



Article

# Temporal Aggregation and Long Memory for Asset Price Volatility

Pierre Perron <sup>1,\*</sup> and Wendong Shi <sup>2</sup>

<sup>1</sup> Department of Economics, Boston University, 270 Bay State Rd., Boston, MA 02215, USA

<sup>2</sup> School of Economics, Renmin University of China, 59 Zhongguancun Street, Beijing 100872, China; shi@ruc.edu.cn

\* Correspondence: perron@bu.edu

Received: 3 August 2020; Accepted: 12 August 2020; Published: 15 August 2020



**Abstract:** The effects of temporal aggregation and choice of sampling frequency are of great interest in modeling the dynamics of asset price volatility. We show how the squared low-frequency returns can be expressed in terms of the temporal aggregation of a high-frequency series. Based on the theory of temporal aggregation, we provide the link between the spectral density function of the squared low-frequency returns and that of the squared high-frequency returns. Furthermore, we analyze the properties of the spectral density function of realized volatility series, constructed from squared returns with different frequencies under temporal aggregation. Our theoretical results allow us to explain some findings reported recently and uncover new features of volatility in financial market indices. The theoretical findings are illustrated via the analysis of both low-frequency daily Standard and Poor's 500 (S&P 500) returns from 1928 to 2011 and high-frequency 1-min S&P 500 returns from 1986 to 2007.

**Keywords:** long memory; stochastic volatility; temporal aggregation; semiparametric estimators; random level shifts

## 1. Introduction

Long-memory processes, especially the possibility of confusing them with structural changes, are of great interest in the field of time series. Applications of long-memory models are numerous, in particular in relation to stock return volatility in financial markets. [Ding et al. \(1993\)](#) argue that stock return volatility series can be well described by long-memory processes. However, it has also been shown that the estimate of the long-memory parameter,  $d$ , is biased away from 0 and the autocovariance function exhibits a slow rate of decay when a stationary short-memory process is contaminated by structural changes in level. In other words, a spurious long-memory process can arise when there are structural changes in a short-memory process. This idea extends that advanced by [Perron \(1989, 1990\)](#), who shows that structural changes and unit roots ( $d = 1$ ) are easily confused in the sense that the estimate of the sum of the autoregressive coefficients is biased towards 1 and that tests of the null hypothesis of a unit root are biased towards non-rejection with a stationary process contaminated by structural changes. Relevant literature on this issue include [Diebold and Inoue \(2001\)](#), [Engle and Smith \(1999\)](#), [Gourieroux and Jasiak \(2001\)](#), [Granger and Hyung \(2004\)](#).

[Perron and Qu \(2007, 2010\)](#) analyze the properties of the autocorrelation function, the periodogram, and the log-periodogram (LP) estimate of the long-memory parameter for short-memory processes with random level shifts (RLS). They show that the autocorrelation function, the periodogram, and the LP estimates for the log squared daily returns for the Standard and Poor's 500 (S&P 500) during 1928–2002 can be explained by a level shift model with a short-memory component, instead of a long-memory process without level shifts. [Lu and Perron \(2010\)](#) present a method to directly estimate

the level shift model using an extension of the Kalman filter and apply it to the log absolute returns for the S&P 500, AMEX, DJIA and NASDAQ stock market return indices. Their point estimates imply few level shifts for all series but once these are taken into account there is no evidence of long-memory in the sense that little serial correlation is found in the remaining noise.

McCloskey and Perron (2013) provide simple trimmed versions of the LP estimator, which are consistent and asymptotically normal with the same limiting variance as the standard LP estimator regardless of whether the underlying long/short-memory process is contaminated by level shifts or deterministic trends. Using these robust LP estimators, they study the log-squared daily return series of the S&P 500, the Dow Jones Industrial Average (DJIA), the NASDAQ and the AMEX stock market indices, which are also examined by Lu and Perron (2010). Their robust LP estimator is 0.007, indicating the (near) absence of long-memory for the S&P 500 volatility series, which contradicts the estimate of 0.51, given by the standard LP estimator. Very similar results emerge for the DJIA and the AMEX. An interesting finding is that the robust LP estimator is 0.51, indicating a highly persistent long-memory process, when using the log of daily realized volatility series constructed from 5-min returns of the S&P 500 futures index from 21 April 1982 to 2 March 2007. The contrasting results raise an interesting puzzle. More recently, Varneskov and Perron (2018) extend the work of Lu and Perron (2010) and adopt a full parametric model including both random level shifts and ARFIMA components. They estimate the long-memory parameter,  $d$ , for the realized volatility series of the S&P 500 data generated from 5-min returns, which is the same data used by McCloskey and Perron (2013). They also estimate the long-memory parameter for the squared daily returns of the S&P 500 data for the time period 1929–2004, using a RLS+ARFIMA  $(0, d, 0)$  model. The estimate for the former is 0.36, indicating the existence of long-memory processes, while it is 0.053 for the latter, indicating the (near) absence of long-memory processes. They also consider other RLS+ARFIMA models and the 30-year T-bonds data and the Dollar-AUS exchange rate, all of which present similar results.

The realized volatility series are essentially the aggregation of the squared high-frequency returns, and the daily return series are the aggregated intraday returns. In this regard, it may be possible to explain the above puzzle by means of temporal aggregation effects. There is a vast amount of papers on temporal aggregation, e.g., Wei (1978) and Chambers (1998), but little attention has been paid to its effects in the frequency domain. In a series of papers, Souza (2005, 2007, 2008) discuss the effect of temporal aggregation and bandwidth selection in estimating the degree of long memory. Ohanissian et al. (2008) propose a test to distinguish between true and spurious long memory by exploiting the invariance of the long-memory parameter to temporal aggregation. Hassler (2011) investigates whether typical frequency domain assumptions made for semiparametric estimation and inference are closed with respect to aggregation.

It is believed by many that various measures of volatility are perturbed fractionally integrated processes, related to stochastic volatility model (Hassler 2011). However, the aggregation of a stochastic volatility model has been little discussed. Bollerslev and Wright (2000) study the effect of the temporal aggregation of a long-memory stochastic volatility model in the time domain. Studying the effect of the temporal aggregation of a stochastic volatility model in the frequency domain is also important. First, the log periodogram (LP) estimator has become one of the most popular memory parameter estimator among empiricists due to simplicity, intuitiveness and ease of use, as discussed by McCloskey and Perron (2013). Therefore, it is necessary to study the aggregation of the stochastic volatility in the frequency domain. Second, by using a stochastic volatility model, we can better understand the difference between the various measures of volatilities, e.g., the realized volatility and squared daily returns.

The aim of the paper is first to show how the squared low-frequency returns can be expressed in terms of the temporal aggregation of a high-frequency series. We build a bridge between the spectral density function of squared low-frequency returns and that of the squared high-frequency returns. Furthermore, we analyze the properties of the spectral density function of realized volatility, constructed from the squared returns with different frequencies under temporal aggregation.

These results will allow us to tackle a second aim and help us explain the following puzzles. First, the trimmed LP estimates for the log squared daily return series are very closed to zero, while they are relatively large (around 0.4) for the log daily realized volatility constructed from high-frequency returns. Second, the trimmed LP estimates for the log squared daily return series are near zero, while they are very large for the log aggregated squared daily returns. Third, the trimmed LP estimates for the log realized return series are large using high-frequency return data, while they are close to zero for the disaggregated original high-frequency returns when a large bandwidth is used. The theoretical findings are illustrated through the analysis of both low-frequency daily S&P 500 returns from 1928 to 2011 and high-frequency 1-min S&P 500 returns from 1986 to 2007. We consider the estimation of the long-memory parameter using the standard and the trimmed LP estimate. Overall, the results indicate that random level shifts are needed to explain the empirical features documented. For the low frequency data on S&P 500 returns, one cannot infer whether the noise is stationary long memory. On the other hand, a long-memory process appears needed to explain the features related to high frequency S&P 500 futures. This is in line with the findings of [Varneskov and Perron \(2018\)](#).

The remainder of this paper is structured as follows. Section 2 introduces the stochastic volatility model and the different aggregation processes for the realized volatility and the squared daily returns. Section 3 analyzes the properties of the long-memory parameter estimates with different aggregation levels and with random level shifts. Section 4 focuses on empirical applications of the theoretical findings to S&P500 data with different sampling frequencies. Section 5 contains brief concluding remarks and comments about potential avenues for future research. A mathematical appendix contains the technical derivations.

## 2. Alternative Volatility Measures

We first review the stochastic volatility models and the aggregation mechanisms. We then present theoretical results about temporal aggregation in the frequency domain.

### 2.1. Stochastic Volatility Model

It is evident that the number of observations for stock market returns during a fixed period of time is inversely related to the length of each return interval. That is, with prices fully available, a shorter interval means that the returns are observed more frequently and therefore more observations can be obtained. High-frequency returns are now available since many financial time series are available on a tick-by-tick basis, which is virtually continuous. In this paper, high-frequency returns are classified according to the numbers of time units included in their intervals. To simplify, the returns obtained every one unit of time period are defined as 1-period returns, and similarly  $k$ -period returns denote the returns observed every  $k$  units of time periods. Here,  $k$  is an integer larger than 1.

Let  $r_{t,n}$  be the  $n$ th intraday log-return (i.e.,  $r_{t,n} = \log(P_{t,n}) - \log(P_{t,n-1})$ , where  $P_{t,n}$  is the price index) at day  $t$  such that

$$r_{t,n} = (1/\sqrt{s})h_{t,n}\varepsilon_{t,n}$$

with  $n = 1, \dots, s$  is the highest sampling frequency and  $t = 1, \dots, T$  is the number of days. The component  $h_{t,n}$  intends to capture the volatility level.

**Assumption 1.** We assume that  $\varepsilon_{t,n}$  is i.i.d. standard normal, and  $h_{t,n}$  and  $\varepsilon_{t,n}$  are mutually independent. It is further assumed that the demeaned squared volatility level  $y_{t,n} = h_{t,n}^2 - E(h_{t,n}^2)$  is covariance stationary long-memory with integrable spectral density  $f_y(\lambda)$ .

**Remark 1.** This assumption considers the case of  $y_{t,n}$  being long memory, with spectral density function  $f_y(y) \approx \lambda^{-2d}$  as  $\lambda \rightarrow 0$ .

## 2.2. Temporal Aggregation

According to the classification of high-frequency returns in Section 2.1, a sample of  $s$  1-period returns contains  $s/k$   $k$ -period returns. We assume that  $k$  is chosen such that  $s = kS$  for some integer  $S$ . Let  $r_{t,p}^{(k)} = \sum_{n=k(p-1)+1}^{pk} r_{t,n} = \sum_{j=0}^{k-1} L^j r_{t+kp}$  denote the continuously compounded  $k$ -period return, so that

$$r_{t,p}^{(k)} = (1/\sqrt{s}) \left( \sum_{j=0}^{k-1} L^j \right) h_{t+kp} \epsilon_{t+kp}$$

for  $p = 1, \dots, S$  and  $t = 1, \dots, T$ . Here,  $L$  is the backshift operator. Note that the backshift operator,  $L$ , is applied to the second subscript, i.e.,  $Lh_{t+kp}\epsilon_{t+kp} = h_{t+kp-1}\epsilon_{t+kp-1}$ . Therefore, the  $k$ -period return  $r_{t,p}^{(k)}$  can be written as

$$r_{t,p}^{(k)} = (1/\sqrt{s}) z_{t,p} \sqrt{\left( \sum_{j=0}^{k-1} L^j \right) h_{t+kp}^2}$$

where  $z_{t,p}$  is i.i.d. standard normal. Then the squared  $k$ -period return  $r_{t,p}^{(k)2}$  is given by

$$r_{t,p}^{(k)2} = (1/s) z_{t,p}^2 \left( \sum_{j=0}^{k-1} L^j \right) h_{t+kp}^2$$

which can be expressed in terms of temporal aggregation, as

$$r_{t,p}^{(k)2} = \left[ (1/s) h_{t,n}^2 z_{t,[n/k]+1}^2 \right]^{(k)}$$

where  $[n/k]$  denotes the integer part of  $n/k$  and  $[\cdot]^{(k)}$  denotes the  $k$ -period non-overlapping temporal aggregation of the series  $x_t$ , i.e.,

$$x_p^{(k)} = \sum_{j=0}^{k-1} x_{kp-j}, \quad p = 1, \dots, P.$$

Note that in the case of aggregated variables, we use a square bracelet operator, i.e.,  $[\cdot]^{(k)}$ . This should not be confused with  $[\cdot]$  without a subscript  $(k)$  which denotes the integer part. Therefore, the squared daily return

$$r_t^2 = \left[ (1/s) h_{t,n}^2 z_t^2 \right]^{(s)}$$

is the temporal aggregation of  $(1/s) h_{t,n}^2 z_t^2$ , over a day. Here,  $z_t$  is i.i.d. standard normal for  $t = 1, \dots, T$ . The realized volatility constructed from the  $k$ -period returns  $r_{t,p}^{(k)}$ ,

$$RV_t = \sum_{p=1}^S r_{t,p}^{(k)2}$$

is the temporal aggregation of the squared  $k$ -period returns  $r_{t,p}^{(k)2}$ , such that

$$RV_t = \left[ r_{t,p}^{(k)2} \right]^{(S)}$$

## 2.3. Temporal Aggregation in the Frequency Domain

**Assumption 2.** We assume that the demeaned squared volatility level  $y_{t,n} = h_{t,n}^2 - E(h_{t,n}^2)$  is covariance stationary with integrable spectral density  $f_y(\lambda)$ .

**Proposition 1.** Under Assumptions 1-2, the spectral density of the squared  $r_{t,p}^{(k)}$  is

$$f_{r_{t,p}^{(k)2}}(\lambda) = \frac{1}{s^2 k} \left| \sin\left(\frac{\lambda}{2}\right) \right|^2 \sum_{j=0}^{k-1} \left| \sin\left(\frac{\lambda + 2j\pi}{2k}\right) \right|^{-2} f_y\left(\frac{\lambda + 2j\pi}{2k}\right) + \frac{1}{s^2 \pi} \left[ k^2 \left( E\left(h_{t,n}^2\right) \right)^2 + \text{Var}\left(h_{t,n}^{(k)}\right) \right] \quad (1)$$

**Proof.** See the Appendix A.  $\square$

The first term on the right hand side of (1) corresponds to the spectral density of the temporal aggregation of the  $k$ -period demeaned squared volatility level,  $[y_{t,n}]^{(k)}$ . The remaining two terms are constant, induced by the noise component  $h_t$ , which does not carry any information about the long-memory.

**Remark 2.** When  $k = s$ , we have the spectral density of the squared daily returns

$$f_{r_t^2}(\lambda) = \frac{1}{s^3} \left| \sin\left(\frac{\lambda}{2}\right) \right|^2 \sum_{j=0}^{s-1} \left| \sin\left(\frac{\lambda + 2j\pi}{2s}\right) \right|^{-2} f_y\left(\frac{\lambda + 2j\pi}{2s}\right) + \frac{1}{s^2 \pi} \left[ s^2 \left( E\left(h_{t,n}^2\right) \right)^2 + \text{var}\left(h_{t,n}^{(s)}\right) \right] \quad (2)$$

The first term of (2) is such that

$$\frac{1}{s^3} \left| \sin\left(\frac{\lambda}{2}\right) \right|^2 \sum_{j=0}^{s-1} \left| \sin\left(\frac{\lambda + 2j\pi}{2s}\right) \right|^{-2} f_y\left(\frac{\lambda + 2j\pi}{2s}\right) \rightarrow \frac{1}{s} f_y\left(\frac{\lambda}{2s}\right)$$

as  $\lambda \rightarrow 0$ . Therefore, the spectral density of the demeaned squared daily volatility decreases as  $s$  increases. When  $s$  is large enough, the spectral density of the squared daily returns  $f_{r_t^2}(\lambda)$  will be dominated by the second and third terms of (2), which implies that the squared daily return series is dominated by noise.

**Proposition 2.** The spectral density of the realized volatility obtained from  $k$ -period returns,  $r_{t,p}^{(k)}$ , is

$$f_{[r_{t,p}^{(k)2}]^{(s)}}(\lambda) = \frac{1}{s^3} \left| \sin\left(\frac{\lambda}{2}\right) \right|^2 \sum_{j=0}^{s-1} \left| \sin\left(\frac{\lambda + 2j\pi}{2s}\right) \right|^{-2} f_y\left(\frac{\lambda + 2j\pi}{2s}\right) + \frac{1}{s^2 \pi} \left[ sk \left( E h_{t,n}^2 \right)^2 + s \text{var}\left(h_{t,n}^{(k)}\right) \right] \quad (3)$$

**Proof.** See the Appendix A.  $\square$

**Remark 3.** When  $S = s$ , i.e., the realized volatility is obtained from 1-period returns,  $r_{t,n}$ , we have the following form of the spectral density

$$f_{[r_{t,n}^2]^{(s)}}(\lambda) = \frac{1}{s^3} \left| \sin\left(\frac{\lambda}{2}\right) \right|^2 \sum_{j=0}^{s-1} \left| \sin\left(\frac{\lambda + 2j\pi}{2s}\right) \right|^{-2} f_y\left(\frac{\lambda + 2j\pi}{2s}\right) + \frac{1}{s \pi} \left[ \left( E h_{t,n}^2 \right)^2 + \text{var}\left(h_{t,n}^2\right) \right] \quad (4)$$

For the spectral density of realized volatility, we have from (4)

$$f_{[r_{t,n}^2]^{(s)}}(\lambda) \rightarrow \frac{1}{s \pi} \left[ \pi f_y\left(\frac{\lambda}{2s}\right) + \left( E\left(h_{t,n}^2\right) \right)^2 + \text{var}\left(h_{t,n}\right) \right]$$

as  $\lambda \rightarrow 0$ . This means that the three terms of the spectral density of the realized volatility in equation (4) decrease at the same rate when  $s$  increases. Comparing the spectral density of the squared daily returns (2) with that of the realized volatility (4), note that their first terms are identical, i.e.,

$$\frac{1}{s^3} \left| \sin \left( \frac{\lambda}{2} \right) \right|^2 \sum_{j=0}^{s-1} \left| \sin \left( \frac{\lambda + 2j\pi}{2s} \right) \right|^{-2} f_y \left( \frac{\lambda + 2j\pi}{2s} \right)$$

corresponding to the spectral density of the temporal aggregation of the demeaned squared volatility level,  $y_{t,n} = h_{t,n}^2 - E(h_{t,n}^2)$ , over a day. This is the only part that contains information about the long memory and the remaining terms are simply noise. Therefore, both realized volatility and squared daily returns contain the same information about long memory.

A difference between the spectral density of the squared daily returns and that for the realized volatility series occurs only in the second and the third terms,

$$f_{r_t^2}(\lambda) - f_{[r_{t,p}^{(k)2}]^{(s)}}(\lambda) = \frac{(s-k)}{s\pi} \left( E(h_{t,n}^2) \right)^2 + \frac{1}{s^2\pi} \left[ \text{var}(y_{t,n}^{(s)}) - S\text{var}(h_{t,n}^{(k)}) \right] \quad (5)$$

which is independent of the value of  $\lambda$ . Furthermore,  $\text{Var}([y_{t,n}]^{(s)}) - S\text{var}([h_{t,n}]^{(k)})$  is positive in general because most financial series have positive autocorrelation even at large lag. Therefore, the difference between the spectral density of the squared daily returns and the realized volatility will be positive and larger than  $[(s-k)/s\pi] (E(h_{t,n}^2))^2$ .

### 3. Long-Memory Parameter Estimates across Aggregation Levels

We first describe in Section 3.1 the standard log-periodogram regression, both regular and trimmed as suggested by McCloskey and Perron (2013). Then, in Section 3.2, we show the equivalence of the estimates across aggregation levels.

#### 3.1. Log-Periodogram Regressions

A long-memory process typically has a spectral density function which is proportional to  $\lambda^{-2d}$  as  $\lambda$  goes to zero, where  $d$  is the memory parameter. The fractionally integrated model, proposed by Granger and Joyeux (1980) and Hosking (1981), is a long-memory generalization of an ARMA model whose autocorrelations decay exponentially. When  $d \in (0, 0.5)$ , the autocorrelations decay slowly, a characteristic of long-memory processes. Various estimators of  $d$  have been proposed, among which semiparametric estimators have become widely used as they do not require a distributional assumption on the process generating the difference of order  $d$  of the series. A popular semiparametric estimator is the LP regression estimator proposed by Geweke and Porter-Hudak (1983), which uses only frequencies near zero to avoid possible misspecification caused by high frequency movements. The LP regression estimator was analyzed by, among others, Robinson (1995).

The LP regression estimator is based on the following spectral characterization of a long-memory process:

$$\log f(\lambda) \approx c - 2d \log \lambda$$

as  $\lambda \rightarrow 0_+$ , where  $f$  is the spectral density function of the process. The periodogram of the time series at  $\lambda_j$  is defined as

$$I_x(\lambda_j) \equiv |w_x(\lambda_j)|^2 = w_{x(\lambda_j)} w_{x(\lambda_j)}^*,$$



where  $w_x(\lambda_j)$  is the discrete Fourier transform of  $\{x_t\}_{t=1}^T$  evaluated at the Fourier frequency  $\lambda_j = 2\pi j/T$ , and  $c^*$  denotes the complex conjugate of any complex number  $c$ .  $I_x(\lambda_j)$  can be viewed as a noisy approximation to  $f$ . Therefore, the LP regression is:

$$\log I_x(\lambda_j) = c + dX_j + e_j, \quad j = l, \dots, m,$$

where  $X_j = -\log(2 - 2\cos(\lambda_j))$  for  $j = l, \dots, m$ . The LP regression estimator is

$$\hat{d} = \frac{-0.5 \sum_{j=l}^m (Y_j - \bar{Y}) \log I_j}{\sum_{j=l}^m (Y_j - \bar{Y})^2},$$

where  $Y_j = (1/2) \log(2 - 2\cos(\lambda_j))$  and  $\bar{Y} = (1/(m-l+1)) \sum_{k=l}^m Y_k$ . When  $l = 1$ , this is the standard LP regression estimator. We can trim some of the lower frequencies, as in [McCloskey and Perron \(2013\)](#), to obtain consistency and asymptotic normality with the same limiting variance as the standard LP regression estimator regardless of whether the underlying long/short-memory process is contaminated by level shifts or deterministic trends.

### 3.2. Equivalence of Estimates across Aggregation Levels

**Lemma 1.** Under Assumption 1, for the spectral densities of the aggregated series and the original squared return series, we have

$$f_{\left[r_{t,p}^{(k)2}\right]^{(S)}}(\lambda) - Sf_{r_{t,p}^{(k)2}}(\lambda/S) \rightarrow 0 \text{ as } \lambda \rightarrow 0.$$

**Proof.** See the Appendix A.  $\square$

Lemma 1 implies that the spectral densities of the aggregated series and the original squared return series have the same slope near frequency zero. Hence, aggregation does not change the value of the long-memory parameter, consistent with the results of [Chambers \(1998\)](#), [Souza \(2005\)](#) and [Hassler \(2011\)](#).

**Corollary 1.** The periodogram is a finite sample version of the spectral density; hence, a similar relation holds approximately, i.e.,

$$I_{\left[r_{t,p}^{(k)2}\right]^{(S)}} - SI_{r_{t,p}^{(k)2}} \rightarrow^p 0$$

as  $T \rightarrow \infty$  for  $\lambda \rightarrow 0$ .

A similar result was obtained for stationary long memory series by [Ohanissian et al. \(2008\)](#).

**Remark 4.** Lemma 1 and Corollary 1 do not depend on the stochastic volatility specification. The validity of the results only require that the original process be stationary.

**Remark 5.** As discussed by [Souza \(2008\)](#), a better estimator is not necessarily generated by temporally aggregating a time series, because the same estimate can be obtained when the same bandwidths are used on the original time series, which offers a wider choice of bandwidths. That is, the original time series could provide potentially improved estimates in the sense that it allows for more flexible bandwidth selection.

**Remark 6.** The microstructure noise is not taken into consideration here. However, adding a microstructure noise process will not change the result in Lemma 1, because Assumption 1 will still be satisfied even adding a stationary noise, and the above results hold as long as the time series is stationary.

According to [Perron and Qu \(2010\)](#), the autocorrelation function, the periodogram, and the LP estimate for the log squared daily returns for the S&P 500 can be explained by a simple level

shift model with a short-memory component, instead of a long-memory process without level shifts. Lu and Perron (2010) estimate a random level shifts model and find that few level shifts are present, but once these are accounted for, there is no evidence for the existence of long-memory processes in the sense that little serial correlation is found in the remaining noise. Therefore, random level shifts, which have not been included in Assumption 1, should be considered here to generalize Lemma 1.

We consider the following random level shift model proposed by Perron and Qu (2010),

$$u_{T,t} = \sum_{j=1}^t \delta_{T,j}, \quad \delta_{T,t} = \pi_{T,t} \eta_t \quad (6)$$

where  $\eta_t \sim i.i.d. N(0, \sigma_\eta^2)$  and  $\pi_{T,t} \sim i.i.d. Bernoulli(p/T, 1)$ . It is also assumed that the components  $\pi_{T,t}$  and  $\eta_t$  are mutually independent. According to Proposition 3 of Perron and Qu (2010), the limit of the expectation of the periodogram has the following form

$$\lim_{T \rightarrow \infty} T^{-1} E[I_{u,j}] = \frac{p\sigma_\eta^2}{4\pi^3} \frac{1}{j^2} \quad \text{as } T \rightarrow \infty.$$

**Lemma 2.** Under the data generating process (6), we have the following relation between the  $k$ -period aggregated series and the original random level shift series,

$$\lim_{T \rightarrow \infty} E[I_{u^{(k)},j}] - k \lim_{T \rightarrow \infty} E[I_{u^{(1)},j}] = 0 \quad \text{as } \lambda \rightarrow 0.$$

**Proof.** See Appendix A.  $\square$

Therefore, Remarks 4–6 still hold when random level shifts are taken into consideration.

#### 4. S&P 500 Volatility

We consider two series of returns related to the S&P 500 data, namely low-frequency and high-frequency returns. The low-frequency data consist of 22,000 daily return observations for the period from 13 August 1928 to 30 December 2011. Among these daily returns, the observations from 13 August 1928 to 30 October 2002 were kindly provided by William Schwert. The source of the data for the period 4 January 1928 through 2 July 1962 is Schwert (1990). From 3 July 1962 to 30 October 2002 it is from the CRSP daily returns file, and the returns for the time period after 30 October 2002 were obtained from the Yahoo Finance website. Because of the need to construct various aggregate measures, the effective initial date for estimate is 13 August 1928. The high-frequency data pertain to S&P 500 futures and includes 1-min returns from 7 October 1986 to 2 March 2007, amounting to 5000 trading days in total. These were purchased from <http://www.grainmarketresearch.com/>. These futures contracts expire within one year after their inception. Specifically, contracts incepted in January, April, July, and October expire in March, June, September, and December, respectively. The cleaned version was provided by Shin Ikeda as described in Appendix A of Ikeda (2015). The span of the data was mostly dictated by the data availability, though it conveniently avoids the turbulent period of the great recession. In order to eliminate the effect of outliers in the data, we use a logarithmic transformation of the observations. Since there are some zeros in the original high-frequency and daily data, we demean our data first, as in Deo et al. (2006). Other methods were proposed in the literature, e.g., Perron and Qu (2010) and Lu and Perron (2010), who add a small value to the squared returns.

##### 4.1. Low Frequency Data

We first start our analysis with low frequency data, i.e., the daily data series. Table 1 shows the LP estimates for the log realized  $S$ -day return series, which is simply calculated by cumulating  $S$  neighboring squared daily returns that do not overlap. More specifically,  $S = 1, 5, 10$ , and 20 stands for the squared daily returns, the realized weekly (every five business days), biweekly, and



monthly (every 20 business days) volatilities, respectively. The columns labelled  $[\log(r_t^2)]^{(S)}$ ,  $\log(r_t^2)$  and  $\log\{[r_t^2]^{(S)}\}$  refer to the  $S$ -period aggregation of the logarithmic transformation of squared daily returns, the logarithmic transformation of the original squared daily returns and the logarithmic transformation of the  $S$ -period aggregated squared daily returns, respectively, with  $S$  denoting the aggregation level. Both the standard LP (SLP) and trimmed LP (TLP) estimates are presented for purposes of comparisons. For each series, the standard LP estimate is computed using  $m = N^{0.5}$ , while the trimmed one is constructed using  $(l, m) = (N^{0.65}, N^{0.9})$ , which performs relatively well according to McCloskey and Perron (2013). In all cases,  $N = T/S$ . Note that for the original daily returns series  $\log(r_t^2)$  we consider the LP estimates constructed with different bandwidths to highlight the importance of the bandwidth selection. As the results will show, the empirical estimates are very similar using either the  $S$  periods aggregation with bandwidth  $T/S$  or the original daily series with the same bandwidth  $T/S$ . This is an implication of Lemmas 1–2 so that it should hold whether the process is a RLS or pure long-memory.

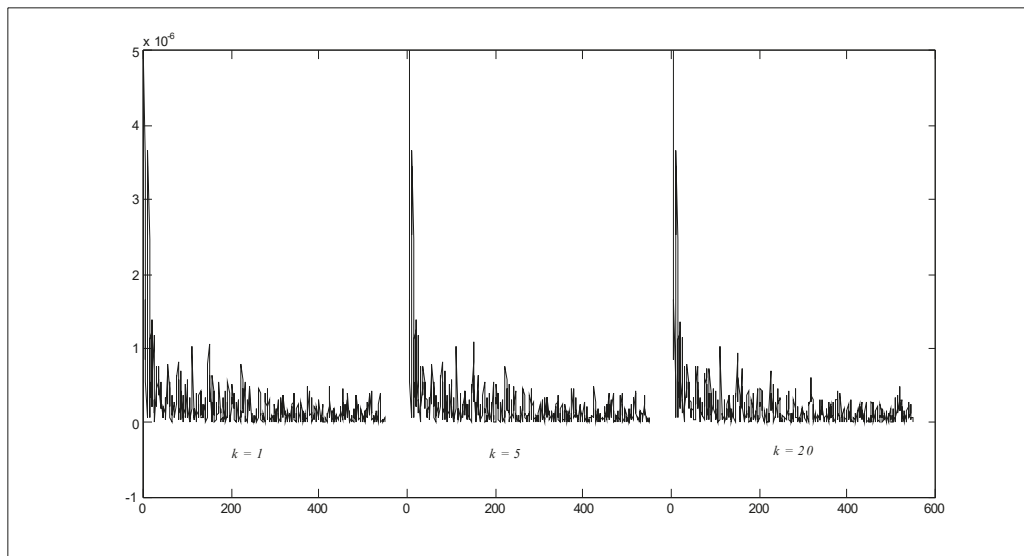
**Table 1.** Long-memory parameter estimates with daily S&P 500 data.

S	$[\log(r_t^2)]^{(S)}$		$\log(r_t^2)$		$\log\{[r_t^2]^{(S)}\}$	
	SLP	TLP	SLP	TLP	SLP	TLP
$S = 1$	0.5659	0.0941	0.5659	0.0941	0.5659	0.0941
$S = 5$	0.6313	0.2570	0.6317	0.2317	0.6236	0.3092
$S = 10$	0.7770	0.3544	0.7762	0.3200	0.7983	0.4359
$S = 20$	0.7440	0.4457	0.7427	0.3915	0.7554	0.4689

Notes: The table reports the standard log-periodogram regression estimates (SLP) and the trimmed log-periodogram regression estimates (TLP) for the degree of fractional integration in the daily S&P 500 returns data. The sample period is from 13 August 1928 to 30 December 2011. The sample size is 22,000. The rows labelled  $S = 1$ ,  $S = 5$ ,  $S = 10$  and  $S = 20$  refer to the squared daily returns, aggregated 5-day squared daily returns, aggregated 10-day squared daily returns and aggregated 20-day squared daily returns, respectively. The columns labelled  $[\log(r_t^2)]^{(S)}$ ,  $\log(r_t^2)$  and  $\log\{[r_t^2]^{(S)}\}$  refer to the  $S$ -periods aggregation of the logarithmic transformation of squared daily returns, the logarithmic transformation of the original squared daily returns and the logarithmic transformation of the  $S$ -periods aggregated squared daily returns, respectively.  $S$  denotes the aggregation level. The estimates are based on the bandwidth  $m = N^{0.5}$  for the log-periodogram regression and  $(l, m) = (N^{0.65}, N^{0.9})$  for the trimmed log-periodogram regression. For the estimates of aggregated series, we set  $N = T/S$ .

**Remark 7.** Our theory applies only to the cases  $[\log(r_t^2)]^{(S)}$  and  $\log(r_t^2)$  but not to  $\log\{[r_t^2]^{(S)}\}$ . However, unreported simulations show that all three measures behave similarly as the aggregation changes. Hence, we conjecture that it would be possible to extend our results to cover the case of  $\log\{[r_t^2]^{(S)}\}$ .

Some interesting results in Table 1 are worth notice. First, with the same bandwidth ( $N = T/S$ ), the estimates for the log realized daily return series, which is the log aggregated squared daily returns, are approximately equal to and always a bit larger than the estimates for the log squared original returns. This finding confirms Corollary 1, i.e., the same long-memory parameter estimate can be obtained for both the aggregated time series and the original series when the same bandwidths are used. Figure 1 shows the periodograms of the squared daily return series (left), the five-period aggregation of the squared daily return series, divided by 5 (middle), and the 20-periods aggregation of the squared daily return series, divided by 20 (right) for frequency indices up to 550. The 550th frequency index corresponds to the frequencies  $\pi/20$ ,  $\pi/4$ , and  $\pi$  for the squared daily returns, the five-period aggregation of the squared daily returns, and the 20-period aggregation of the squared daily returns, respectively. Note that they have almost identical pattern near frequency zero. This finding confirms Lemmas 1–2 again and implies the same long-memory parameter estimate for the aggregated and original series when the same bandwidths are used.



**Figure 1.** The periodogram of the squared daily returns (**left**), the 5-periods aggregation of the squared daily returns divided by 5 (**middle**), and the 20-periods aggregation of the squared daily returns divided by 20 (**right**) for the daily S&P 500 returns data. The sample period is from 13 August 1928 to 30 December 2011.

Second, when  $S = 1$ , i.e., the daily return series, note that both the standard and trimmed LP estimators are very different. In particular, the trimmed LP estimates are close to zero when  $S = 1$ , indicating the (near) absence of long-memory processes. On the other hand, the standard LP estimate is 0.57. However, for the realized 5-day return series, the standard and trimmed LP estimators are 0.62 and 0.31, respectively. In addition, a feature of interest is the fact that both the standard and trimmed LP estimates increase as  $S$  increases. As shown in Remark 5, the same estimate of the long-memory parameter for the aggregated series should be obtained when using the same bandwidths as those on the original time series. Therefore, the apparent difference could simply be caused by the bandwidth selection. We actually use a relative small bandwidth for the aggregated series, compared with the original series, because the number of observation in the aggregated series is smaller. The issue of interest is whether the documented feature is more likely to occur with a RLS model or with a pure long-memory process. As discussed in Perron and Qu (2010), the LP estimate increases as  $m$  decreases when considering the RLS plus white noise model. Hence, a larger LP estimate is expected for the aggregated series under RLS. In particular, it is expected to be greater than 0.5, i.e., in the non-stationary region. No such increase towards values in the non-stationary region is expected, as the bandwidth decreases, with a pure long-memory process. Hence, these results are more consistent with a RLS being the data-generating process.

#### 4.2. High Frequency Data

We now consider high-frequency data, for which the unit of time is one minute. Table 2 shows the LP estimates for the log realized daily volatility constructed from  $k$ -period high-frequency data, and the log squared original returns. Here,  $k = 1, 5, 30$ , and 330 correspond to the case of 1-min, 5-min, 30-min, and daily returns, respectively. The columns labelled  $[\log(r_{t,n}^{(k)2})]^{(S)}$ ,  $\log[r_{t,n}^{(k)2}]$  and  $\log[r_{t,n}^{(k)2}]^{(S)}$  refer to the  $S$ -period aggregation of the logarithmic transformation of squared  $k$ -min returns, the logarithmic transformation of squared  $k$ -min returns and logarithmic transformation of the realized daily volatility aggregated by squared  $k$ -min returns over a day, respectively.  $S$  denotes the number of  $k$ -min returns per day, with  $s = kS$ . Both the standard and trimmed LP estimates are presented for purposes of comparisons. For each series, the standard LP estimate is constructed using  $m = N^{0.5}$ , and the trimmed one using  $(l, m) = (N^{0.65}, N^{0.9})$ . For the log realized return volatility series, we let

$N = T$ , so that the number of return observations equals the number of days on which prices are available. However, we let  $N = TS$  for the log squared return series, which means that the total number of return observations is equal to the product of the number of days and the number of observations in each day. For comparison purposes, we also include the estimates for the log squared original returns with  $N = T$ .

**Table 2.** Long-memory parameter estimates with high-frequency S&P 500 futures data.

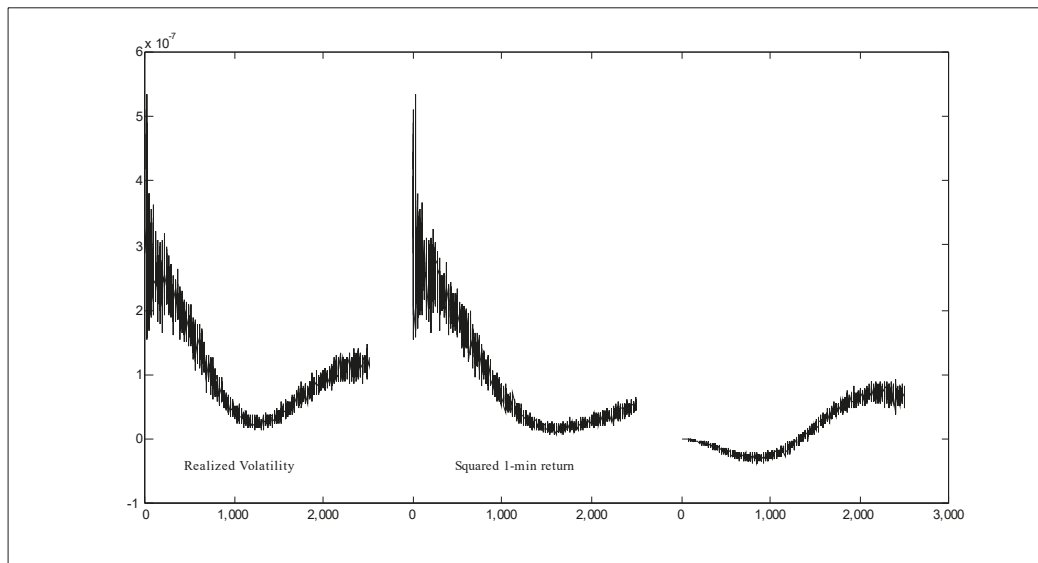
	$\left[\log \left(r_{t,n}^{(k)2}\right)\right]^{(S)}$		$\log \left[r_{t,n}^{(k)2}\right]$		$\log \left\{\left[r_{t,n}^{(k)2}\right]^{(S)}\right\}$			
	$N = T$		$N = T$		$N = TS$		$N = T$	
$k$	SLP	TLP	SLP	TLP	SLP	TLP	SLP	TLP
$k = 1$	0.6231	0.3949	0.6218	0.3765	0.3964	0.0676	0.6768	0.4842
$k = 5$	0.6336	0.3432	0.6323	0.3260	0.4595	0.0603	0.6497	0.4269
$k = 30$	0.6439	0.2421	0.6443	0.2203	0.5275	0.0679	0.6637	0.3072
$k = 330$	0.4835	0.0415	0.4835	0.0415	0.4835	0.0415	0.4835	0.0415

Notes: The table reports the standard log-periodogram regression estimates (SLP) and the trimmed log-periodogram regression estimates (TLP) for the degree of fractional integration for the high-frequency S&P 500 futures returns data. The sample period is from 7 October 1986 to 2 March 2007. The sample size for 1-min returns is 1,650,000 ( $T = 5000$  and  $s = 330$ ). The rows labelled  $k = 1, k = 5, k = 30$  and  $k = 330$  refer to the 1-min return data, 5-min return data, 30-min return data and daily return data, respectively. The columns labelled  $[\log(r_{t,n}^{(k)2})]^{(S)}$ ,  $\log[r_{t,n}^{(k)2}]$  and  $\log[r_{t,n}^{(k)2}]^{(S)}$  refer to the  $S$ -period aggregation of the logarithmic transformation of squared  $k$ -min returns, the logarithmic transformation of the original squared  $k$ -min returns and logarithmic transformation of the realized daily volatility aggregated from squared  $k$ -min returns over a day, respectively.  $S$  denotes the number of  $k$ -min returns per day and we have  $s = kS$ . The estimates are based on the bandwidth  $m = N^{0.5}$  for the log-periodogram regression and  $(l, m) = (N^{0.65}, N^{0.9})$  for the trimmed log-periodogram regression. For the estimates of aggregated log squared  $k$ -min returns and realized volatility, we set  $N = T$ . We report results for both  $N = T \times S$  and  $N = T$  for the estimate obtained using the log squared return series.

Some interesting results in Table 2 are worth notice. First, similar to the results in Table 1, with the same bandwidth set to  $N = T$ , the estimates for the log realized volatility series, which is the aggregated squared returns, are approximately equal to the estimates for the log squared original returns. For instance, when  $k = 1$  and  $S = 330$ , the trimmed LP estimator for the log realized return volatility is 0.68 while the corresponding estimate for the log squared original returns is 0.62.

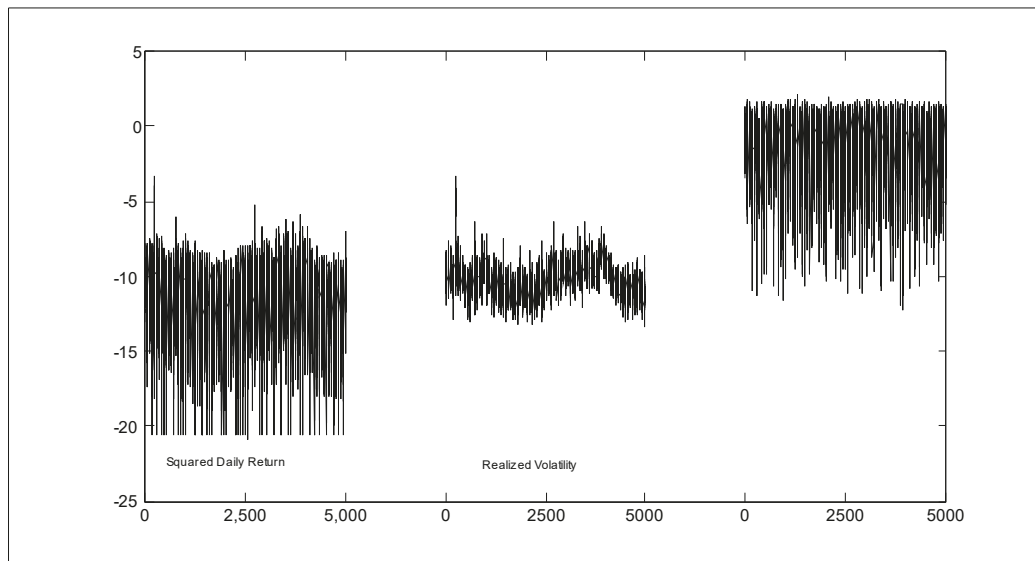
Figure 2 shows the periodograms of the realized volatility obtained from 1-min return series (left),  $s$  times the periodograms of the squared 1-min return series (middle), as well as the difference between them (right), with 2500 frequency indices. The 2500th frequency index corresponds to the frequencies  $\pi$  and  $\pi/330$  for the realized volatility and the squared 1-min returns, respectively. The periodogram of the realized volatility (left) and that of the squared 1-min returns (middle) exhibit very similar values for frequency indices smaller than 1000, especially for frequency indices close to zero. The difference between the periodogram for the realized volatility and the 1-min returns (right) is slight. These results can be explained by the fact that the realized volatility series is here the 330-period non-overlapping aggregation of the squared 1-min return series, and their periodograms exhibit approximately the same values for small frequency indices, as stated in Lemmas 1–2.

Second, for the log realized volatility series, the LP estimates decrease as  $k$  increases, i.e., they are smaller when the return interval is longer. With daily returns ( $k = 330, S = 1$ ), so that the log realized volatility series is the log squared daily return series, the standard and trimmed LP estimates are 0.48 and 0.04. In particular, the trimmed LP estimate is close to zero, indicating the (near) absence of long-memory, while the standard LP estimate is large, consistent with a RLS process. Of interest is the fact that for  $k = 1, 5, 30$  all estimates are similar when the same bandwidth is used. This accords with the theoretical results that the estimates are invariant to the aggregation level. When  $k = 330$ , i.e., daily data, the estimates are somewhat smaller. This feature can be explained by the fact that as the aggregation level increases the spectral density function is contaminated by noise, which henceforth reduces the estimate, see Remark 2.

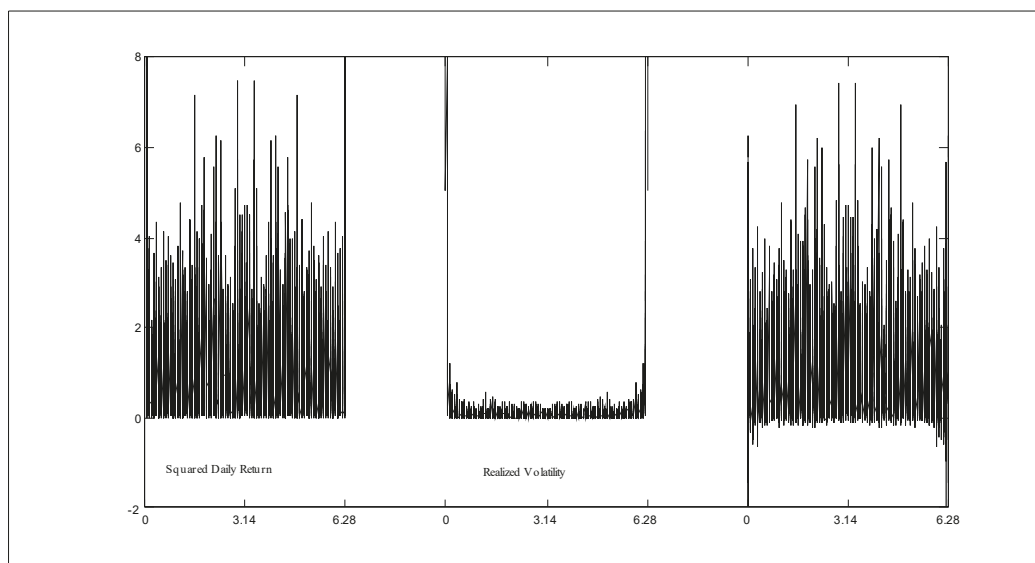


**Figure 2.** The periodogram of the realized volatility obtained from 1-min returns (**left**), the squared 1-min returns multiplied by 330 (**middle**), and the difference between them (**right**) for the high-frequency S&P 500 returns data. The sample period is from 7 October 1986 to 2 March 2007.

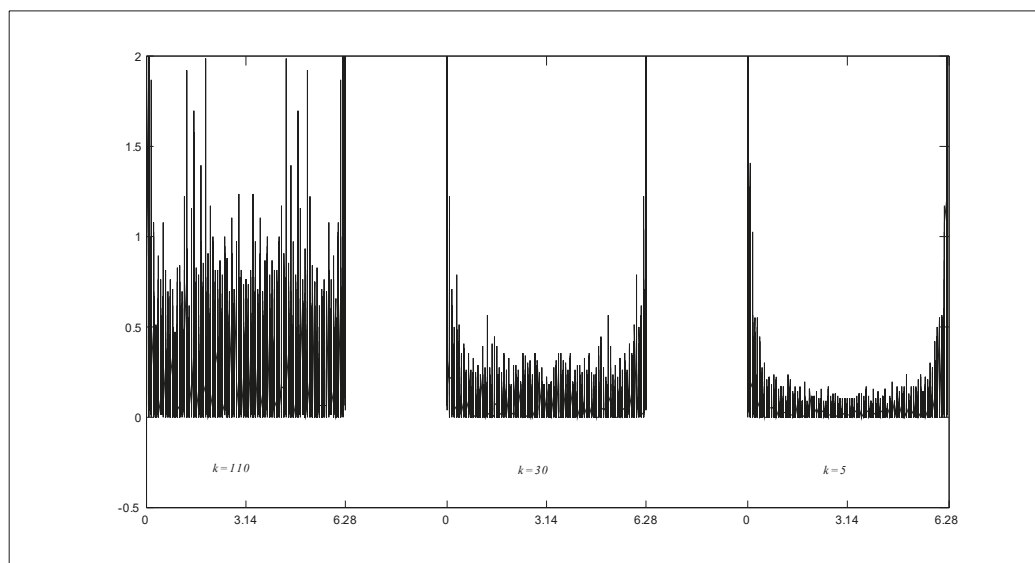
Combining the equation for the squared daily returns (2) and that for the realized volatility (4), we know, as discussed in Section 3, that both the realized volatility and the squared daily returns contain the same information about long memory. However, the squared daily returns contain a larger noise component than does the realized volatility. Figure 3 shows the log squared daily return series (left), the log realized volatility obtained from 1-min return series (middle), and the difference between them (right). We can see that the log squared daily return series (left) exhibits larger variance than the log realized volatility series (middle). No pattern is seen in the difference (right), and it simply seems to be noise. Figure 4 shows the periodograms of the log squared daily returns (left) and the log realized volatility obtained from 1-min return (middle), as well as the difference between them (right). Note that the periodogram of the log squared daily returns (left) is much larger than that of the log realized volatility constructed from 1-min returns (middle) except for the first few frequencies near zero. These results are consistent with Equation (5) and with the presence of RLS given the very large values near frequency zero. For those frequencies near zero, both periodograms show very large values. Similar to what was shown in Figure 3, the difference (right) appears to be caused by a white noise process. In addition, we can see that the white noise process is dominant in the periodogram of the log squared daily return series (left). As shown in Figure 5, similar results occur for the periodograms of the log realized volatility obtained from 110-min returns (left), 30-min returns (middle) and 5-min returns (right). The periodograms exhibit smaller values when higher frequency data are used, except for the first few frequencies indices near zero. For these, the periodograms are likely determined by low-frequency contamination, for example, random level shifts (Perron and Qu 2010 and McCloskey and Perron 2013). In general, the values of the periodograms are much larger for frequency indices near zero, which can be explained by the fact that the impact of the random level shifts dominates that of the noise process.



**Figure 3.** Log squared daily returns (**left**), log realized volatility obtained from 1-min returns (**middle**) and the difference (**right**) for the high-frequency S&P 500 returns data. The sample period is from 7 October 1986 to 2 March 2007.



**Figure 4.** The periodogram of the log squared daily returns (**left**), the log realized volatility obtained from 1-min returns (**middle**), and the difference between them (**right**) for the high-frequency S&P 500 returns data. The sample period is from 7 October 1986 to 2 March 2007.



**Figure 5.** The periodogram of the log realized volatility obtained from 110-min returns (**left**), 30-min returns (**middle**), and 5-min returns (**right**) for the high-frequency S&P 500 returns data. The sample period is from 7 October 1986 to 2 March 2007.

Third, when larger bandwidths ( $N = T \times S$ ) are used to estimate the memory parameter for the log squared original returns, the trimmed LP estimates are close to zero, indicating the (near) absence of long memory, while the standard LP estimates are near the non-stationary region, regardless of the length of the return intervals. These results are consistent with those of [Perron and Qu \(2010\)](#), [Lu and Perron \(2010\)](#), [McCloskey and Perron \(2013\)](#) and [Varneskov and Perron \(2018\)](#). This is an important feature, which shows the importance of the bandwidth selection, in particular in selecting a value large enough. The use of the aggregated squared  $k$  minutes returns allow much more flexibility in the possible choice of the bandwidth, so that we can always use  $N = TS$ , which was suggested by [Souza \(2008\)](#) as leading to improved estimates. Such choices are not possible when using log realized volatility so that the estimator is much more influenced by the low frequencies thereby inducing larger estimates, regardless of whether the true process is RLS or pure long-memory. Hence, we view the estimates obtained with the aggregated squared  $k$  minutes returns and a large bandwidth as being the most reliable, indicating the near absence of long-memory. This is reinforced by the fact that these estimates are (very nearly) the same across aggregation levels, showing robustness to aggregation at any level.

## 5. Conclusions

We showed in this paper that the squared low-frequency returns can be expressed in terms of the temporal aggregation of a high-frequency series. We built a bridge between the spectral density function of squared low-frequency and high-frequency returns. Furthermore, we analyzed the properties of the spectral density function of realized volatility, constructed from squared returns with different frequencies under temporal aggregation. The theoretical findings were illustrated through the analysis of both low-frequency daily S&P 500 returns from 1928 to 2011 and high-frequency 1-min S&P 500 returns from 1986 to 2007. Overall, the results indicate that random level shifts are needed to explain the empirical features documented. For the low frequency data on S&P 500 returns, one cannot infer whether the noise is stationary long memory. On the other hand, a long-memory process appears needed to explain the features related to high frequency S&P 500 futures. The differences may be due to the difference in the asset, i.e., futures versus spot returns. It is a feature for which we have no explanation. More work is needed to better understand the differences. This is important for various aspects of financial modeling, including the pricing of options.



**Author Contributions:** Conceptualization, P.P. and W.S.; formal analysis, P.P. and W.S.; investigation, P.P. and W.S.; methodology, P.P. and W.S.; writing—original draft, P.P. and W.S.; writing—review & editing, P.P. and W.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** We wish to thank Shinsuke Ikeda for kindly sharing high frequency cleaned data for the S&P 500 futures. We are also grateful to Zhongjun Qu for useful comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

We first state a lemma that will be used in subsequent proofs.

**Lemma A1** (Souza (2005), Hassler (2011)). Let  $v_t$  be a covariance stationary discrete-time process with spectral density function  $f_v(\lambda)$ , the spectral density of the  $k$ -period aggregation,  $v^{(k)}$ , is,

$$f_{v^{(k)}}(\lambda) = \frac{1}{k} \left| \sin\left(\frac{\lambda}{2}\right) \right|^2 \sum_{j=0}^{k-1} \left| \sin\left(\frac{\lambda + 2j\pi}{2k}\right) \right|^{-2} f_v\left(\frac{\lambda + 2j\pi}{2k}\right)$$

for  $\lambda \in [0, 2\pi]$ .

**Proof of Proposition 1.** As shown in Section 2, the squared  $k$ -period return  $r_{t,p}^{(k)2}$  is given by

$$\begin{aligned} r_{t,p}^{(k)2} &= \frac{1}{s} z_{t,p}^2 \left( \sum_{j=0}^{k-1} L^j \right) h_{kp}^2 \\ &= \frac{1}{s} \left[ \left( \sum_{j=0}^{k-1} L^j \right) h_{kp}^2 + (z_{t,p}^2 - 1) \left( \sum_{j=0}^{k-1} L^j \right) h_{kp}^2 \right] \\ &= \frac{1}{s} \left[ \left( \sum_{j=0}^{k-1} L^j \right) (h_{t,kp}^2 - E h_{t,kp}^2) + \left( \sum_{j=0}^{k-1} L^j \right) E h_{t,kp}^2 \right] \\ &\quad + \frac{1}{s} \left[ (z_{t,p}^2 - 1) \left( \sum_{j=0}^{k-1} L^j \right) (h_{t,kp}^2 - E h_{t,kp}^2) + (z_{t,p}^2 - 1) \left( \sum_{j=0}^{k-1} L^j \right) E h_{t,kp}^2 \right] \\ &= \frac{1}{s} \left[ \sum_{j=0}^{k-1} L^j y_{t,kp} + \sum_{j=0}^{k-1} L^j E h_{t,kp}^2 \right] \\ &\quad + \frac{1}{s} \left[ (z_{t,p}^2 - 1) \sum_{j=0}^{k-1} L^j y_{t,kp} + (z_{t,p}^2 - 1) \sum_{j=0}^{k-1} L^j E h_{t,kp}^2 \right] \end{aligned}$$

which can be expressed as

$$r_{t,p}^{(k)2} = \frac{1}{s} \left[ [y_{t,n}]_p^{(k)} + k E h_{t,n}^2 + [y_{t,n}]_p^{(k)} (z_{t,p}^2 - 1) + k E h_{t,n}^2 (z_{t,p}^2 - 1) \right]$$

From Lemma A1, the spectral density of the squared  $k$ -period return  $r_{t,p}^{(k)2}$  is given by

$$\begin{aligned} f_{r_{t,p}^{(k)2}} &= \frac{1}{s^2 k} \left| \sin\left(\frac{\lambda}{2}\right) \right|^2 \sum_{j=0}^{k-1} \left| \sin\left(\frac{\lambda + 2j\pi}{2k}\right) \right|^{-2} f_y\left(\frac{\lambda + 2j\pi}{2k}\right) \\ &\quad + \frac{1}{s^2 \pi} \left[ \text{var}\left(y_{t,n}^{(k)}\right) + k^2 \left(E h_{t,n}^2\right)^2 \right]. \end{aligned}$$

□

**Proof of Proposition 2.** The realized volatility constructed from the  $k$ -period returns  $r_{t,p}^{(k)}$  is

$$\begin{aligned} RV_t &= \sum_{p=1}^S r_{t,p}^{(k)2} \\ &= \frac{1}{s} \sum_{p=1}^S \left[ \sum_{j=0}^{k-1} L^j y_{t,kp} + \sum_{j=0}^{k-1} L^j Eh_{t,kp}^2 \right] \\ &\quad + \frac{1}{s} \sum_{p=1}^S \left[ \left( z_{t,p}^2 - 1 \right) \sum_{j=0}^{k-1} L^j y_{t,kp} + \left( z_{t,p}^2 - 1 \right) \sum_{j=0}^{k-1} L^j Eh_{t,kp}^2 \right] \\ &= \frac{1}{s} \left[ \sum_{p=1}^S \sum_{j=0}^{k-1} L^j y_{t,kp} + \sum_{p=1}^S \sum_{j=0}^{k-1} L^j Eh_{t,kp}^2 \right] \\ &\quad + \frac{1}{s} \left\{ \sum_{p=1}^S \left[ \left( z_{t,p}^2 - 1 \right) \sum_{j=0}^{k-1} L^j y_{t,kp} \right] + \sum_{p=1}^S \left[ \left( z_{t,p}^2 - 1 \right) \sum_{j=0}^{k-1} L^j Eh_{t,kp}^2 \right] \right\} \\ &= \frac{1}{s} \left[ \sum_{p=1}^S L^j y_{t,n} + s Eh_{t,kp}^2 + \sum_{p=1}^S \left( z_{t,p}^2 - 1 \right) \left( y_{t,n}^{(k)} \right)_p + k Eh_{t,kp}^2 \sum_{p=1}^S \left( z_{t,p}^2 - 1 \right) \right] \end{aligned}$$

which can be expressed as

$$RV_t = \frac{1}{s} \left\{ [y_{t,n}]_p^{(s)} + s Eh_{t,kp}^2 + \left[ \left( z_{t,p}^2 - 1 \right) y_{t,n}^{(k)} \right]^{(s)} + k Eh_{t,kp}^2 \left[ \left( z_{t,p}^2 - 1 \right) \right]^{(s)} \right\}.$$

From Lemma A1, the spectral density of the realized volatility constructed from the  $k$ -period returns  $r_{t,p}^{(k)}$ ,  $RV_t$ , is given by

$$\begin{aligned} f_{RV} &= \frac{1}{s^3} \left| \sin \left( \frac{\lambda}{2} \right) \right|^2 \sum_{j=0}^{s-1} \left| \sin \left( \frac{\lambda + 2j\pi}{2s} \right) \right|^{-2} f_y \left( \frac{\lambda + 2j\pi}{2s} \right) \\ &\quad + \frac{1}{s^2 \pi} \left[ Svar \left( y_{t,n}^{(k)} \right) + Sk^2 \left( Eh_{t,n}^2 \right)^2 \right]. \end{aligned}$$

□

**Proof of Lemma 1.** We have the following relation

$$\begin{aligned} &f_{[r_{t,p}^{(k)2}]^{(s)}}(\lambda) - S f_{r_{t,p}^{(k)2}}(\lambda/S) = \\ &\frac{1}{s} \left| \sin \left( \frac{\lambda}{2} \right) \right|^2 \sum_{j=0}^{s-1} \left| \sin \left( \frac{\lambda + 2j\pi}{2s} \right) \right|^{-2} f_y \left( \frac{\lambda + 2j\pi}{2s} \right) \\ &\quad - \frac{S}{k} \left| \sin \left( \frac{\lambda}{2S} \right) \right|^2 \sum_{j=0}^{k-1} \left| \sin \left( \frac{\lambda + 2j\pi}{2s} \right) \right|^{-2} f_y \left( \frac{\lambda + 2j\pi}{2s} \right) \\ &= \frac{1}{s^2} [U(\lambda) + V(\lambda)] \end{aligned}$$

where

$$\begin{aligned} V(\lambda) &= \frac{1}{s} \left| \sin \left( \frac{\lambda}{2} \right) \right|^2 \sum_{j=1}^{s-1} \left| \sin \left( \frac{\lambda + 2j\pi}{2s} \right) \right|^{-2} f_y \left( \frac{\lambda + 2j\pi}{2s} \right) \\ &\quad - \frac{S}{k} \left| \sin \left( \frac{\lambda}{2S} \right) \right|^2 \sum_{j=1}^{k-1} \left| \sin \left( \frac{\lambda + 2j\pi}{2ks} \right) \right|^{-2} f_y \left( \frac{\lambda + 2j\pi}{2ks} \right) \\ &\rightarrow 0 \end{aligned}$$

since  $|\sin(\lambda/2)| \rightarrow 0$  for  $\lambda \rightarrow 0$ , and

$$\begin{aligned}
U(\lambda) &= \frac{1}{s} \left| \sin\left(\frac{\lambda}{2}\right) \right|^2 \left| \sin\left(\frac{\lambda}{2s}\right) \right|^{-2} f_y\left(\frac{\lambda}{2s}\right) - \frac{S}{k} \left| \sin\left(\frac{\lambda}{2S}\right) \right|^2 \left| \sin\left(\frac{\lambda}{2s}\right) \right|^{-2} f_y\left(\frac{\lambda}{2ks}\right) \\
&= \frac{1}{s} \left(\frac{\lambda}{2}\right)^2 \left| \sin\left(\frac{\lambda}{2s}\right) \right|^{-2} f_y\left(\frac{\lambda}{2s}\right) - \frac{S}{k} \left(\frac{\lambda}{2S}\right)^2 \left| \sin\left(\frac{\lambda}{2s}\right) \right|^{-2} f_y\left(\frac{\lambda}{2ks}\right) \\
&\rightarrow 0
\end{aligned}$$

since  $|\sin(\lambda/2)| \rightarrow^p \lambda/2$  and  $\sin(\lambda/2S) \rightarrow^p \lambda/2S$  for  $\lambda \rightarrow 0$ .  $\square$

**Proof of Lemma 2.** After the  $k$ -period non-overlapping temporal aggregation, the random level shift component,  $\eta_q \sim i.i.d. \left(0, k^2 \sigma_\eta^2\right)$ , and the total number of level shifts is unchanged as long as the sample size is large enough. Therefore,

$$u_{T,q}^{(k)} = \sum_{j=1}^t \delta_{T,q}, \quad \delta_{T,q} = \pi_{T,q} \eta_q$$

where  $\eta_q \sim iid \left(0, k^2 \sigma_\eta^2\right)$  and  $\pi_{T,q} \sim i.i.d. Bernoulli(kp/T, 1)$ .

According to Proposition 3 of Perron and Qu (2010), the limit of the expectation of the periodogram of the high-frequency time series has the following form

$$\lim_{T \rightarrow \infty} T^{-1} E \left[ I_{u^{(1)},j} \right] = \frac{p \sigma_\eta^2}{4\pi^3} \frac{1}{j^2}$$

Hence, after  $k$ -period non-overlapping temporal aggregation, the limit of the expectation of the periodogram is

$$\lim_{T \rightarrow \infty} T^{-1} E \left[ I_{u^{(k)},j} \right] = k \frac{p \sigma_\eta^2}{4\pi^3} \frac{1}{j^2}$$

which implies that

$$\lim_{T \rightarrow \infty} E \left[ I_{u^{(k)},j} \right] - k \lim_{T \rightarrow \infty} E \left[ I_{u^{(1)},j} \right] = 0 \quad \text{for } \lambda \rightarrow 0.$$

$\square$

## References

- Bollerslev, Tim, and Jonathan H. Wright. 2000. Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data. *Journal of Econometrics* 98: 81–106. [\[CrossRef\]](#)
- Chambers, Marcus J. 1998. Long-memory and aggregation in macroeconomic time series. *International Economic Review* 39: 1053–72. [\[CrossRef\]](#)
- Deo, Rohit, Clifford Hurvich, and Yi Lu. 2006. Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics* 131: 29–58. [\[CrossRef\]](#)
- Diebold, Francis X., and Atsushi Inoue. 2001. Long memory and regime switching. *Journal of Econometrics* 105: 131–59. [\[CrossRef\]](#)
- Ding, Zhuanxin, Clive W. J. Granger, and Robert F. Engle. 1993. A long memory property of stock market return and a new model. *Journal of Empirical Finance* 1: 86–106. [\[CrossRef\]](#)
- Engle, Robert F., and Aaron D. Smith. 1999. Stochastic permanent breaks. *Review of Economics and Statistics* 81: 533–74. [\[CrossRef\]](#)
- Geweke, John, and Susan Porter-Hudak. 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4: 221–38. [\[CrossRef\]](#)
- Gourieroux, Christian, and Joann Jasiak. 2001. Memory and infrequent breaks. *Economics Letters* 70: 29–41. [\[CrossRef\]](#)

- Granger, Clive W. J., and Namwon Hyung. 2004. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* 11: 399–421.
- Granger, Clive W. J., and Roselyne Joyeux. 1980. An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1: 15–39. [\[CrossRef\]](#)
- Hassler, Uwe. 2011. Estimation of fractional integration under temporal aggregation. *Journal of Econometrics* 162: 240–47. [\[CrossRef\]](#)
- Hosking, Jonathan R. M. 1981. Fractional differencing. *Biometrika* 68: 165–76. [\[CrossRef\]](#)
- Ikeda, Shin S. 2015. Two scale realized kernels: A univariate case. *Journal of Financial Econometrics* 13: 126–65. [\[CrossRef\]](#)
- Lu, Yang K., and Pierre Perron. 2010. Modeling and forecasting stock return volatility using a random level shift model. *Journal of Empirical Finance* 17: 138–56. [\[CrossRef\]](#)
- McCloskey, Adam, and Pierre Perron, P. 2013. Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory* 29: 1196–237. [\[CrossRef\]](#)
- Ohanissian Arek, Jeffrey R. Russell, and Ruey S. Tsay. 2008. True or spurious long memory? a new test. *Journal of Business and Economic Statistics* 26: 161–75. [\[CrossRef\]](#)
- Perron, Pierre. 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57: 1361–401. [\[CrossRef\]](#)
- Perron, Pierre. 1990. Testing for a unit root in a time series regression with a changing mean. *Journal of Business and Economic Statistics* 8: 153–62.
- Perron, Pierre, and Zhongjun Qu. 2007. *An Analytical Evaluation of the Log-Periodogram Estimate in the Presence of Level Shifts*. Working paper. Boston: Boston University.
- Perron, Pierre, and Zhongjun Qu. 2010. Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business and Economic Statistics* 28: 275–90. [\[CrossRef\]](#)
- Robinson, Peter M. 1995. Log periodogram regression of time series with long range dependence. *Annals of Statistics* 23: 1048–72. [\[CrossRef\]](#)
- Souza, Leonardo Rocha. 2005. A note on Chambers's "Long-memory and aggregation in macroeconomic time series". *International Economic Review* 46: 1059–62. [\[CrossRef\]](#)
- Souza, Leonardo Rocha. 2007. Temporal aggregation and bandwidth selection in estimating long memory. *Journal of Time Series Analysis* 28: 701–22. [\[CrossRef\]](#)
- Souza, Leonardo Rocha. 2008. Why aggregate long memory time series? *Econometric Reviews* 27: 298–316. [\[CrossRef\]](#)
- Schwert, G. William. 1990. Indexes of United States stock prices, 1802–1987. *Journal of Business* 63: 399–426. [\[CrossRef\]](#)
- Varneskov, Rasmus T., and Pierre Perron. 2018. Combining long memory and level shifts in modeling and forecasting the volatility of asset returns. *Quantitative Finance* 18: 371–93. [\[CrossRef\]](#)
- Wei, William W. S. 1978. Some consequences of temporal aggregation in seasonal time series models. In *Seasonal Analysis of Economic Time Series*. Edited by A. Zeller. Washington, DC: US Department of Commerce, Bureau of Census.

