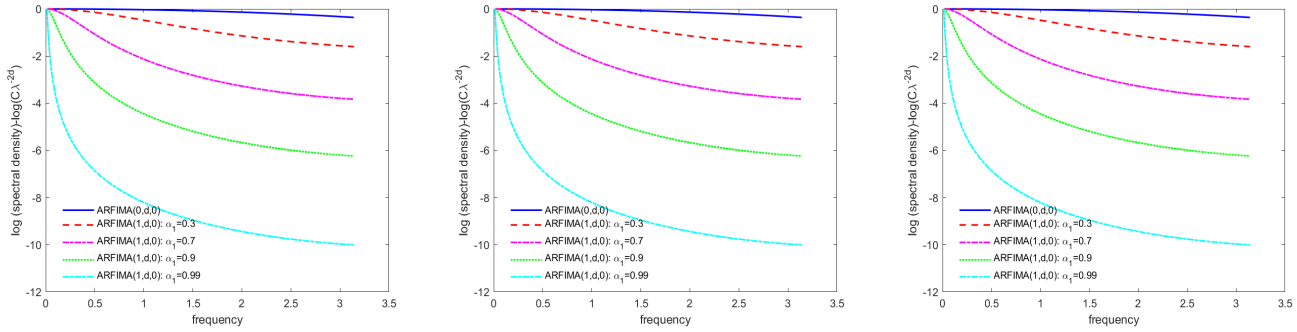
			RV				log RV			
Name	Start date	Obs.	Mean	Std.	Skew.	Kurto.	Mean	Std.	Skew.	Kurto.
Market Futures or ETFs										
S&P 500 ETF (SPY)	05-Jan-2010	2757	0.10	0.07	4.07	30.67	-2.40	0.50	0.66	4.04
S&P 500 E-min Futures (ES)	05-Jan-2010	2671	0.14	0.08	3.00	19.17	-2.11	0.49	0.47	3.52
Nikkei 225 Futures CME (NK)	28-Jul-2015	1383	0.18	0.12	3.68	24.04	-1.85	0.48	0.74	4.31
NASDAQ 100 E-min Futures (NQ)	28-Jul-2015	1305	0.16	0.10	3.49	24.99	-1.94	0.48	0.58	3.66
Dow Jones E-mini Futures (YM)	05-Jan-2010	2627	0.13	0.09	4.68	42.64	-2.16	0.49	0.70	4.16
Industry ETFs										
Financial (XLF)	05-Jan-2010	2724	0.14	0.10	10.98	268.32	-2.04	0.43	0.95	5.31
Industrial (XLI)	05-Jan-2010	2724	0.13	0.08	3.78	26.27	-2.18	0.45	0.75	4.32
Technology (XLK)	05-Jan-2010	2723	0.13	0.08	5.70	70.26	-2.18	0.46	0.78	4.65
Health Care (XLV)	05-Jan-2010	2723	0.12	0.07	6.20	81.80	-2.25	0.41	0.98	5.27
Consumer Discretionary (XLY)	05-Jan-2010	2722	0.12	0.07	3.60	25.11	-2.22	0.46	0.73	4.01
Exchange rate futures										
New Zealand Dollar Futures (NE)	05-Jan-2010	2733	0.11	0.05	2.57	18.72	-2.26	0.36	0.22	3.74
Swiss Franc Futures (SF)	05-Jan-2010	2733	0.09	0.06	23.09	884.61	-2.51	0.40	0.53	5.34
Euro FX E-mini Futures (UROM)	28-Jul-2015	1381	0.07	0.03	3.53	30.49	-2.69	0.35	0.37	4.56

Table 9: Model estimation results

Name (Ticker)	MPL		Whittle		LWE		LPE	
	$\hat{d}$	α	$\hat{d}$	α	$\hat{d}$	α	$\hat{d}$	α
Market Indexes: 01/01/2010 to 21/05/2021								
S&P 500 ETF (SPY)	-0.382	0.995	-0.382	0.995	0.602	-0.006	0.622	-0.027
S&P 500 E-mini Futures (ES)	-0.339	0.994	-0.339	0.994	0.586	0.072	0.616	0.039
Nikkei 225 Futures CME (NK)	-0.393	0.994	-0.391	0.992	0.623	-0.040	0.639	-0.055
NASDAQ 100 E-mini Futures (NQ)	-0.264	0.985	-0.262	0.984	0.571	0.190	0.529	0.237
Dow Jones E-mini Futures (YM)	-0.304	0.992	-0.304	0.992	0.582	0.114	0.607	0.085
Industry ETF: 01/01/2010 to 21/05/202								
Financial (XLF)	-0.427	0.998	-0.427	0.997	0.607	-0.053	0.599	-0.046
Industrial (XLI)	-0.422	0.998	-0.423	0.997	0.611	-0.056	0.625	-0.070
Technology (XLK)	-0.440	0.998	-0.439	0.997	0.540	0.004	0.543	0.001
Health Care (XLV)	-0.426	0.995	-0.426	0.994	0.601	-0.061	0.553	-0.012
Consumer Discretionary (XLY)	-0.434	0.998	-0.434	0.997	0.607	-0.070	0.586	-0.050
Exchange Rate Futures: 01/01/2010 to 21/05/2021								
New Zealand Dollar Futures (NE)	0.469	-0.076	0.474	-0.077	0.610	-0.186	0.607	-0.184
Swiss Franc Futures (SF)	0.472	-0.057	0.482	-0.061	0.635	-0.182	0.635	-0.183
Euro FX E-Mini Futures (UROM)	0.437	-0.059	0.444	-0.059	0.625	-0.200	0.664	-0.223

Figure 1: The theoretical spectral densities of ARFIMA models: $\log(f(\lambda)) - \log(C\lambda^{-2d})$

(a) $d = -0.4$ (b) $d = 0$ (c) $d = 0.4$



(d) $\alpha = 0$ (e) $\alpha = 0.3$ (f) $\alpha = 0.9$

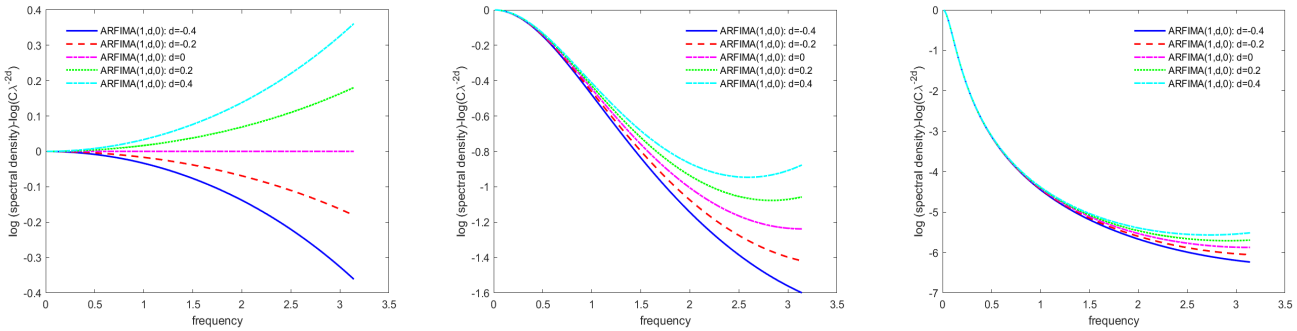
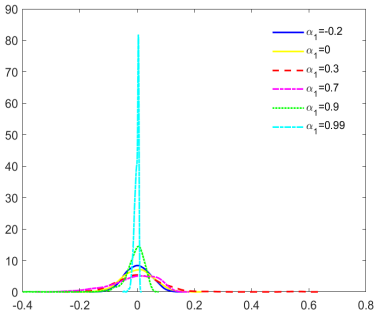
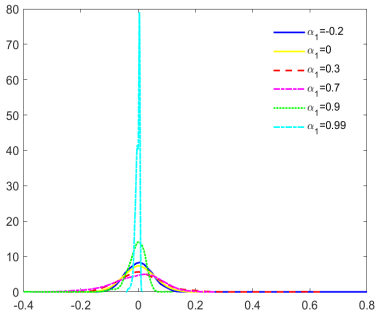


Figure 2: The kernel densities of the MPL and Whittle estimates of $\hat{\alpha}_1 - \alpha_1$

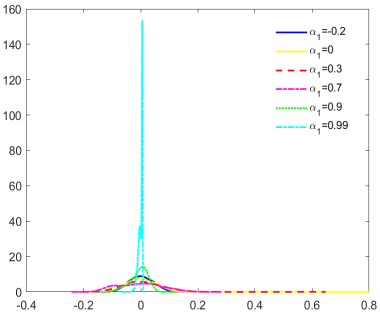
(a) MPL: $d = -0.4$



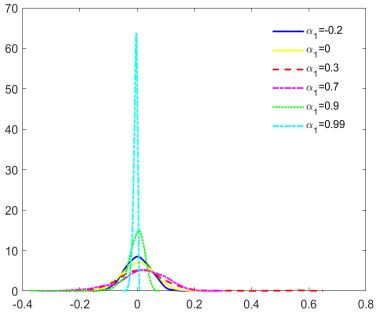
(b) MPL: $d = 0$



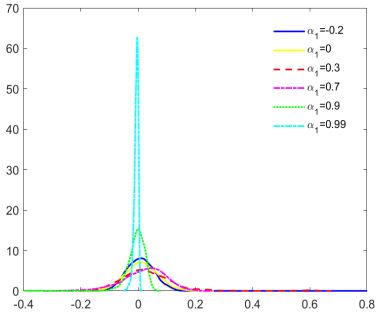
(c) MPL: $d = 0.4$



(d) Whittle: $d = -0.4$



(e) Whittle: $d = 0$



(f) Whittle: $d = 0.4$

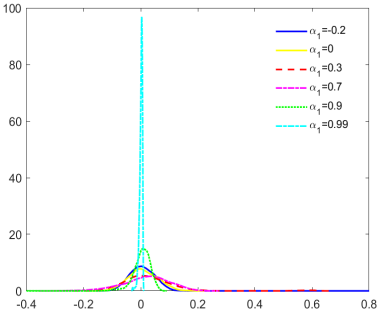
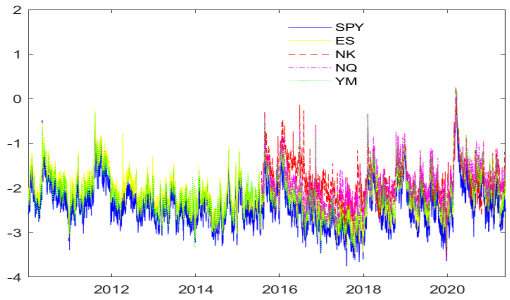
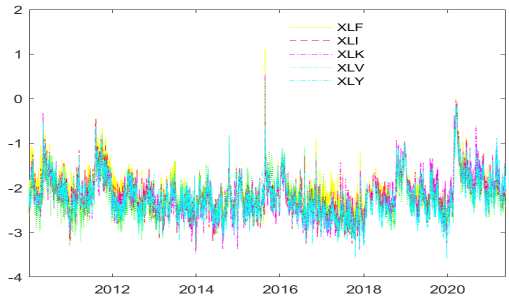


Figure 3: The log RV of various financial assets

(a) Markets



(b) Industry ETF



(c) FX

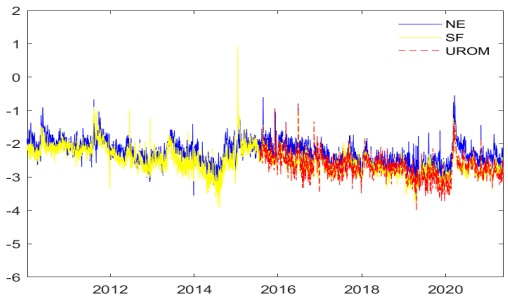
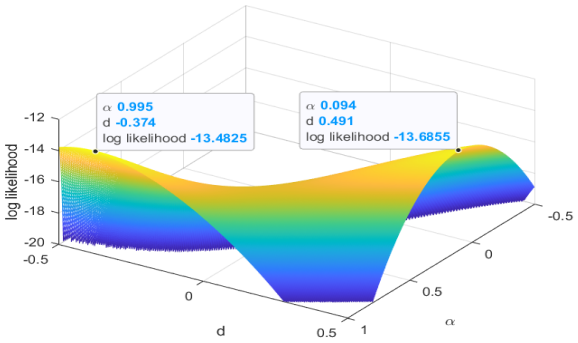


Figure 4: The Whittle log likelihood surface of log RVs

(a) SPY



(b) UROM

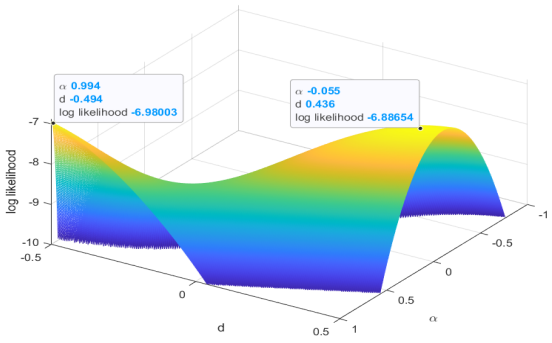
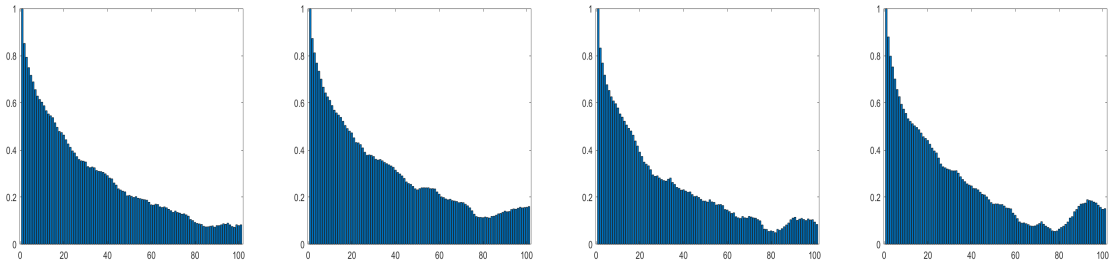
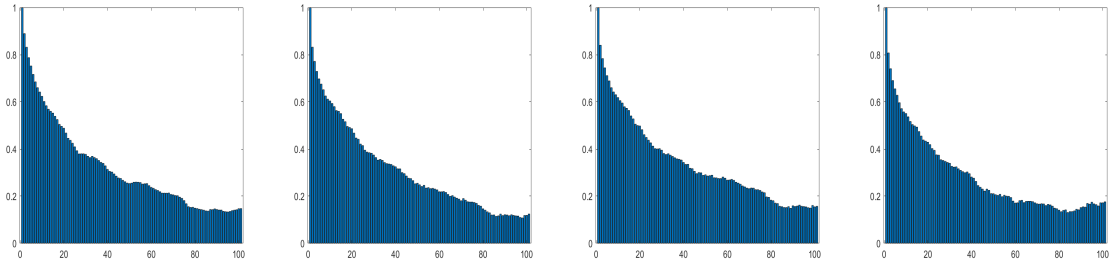


Figure 5: The autocorrelation functions of the logarithmic volatility of the thirteen financial assets

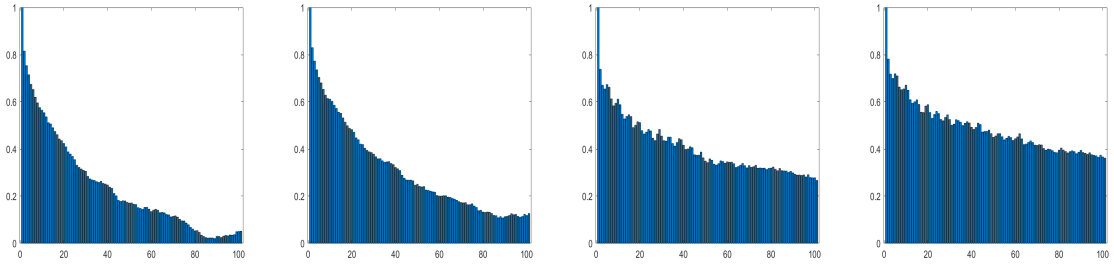
(a) SPY (b) ES (c) NK (d) NQ



(e) YM (f) XLF (g) XLI (h) XLK



(i) XLV (j) XLY (k) NE (l) SF



(m) UROM