Conditions for the Propagation of Memory Parameter from Durations to Counts and Realized Volatility

Rohit Deo* Clifford M. Hurvich* Philippe Soulier† Yi Wang*

May 16, 2007

Abstract

We establish sufficient conditions on durations that are stationary with finite variance and memory parameter $d \in [0, 1/2)$ to ensure that the corresponding counting process $N(t)$ satisfies $\operatorname{Var} N(t) \sim Ct^{2d+1}$ $(C > 0)$ as $t \to \infty$, with the same memory parameter $d \in [0, 1/2)$ that was assumed for the durations. Thus, these conditions ensure that the memory parameter in durations propagates to the same memory parameter in the counts. We then show that any Autoregressive Conditional Duration ACD(1,1) model with a sufficient number of finite moments yields short memory in counts, while any Long Memory Stochastic Duration model with $d > 0$ and all finite moments yields long memory in counts, with the same d. Next, we present a result implying that the only way for a series of counts aggregated over a long time period to have nontrivial autocorrelation is for the counts to have long memory. In other words, aggregation ultimately destroys all autocorrelation in counts, if and only if the counts have short memory. Finally, under assumptions on the pure-jump price process, we show that the memory parameter in durations propagates all the way to the realized volatility, under both calendar-time sampling and transaction-time sampling.

KEYWORDS: Long Memory Stochastic Duration, Autoregressive Conditional Duration, Rosenthal-type Inequality.

*New York University, 44 W. 4'th Street, New York NY 10012 USA

†Université Paris X, 200 avenue de la République, 92001 Nanterre cedex, France

The authors thank Xiaohong Chen and Raymond Thomas for helpful comments and suggestions.

I Introduction

There is a growing literature on long memory in volatility of financial time series. See, e.g., Robinson (1991), Robinson and Henry (1999), Deo and Hurvich (2001), Hurvich, Moulines and Soulier (2005). Long memory in volatility, which has been repeatedly found in the empirical literature (see Breidt, Crato and de Lima 1998, Baillie, Bollerslev and Mikkelsen 1996, Bollerslev and Mikkelsen 1996, Andersen and Bollerslev 1997), plays a key role in the forecasting of realized volatility, defined as the sum of squared high-frequency returns (Andersen, Bollerslev, Diebold and Labys 2001, Barndorff-Nielsen and Shephard 2006, Deo, Hurvich and Lu 2006), and has important implications on option pricing (see Comte and Renault 1998).

Given the increasing availability of transaction-level data it is of interest to explain phenomena observed at longer time scales from equally-spaced returns in terms of more fundamental properties at the transaction level. Engle and Russell (1998) proposed the Autoregressive Conditional Duration (ACD) model to describe the durations between trades, and briefly explored the implications of this model on volatility of returns in discrete time, though they did not determine the persistence of this volatility, as measured, say, by the decay rate of the autocorrelations of the squared returns. Deo, Hsieh and Hurvich (2007) proposed the Long-Memory Stochastic Duration (LMSD) model, and carried out an empirical exploration of the question as to which properties of durations lead to long memory in volatility.

In principle, why should there be a link between the dependence properties of durations and volatility? The durations between transactions determine the number of transactions (counts) in a given interval of time. Thus, any dependence in durations will affect the dependence in the time series of counts. Inspired by the work of Clark (1976), and other related work (Tauchen and Pitts 1983, Epps and Epps 1976, Bollerslev and Jubinski 1999) on the Mixture of Distributions Hypothesis (MDH), financial econometricians now generally accept that dependence in counts is related to dependence in volatility. Overall, then, dependence in durations and volatility are closely linked.

The specification of a model for the durations, together with a model for the transaction-level price changes, constitutes a pure-jump model for the prices. In such a model, it is the transactions that generate price changes. Therefore, the (logarithmic) price change in a given time interval is the accumulation of transaction-level price changes. Furthermore, the total number of transactions up to a given time is simply the counting process induced by the series of intertrade durations. Such models have been considered recently by Oomen (2006), Rydberg and Shephard (2003), among others. Clearly, then, properties of realized volatility calculated from high-frequency returns generated from such a process would also depend on the properties of durations. It is therefore essential to get a better understanding of the theoretical link between durations, counts and realized volatility so that price process models can be compatible with what is observed in empirical data.

We present first a few basic definitions. The collection of time points $\cdots t_{-1} < t_0 \leq 0 < t_1 < t_2 < \cdots$ at which a transaction (say, a trade of a particular stock on a specific market) takes place, comprises a point process, a fact which was exploited by Engle and Russell (1988). These event times $\{t_k\}$ determine a *counting process*,

$$N(t) = Number\ of\ Events\ in\ (0, t].$$

For any fixed time spacing $\Delta t > 0$, one can define the *counts* $\Delta N_s = N(s\Delta t) - N((s-1)\Delta t)$, the number of events in the s'th time interval of width Δt, where $s = 1, 2, \cdots$. The event times $\{t_k\}_{k=-\infty}^{\infty}$ also determine the *durations*, given by $\{\tau_k\}_{k=-\infty}^{\infty}$, $\tau_k = t_k - t_{k-1}$. In this paper, we will say that a stationary series in discrete time (either counts or durations) has long memory with memory parameter $d \in (0, 1/2)$ (or short memory if $d = 0$) if the partial sum of n contiguous values of the series has a variance that behaves as Cn^{2d+1} as $n \to \infty$, where $C > 0$ is a constant. We will say that a stationary counting process $N(t)$ has long memory with memory parameter $d \in (0, 1/2)$ (or short memory if $d = 0$) if $\text{Var}N(t) \sim Ct^{2d+1}$ as $t \to \infty$ where $C > 0$ is a constant. Note that if $N(t)$ has memory parameter $d \in [0, 1/2)$, then so does the series of counts, with the same memory parameter, for any value of Δt.

Deo, Hsieh and Hurvich (2007) analyzed transaction-level data on ten stocks traded on the NYSE. They estimated the memory parameter of the durations between transactions, as well as the counts and squared returns for various choices of Δt, and the daily realized volatility. The estimates were all roughly similar, taking the standard errors into account, typically lying in the range from 0.3 to 0.45. If it could be established theoretically that under certain reasonable conditions the memory parameter (be it positive or zero) propagates unchanged from the durations to the counting process and then to realized volatility, this would have strong implications on the choice of an appropriate model for durations.

Both the ACD and LMSD models imply that the doubly infinite sequence of durations $\{\tau_k\}_{k=-\infty}^{\infty}$ is a stationary time series, i.e., there exists a probability measure P^0 under which the joint distribution of any subcollection of the $\{\tau_k\}$ depends only on the lags between the entries. On the other hand, a point process N on the real line is stationary under the measure P if for every positive integer r the joint distribution of $\{N(A_1+t),\ldots,N(A_r+t)\}$ does not depend on t, where $A_1,\ldots,A_r$ are bounded Borel sets. A fundamental fact about point processes is that in general (a notable exception is the Poisson process) there is no single measure under which both the point process N and the durations $\{\tau_k\}$ are stationary, i.e., in general P and P^0 are not the same. Nevertheless, there is a one-to-one correspondence between the class of measures P^0 that determine a stationary duration sequence and the class of measures P that determine a stationary point process. The measure P^0 corresponding to P is called the *Palm distribution.* The counts are stationary under P, while the durations are stationary under P^0.

An economic interpretation of the distinction between P and P^0 is as follows. If the cumulative number of transactions $N(t)$ is calculated from the opening bell (e.g., 9:00 E.T., Wednesday, July 12'th 2006), then the appropriate measure is P. By contrast, if $N(t)$ is calculated from the first transaction of that trading day, then the appropriate measure would be P^0. For thinly traded stocks, the difference between these two starting points may be quite large.

Deo, Hsieh and Hurvich (2007) pointed out, using a theorem of Daley, Rolski and Vesilo (2000) that if durations are generated by an *ACD* model and if the durations have tail index $\kappa \in (1,2)$ under P^0, then the resulting counting process $N(t)$ has long memory with memory parameter $d \geq 1-\kappa/2$, under P. An open question, however, is the determination of conditions under which the memory parameter of a finite-variance duration sequence propagates unchanged to the counting process. This question is a nontrivial one, since even a time-instantaneous transformation of a Gaussian long-memory process may reduce the memory parameter. See Dittmann and Granger (2002).

In this paper, we will establish sufficient conditions on durations that are stationary with finite variance and memory parameter $d \in [0, 1/2)$ under P^0 to ensure that the corresponding counting process $N(t)$ satisfies $\operatorname{Var} N(t) \sim Ct^{2d+1}$ $(C > 0)$ as $t \to \infty$ under P, with the same memory parameter $d \in [0, 1/2)$ that was assumed for the durations. Thus, these conditions, given in Theorem 1, ensure that the memory parameter in durations propagates to the same memory parameter in counts. Moreover, we show that under a particular model for prices, the same memory parameter propagates further from the counting process to the realized volatility.

We will verify that the sufficient conditions of our Theorem 1 are satisfied for the ACD(1,1) model assuming finite $3 + \epsilon$ moment $(\epsilon > 0)$ of the durations under P^0, and for the LMSD model with any $d \in [0, 1/2)$ assuming that the multiplying shocks have all moments finite. Thus, any ACD(1,1) model with a sufficient number of finite moments yields short memory in the counting process, while any LMSD model with $d > 0$ and all finite moments yields long memory in the counting process. These results for the LMSD and ACD(1,1) models are given in Theorems 2 and 3, respectively. Proposition 1, which is used in proving Theorem 2, provides a Rosenthal-type inequality for moments of absolute standardized partial sums of durations under the LMSD model, and is of interest in its own right.

We then present a result (Theorem 4) implying that if counts have memory parameter $d \in [0, 1/2)$ then further aggregations of these counts to longer time intervals will have a lag-1 autocorrelation that tends to $2^{2d} - 1$ as the level of aggregation grows. Interestingly, this limit is zero if and only if $d = 0$. Thus, one of the important functions of long memory in counts is that it allows the counts to have a non-vanishing autocorrelation even as Δt grows, as was found by Deo, Hsieh and Hurvich (2007) to occur in empirical data. By contrast, short memory in counts implies that counts at long time scales (large Δt) are essentially uncorrelated, in contradiction to what is seen in actual data. To summarize, aggregation ultimately destroys all autocorrelation in counts, if and only if the counts have short memory.

Finally, we provide the link between counts and realized volatility. Using a simple pure-jump model for prices, we show that if the durations have long memory then realized volatility has long memory with the same memory parameter. We also show that if the durations have short memory and finite variance then realized volatility has short memory. Thus, the properties of durations have direct implications on the properties of realized volatility.

The pure-jump model for prices mentioned above assumes that the transaction-level price changes are independent. However, as we discuss in Section IV, the observed transaction-level returns exhibit autocorrelation at short lags. Hence we relax the independence to the far weaker assumption that the price changes are q-dependent, and then replacing the realized volatility by the transaction-time-sampling version advocated by Oomen (2006), we show that this alternative version of realized volatility inherits the same memory parameter as the durations.

II Theorems on the propagation of the memory parameter

Let E, E^0, Var, Var^0 denote expectations and variances under P and P^0, respectively. Define $\mu = E^0(\tau_k)$ and $\lambda = \frac{1}{\mu}$. Our main theorem uses the assumption that P^0 is $\{\tau_k\}$-mixing, defined as follows. Let $\mathcal{N} = \sigma(\{\tau_k\}_{k=-\infty}^{\infty})$ and $\mathcal{F}_n = \sigma(\{\tau_k\}_{k=n}^{\infty})$. Following Nieuwenhuis (1989, p. 597), we say that P^0 is $\{\tau_k\}$-mixing if

$$\lim_{n\to\infty} \sup_{B\in\mathcal{N}\cap\mathcal{F}_n} |P^0(A\cap B) - P^0(A)P^0(B)| = 0$$

for all $A \in \mathcal{N}$.

Our theorem also uses the notion of α-mixing, also known as strong mixing (see Bradley 2005, Section 2.1).

Theorem 1 *Let $\{\tau_k\}$ be a duration process such that the following conditions hold:*

i) $\exists\ d \in [0, \frac{1}{2})$ *such that*

$$Y_n(s) = \frac{\sum_{k=1}^{\lfloor ns\rfloor}(\tau_k - \mu)}{n^{1/2+d}}, \qquad s \in [0,1]$$

converges weakly to $\sigma B_{1/2+d}(\cdot)$ under P^0, where $\sigma > 0$ and $B_{1/2+d}(\cdot)$ is fractional Brownian motion if $0 < d < \frac{1}{2}$ or standard Brownian motion $B_{1/2} = B$ if $d = 0$.

ii) Consider the d in i) above. If $d \in (0, 1/2)$, then P^0 is $\{\tau_k\}$-mixing. If $d = 0$, then $\{\tau_k\}$ is exponential strong mixing.

iii) For the d in i) above,

$$\sup_n E^0 \left| \frac{\sum_{k=1}^n (\tau_k - \mu)}{n^{1/2+d}} \right|^p < \infty \qquad \begin{cases} for\ all\ p > 0,\ if\ d \in (0, \frac{1}{2}) \\ for\ p = 3 + \epsilon,\ \epsilon > 0,\ if\ d = 0 \end{cases} .$$

Then the induced counting process $N(t)$ satisfies Var$N(t) \sim Ct^{2d+1}$ under P as $t \to \infty$ where $C > 0$.

Remark 1: Assumption *iii*) implies that $Y_n^2(1)$ is uniformly integrable. This, together with *i*) implies that $\text{Var}^0[Y_n(1)] \sim C$ for some $C > 0$, and hence that the durations have long (or short) memory with memory parameter $d \in [0, 1/2)$.

Remark 2: If $d = 0$, the proof of Theorem 1 implies that $\text{Var}^0 N(t) \sim Ct$ (i.e., under the Palm measure P^0) even when Assumption iii) is weakened to $p = 2 + \epsilon$ for $\epsilon > 0$. Nevertheless, we find it preferable to present and prove the result in Theorem 1 in terms of Var$N(t)$, i.e., under the measure P, since it is under this measure that $N(t)$ is a stationary point process. The need for the more stringent condition in our Theorem 1 can be explained as follows. In order for $E[\tau_k^{2+\epsilon}] < \infty$, it is necessary in general to have $E^0[\tau_k^{3+\epsilon}] < \infty$. See Equation (1.2.25) of Baccelli and Brémaud (2003), which we use in our proof.

Remark 3: Inspection of the proof of Theorem 1 reveals that if $d > 0$, only $\frac{5-2d}{1-2d} + \delta$ finite moments are needed, where δ is arbitrarily small. The closer d is to 1/2, the larger the number of finite moments required.

Remark 4: As pointed out by Nieuwenhuis (1989), if $\{\tau_k\}$ is strong mixing under P^0 then P^0 is $\{\tau_k\}$-mixing. Nevertheless, this weaker form of mixing is essential for our purposes in the case $d > 0$ since even Gaussian long-memory processes are not strong mixing. See Viano, Deniau and Oppenheim (1995).

We now present two useful models for durations and give results stating that the conditions of Theorem

1 apply for these models, thereby yielding the properties of the variance of the induced counting processes.

A LMSD Process

A latent-variable model for durations is the Stochastic Duration (SCD) model of Bauwens and Veredas (2004). The model is given by

$$\tau_k = e^{h_k}\epsilon_k,$$

$$h_k = \omega + \beta h_{k-1} + e_k,$$

where $\omega \in \mathbb{R}$, $|\beta| < 1$, the $\{e_k\}$ are *iid* $N(0, \sigma^2)$ and the $\{\epsilon_k\}$ are *iid* with unit mean and positive support. This model is analogous to the widely-used Stochastic Volatility (SV) model for returns (see Harvey, 1998). Bauwens and Veredas (2004) estimated the SCD model to data and found that the autoregressive coefficient β in the latent process $\{h_k\}$ is typically extremely close to 1. Based on a semiparametric analysis, Deo, Hsieh and Hurvich (2007) concluded that the intertrade durations for stock prices have long memory, so they proposed the Long Memory Stochastic Duration (LMSD) model, in which the $\{h_k\}$ possess long memory. The generalization from the SCD to the LMSD model for durations is analogous to that from the SV to the Long Memory Stochastic Volatility (LMSV) model for returns of Harvey (1998) and Breidt, Crato and de Lima (1998).

Following Deo, Hsieh and Hurvich (2007), we now define the LMSD process $\{\tau_k\}_{k=-\infty}^{\infty}$ for $d \in [0, \frac{1}{2})$ as

$$\tau_k = e^{h_k}\epsilon_k$$

where under P^0 the $\epsilon_k \geq 0$ are *i.i.d.* with all moments finite, and $h_k = \sum_{j=0}^{\infty} b_j e_{k-j}$, the $\{e_k\}$ are *i.i.d.* Gaussian with zero mean, independent of $\{\epsilon_k\}$, and

$$b_j \sim \begin{cases} Cj^{d-1} & if\ d \in (0, \frac{1}{2}) \\ Ca^j,\ |a| < 1 & if\ d = 0 \end{cases}$$

$(C \neq 0)$ as $j \to \infty$; for $d = 0$ we further assume that the spectral density of $\{h_k\}$ is bounded away from zero. Any stationary, invertible ARFIMA process $\{h_k\}$ with $d \in [0, 1/2)$ would satisfy the above assumptions. Note that for convenience, we nest the short-memory case ($d = 0$) within the LMSD model, so that the allowable values for d in this model are $0 \leq d < 1/2$. In their empirical analysis, Deo, Hsieh and Hurvich (2007) assumed that the $\{h_k\}$ follow an $ARFIMA(1, d, 0)$ process. In the theoretical results of this paper, we do not make any specific parametric assumption on the $\{h_k\}$.

The following theorem establishes that long memory propagates unchanged from LMSD durations to the counting process.

Theorem 2 *If the durations $\{\tau_k\}$ are generated by the LMSD process with $d \in [0, 1/2)$, then the conditions of Theorem 1 are satisfied and therefore the induced counting process $N(t)$ satisfies $VarN(t) \sim Ct^{2d+1}$ under P as $t \to \infty$ where $C > 0$.*

To establish Theorem 2, we will use the following Rosenthal-type inequality, which is of independent interest.

Proposition 1 *For durations $\{\tau_k\}$ generated by the LMSD process with $d \in [0, \frac{1}{2})$, for any fixed positive integer $p \geq 2$, $E^0\{|y_n - E^0(y_n)|^p\}$ is bounded uniformly in n, where*

$$y_n = \frac{\sum_{k=1}^{n} \tau_k}{n^{1/2+d}} \qquad .$$

B ACD(1,1) Process

The ACD process was introduced by Engle and Russell (1998). While the LMSD model builds dependence in the durations through an unobservable latent variable process, the ACD model is observation driven. The ACD model treats the conditional mean of durations in the same way that the GARCH model treats the conditional variance of returns. The simplest version is the ACD(1,1) process given by $\{\tau_k\}_{k=-\infty}^{\infty}$ as

$$\begin{aligned} \tau_k &= \psi_k \epsilon_k \\ \psi_k &= \omega + \alpha \tau_{k-1} + \beta \psi_{k-1} \end{aligned}$$

with $\omega > 0, \alpha > 0, \beta \geq 0$ and $\alpha + \beta < 1$, where under P^0, $\epsilon_k \geq 0$ are *i.i.d.* with mean 1. We will assume further that under P^0, ϵ_k has a density g_ϵ such that $\int_0^\theta g_\epsilon(x)dx > 0, \forall\, \theta > 0$ and $E^0(\tau_k^{3+\epsilon}) < \infty$ for some $\epsilon > 0$. This last assumption entails further restrictions on α and β. See Nelson (1990).

Nelson (1990) guarantees the existence of the doubly-infinite ACD(1,1) process $\{\tau_k\}_{k=-\infty}^{\infty}$, which in our terminology is stationary under P^0.

The next theorem shows that short memory propagates from $ACD(1,1)$ durations (with sufficiently many finite moments) to the counting process.

Theorem 3 *Suppose that the durations $\{\tau_k\}$ are generated by the ACD(1,1) model, with the additional assumptions stated above. Then the conditions of Theorem 1 are satisfied and therefore the induced counting process $N(t)$ satisfies $VarN(t) \sim Ct$ under P as $t \to \infty$ where $C > 0$.*

III Autocorrelation of Aggregated Counts

The following elementary result relates the memory parameter of a stationary process to the lag-1 autocorrelation of partial sums as the level of aggregation grows.

Theorem 4 *Let $\{X_t\}$ be a stationary process such that $Var(\sum_{t=1}^n X_t) \sim Cn^{1+2d}$ as $n \to \infty$, where $C \neq 0$ and $d \in [0, 1/2)$. Then*

$$\lim_{n\to\infty} Corr\left[\sum_{t=1}^{n} X_t, \sum_{t=n+1}^{2n} X_t\right] = 2^{2d} - 1.$$

Proof:

$$\text{Var}\left[\sum_{t=1}^{2n} X_t\right] = 2\,\text{Var}\left[\sum_{t=1}^{n} X_t\right] + 2\,\text{Cov}\left[\sum_{t=1}^{n} X_t, \sum_{t=n+1}^{2n} X_t\right].$$

Thus,

$$\text{Cov}\left[\sum_{t=1}^{n} X_t, \sum_{t=n+1}^{2n} X_t\right] = .5\left(\text{Var}\left[\sum_{t=1}^{2n} X_t\right] - 2\text{Var}\left[\sum_{t=1}^{n} X_t\right]\right).$$

The result follows by noting that $\lim_{n\to\infty} n^{-2d-1}\mathrm{Var}(\sum_{t=1}^n X_t) = C$, where $C \neq 0$. $\square$

This theorem has an interesting practical interpretation. If we write $X_k = N[k\Delta t] - N[(k-1)\Delta t]$ where $\Delta t > 0$ is fixed, then X_k represents the number of events (count) in a time interval of width Δt, e.g. one minute. Thus, $\sum_{k=1}^n X_k$ is the number of events in a time interval of length n minutes, e.g. one day. The theorem implies that as the level of aggregation (n) increases, the lag-1 autocorrelation of the aggregated counts will approach a nonzero constant if and only if the non-aggregated count series $\{X_k\}$ has long memory. In other words, the only way for a series of counts over a long time period to have nontrivial autocorrelation is for the short-term counts to have long memory. Since in practice long-term counts do have substantial autocorrelation (see Deo, Hsieh and Hurvich 2007), it is important to use only the models for durations that imply long memory in the counting process. Examples of such models include the LMSD model (see Theorem 2), and ACD models with infinite variance (see Daley, Rolski and Vesilo, 2000).

IV The Link Between Counts and Realized Volatility

To establish a link between counts and volatility it is necessary first to assume a continuous-time model for prices which incorporates the counting process. We will start with a simple pure-jump model (to be generalized subsequently),

$$\log P(t) = \log P(0) + \sum_{j=1}^{N(t)} \xi_j, \tag{1}$$

where $P(t)$ is the price at time t, $N(t)$ is the number of transactions up to time t, and the $\{\xi_j\}$ are *i.i.d.*, independent of $N(\cdot)$, with zero mean, variance $\mu_2 < \infty$, and fourth moment $\mu_4 < \infty$.

Models related to (1) have been considered in the economic literature. Clark (1973) wrote the model

$$\log P(t) = \log P(0) + B(\tilde{N}(t)),$$

where B is Brownian motion, and $\tilde{N}(t)$ is a nondecreasing positive stochastic process with independent increments, independent of B. Our model generalizes that of Clark (1973) in that it allows for non-independent increments of $N(\cdot)$ and non-Gaussian price changes, however in our model $N(\cdot)$ is restricted to be a pure-jump process since it is derived from the underlying duration process. Press (1967) considered a model of form (1), but included an additional continuous component, assumed normality of price changes, and assumed that $N(\cdot)$ is a Poisson process. Oomen (2006) has generalized the pure-jump version of the model of Press (1967) to allow for time-varying intensity of the Poisson process, and to allow for non-independent price changes so as to describe market microstructure effects (see below).

The existing literature on realized volatility generally assumes that the logarithmic price process is a diffusion given by

$$\log P(t) = \log P(0) + \int_0^t \sigma(u) dW(u)$$

where W is a Brownian motion and σ, the instantaneous volatility, is a positive càdlàg process. Though the diffusive and pure-jump frameworks appear to be very different, Oomen (2006) points out that they may have similar implications for realized volatility. Nevertheless, the diffusive models by themselves do not yield a mechanism for generating transaction times, or therefore, durations and counts. We choose to use a pure-jump framework here since our goal in this paper is to link the properties of the observable durations, counts and realized volatility.

In (1) and henceforth, we adopt the convention that a sum is taken to be zero if the upper limit is less than the lower limit. Since the price changes are independent of each other and of the process $N(\cdot)$, the log prices under (1) are a Martingale.

The model (1) implies that the returns r_s at equally-spaced clock-time intervals of width $\Delta t > 0$ may be expressed as

$$r_s = \sum_{j=N[(s-1)\Delta t]+1}^{N(s\Delta t)} \xi_j \quad , \quad s = 1, 2, \cdots \quad . \tag{2}$$

Recall that the counts $\{\Delta N_s\}$ are given by $\Delta N_s = N(s\Delta t) - N[(s-1)\Delta t]$, that is, the number of events occurring within the given equally-spaced intervals of clock time. The following theorem shows that under the model (1) the memory parameter of the durations propagates unchanged to the realized volatility, $\sum_{s=1}^{n} r_s^2$.

Theorem 5 *For durations $\{\tau_k\}$ satisfying the assumptions of Theorem 1 with $d \in [0, 1/2)$, we have*

$$Var(\sum_{s=1}^{n} r_s^2) \sim Cn^{2d+1}$$

as $n \to \infty$, where $\{r_s\}$ are the returns given by (2).

Remark 5: The conclusions of Theorem 5 would continue to hold if instead of making assumptions on the durations, we assume that $\mathrm{Var}\, N(t) \sim Ct^{2d+1}$.

Remark 6: It follows from Theorem 5 that if $N(\cdot)$ is a Poisson process, the realized volatility has a variance that is asymptotically proportional to n, i.e., short memory.

Remark 7: The tick-time return process $\{\xi_j\}$ in Thereom 5 could be extended to a sum $\{\xi_j + \eta_j\}$ where $\{\eta_j\}$ are *iid*, and independent of $\{\xi_j\}$.

The model (1) implies that clock-time returns are uncorrelated, whereas there is considerable empirical evidence to the contrary (see, e.g., Roll 1984). This autocorrelation is often attributed to microstructure effects such as bid-ask bounce. We therefore consider a generalization that allows returns in tick time to be autocorrelated, which in turn allows for autocorrelation in the clock-time returns. The generalization we consider is in keeping with the literature for both pure-jump models (Oomen 2006) and diffusion models (Bandi and Russell 2004, Hansen and Lunde 2006, Zhang, Mykland and Aït-Sahalia 2005). The generalization of model (1) takes the form

$$\log P(t) = \log P(0) + \sum_{j=1}^{N(t)} (\xi_j + \eta_j), \tag{3}$$

where $\{\eta_j\}$ is a stationary, zero-mean q-dependent process with finite fourth moment, independent of the counting process $N(\cdot)$. We assume also that the process $\{\xi_j + \eta_j\}$ is q-dependent. We can view the process $\{\eta_j\}$ as representing microstructure noise. Note that we do not require the $\{\eta_j\}$ to be independent of the efficient price shocks $\{\xi_j\}$. Model (3) covers the case considered by Oomen (2006), who assumed that the $\{\eta_j\}$ are a Gaussian q'th-order moving average with respect to the difference of an *iid* process.

We have so far been unable to derive a result corresponding to Theorem 5 for the variance of the realized volatility based on the clock-time returns generated by the price process in the presence of microstructure noise given by (3). Obtaining such a result would require knowledge of the asymptotic behavior of the variance of the number of zero counts in a sequence of n counts, which is not known as far as we are aware. While this is unfortunate, there are alternative sampling schemes for the construction of realized volatility that have been proposed in the literature, which have desirable properties. We consider the transaction time sampling scheme, TTS, proposed by Oomen (2006), under which the returns are measured every K (tick-time) transactions, and then the squares of these returns are aggregated to form a realized volatility. This contrasts with calendar time sampling, CTS (i.e., sampling in fixed intervals of clock time), which results in the realized volatility studied in Theorem 5. Oomen (2006) found that transaction time sampling leads to realized volatility that has superior performance relative to that based on calendar time sampling.

Define $z_j = \xi_j + \eta_j$ for $j = 1, 2, \ldots$. Sampling every K transactions yields the sampled returns $\tilde{r}_j = z_{(j-1)K+1} + \ldots + z_{jK}$. Since $\{z_j\}$ is q-dependent, $\{\tilde{r}_j\}$ is also q-dependent. Aggregating the available squared sampled returns up to time T yields the TTS realized volatility

$$\widetilde{RV}_T = \sum_{j=1}^{\lfloor N(T)/K \rfloor} \tilde{r}_j^2. \tag{4}$$

The following theorem shows that under the model (3) the memory parameter of the durations propagates unchanged to the TTS realized volatility, $\widetilde{RV}_T$.

Theorem 6 *Let the durations $\{\tau_k\}$ satisfy the assumptions of Theorem 1 with $d \in [0, 1/2)$. Furthermore, when $d = 0$, assume that the spectral density of $\{\tilde{r}_j^2\}$ is positive at zero frequency. Then*

$$\mathrm{var}(\widetilde{RV}_T) \sim CT^{2d+1}$$

as $T \to \infty$ for some $C > 0$, where $\widetilde{RV}_T$ is defined in (4).

Remark 8: Gaussianity of $\{\tilde{r}_j\}$ is a sufficient condition for the positivity of the spectral density of $\{\tilde{r}_j^2\}$ at zero frequency. This follows since if $\{\tilde{r}_j\}$ is Gaussian, $cov(\tilde{r}_j, \tilde{r}_k) \geq 0$ for all j, k by the formula of Isserlis (1918), and since the spectral density at zero is the variance plus twice the sum of the autocovariances at nonzero lags.

V Discussion

In their empirical study, Deo, Hsieh and Hurvich (2007) found that the high frequency data on durations, counts and realized volatility possesses long memory and the memory parameter is apparently identical across the three series. In this paper, we have established conditions on the durations that guarantee the propagation of the memory parameter, without change, from the durations to the counting process and then to the realized volatility resulting from the return models considered here. Our theoretical results imply that short-memory models, such as the finite-variance ACD model, and the SCD model of Bauwens and Veredas (2004), cannot generate long memory in the resulting counts or realized volatility.

Two duration models which can yield long memory in counts are the LMSD model (see Theorem 2), and ACD models with infinite variance (see Daley, Rolski and Vesilo, 2000). Although these two models could generate identical second-order dependence properties in counts, the realizations in general would look very different. Under the infinite-variance ACD model, realizations of the counting process would tend to be dominated by a few very long durations, and hence the resulting series of counts would tend to have long strings of zeros. On the other hand, this behavior would not be displayed by the LMSD model. Nevertheless, there is no known statistical procedure for distinguishing between finite and infinite variance of durations based on runs of zeros in the counts. One could consider estimating the tail index directly from durations. Although there exist semiparametric tail index estimators, they are known to be badly behaved even when the series is independent (Resnick 1997), to say nothing of strong dependence, as in the case of transaction durations.

We also presented a result implying that the only way for a series of counts aggregated over a long time period to have nontrivial autocorrelation is for the short-term counts to have long memory. In other words, aggregation ultimately destroys all autocorrelation in counts, if and only if the counts have short memory. Deo, Hsieh and Hurvich (2007) found that the empirical lag-1 autocorrelations of the aggregated counts did not decay to zero with increasing aggregation. They also found, in accordance with Theorem 4, that the lag-1 autocorrelations of counts generated from simulated durations generated from the empirically estimated LMSD models also did not decay to zero whereas those from the estimated exponential and Weibull ACD models (whose estimated parameters implied finite variance though this restriction was not imposed in the estimation) did decay to zero, in contradiction to what was observed in the data. This lends further support to the conclusion that LMSD models are more appropriate for intertrade durations than finite-variance ACD models.

References

[1] Andersen, T.G., and Bollerslev, T. (1997), Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns, *Journal of Finance* **LII**, 975–1005.

[2] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001), The Distribution of Exchange Rate Volatility, *Journal of the American Statistical Association* **96**, 42–55.

[3] Andrasfai, B. (1977), *Introductory Graph Theory.* Bristol, UK: Adam Hilger.

[4] Baccelli, F. and Brémaud, P. (2003), *Elements of Queueing Theory.* New York: Springer.

[5] Bandi, F.M. and Russell, J.R. (2004), Separating Microstructure Noise from Volatility, *Journal of Financial Economics*, forthcoming.

[6] Barndorff-Nielsen, O. and Shephard, N. (2006), Impact of Jumps on Returns and Realised Variances: Econometric Analysis of Time-Deformed Lévy Processes, *Journal of Econometrics* **131**, 217–252.

[7] Bauwens, L. and Veredas, D. (2004), The Stochastic Conditional Duration Model: a Latent Variable Model for the Analysis of Financial Durations, *Journal of Econometrics* **119**, 381–412.

[8] Baillie, R. Bollerslev, T. and Mikkelsen, H. (1996), Fractionally integrated generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* **74,** 3–30.

[9] Billingsley, P. (1986), *Probability and Measure*, 2'nd Ed. New York: Wiley.

[10] Bollerslev, T. and Mikkelsen, H.O. (1996), Modeling and Pricing Long Memory in Stock Market Volatility, *Journal of Econometrics* **73**, 151–184.

[11] Bollerslev, T. and Jubinski, D. (1999), Equity Trading Volume and Volatility: Latent Information Arrivals and Common Long-Run Dependencies, *Journal of Business and Economic Statistics* **17**, 9–21.

[12] Bradley, R. (2005), Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions, *Probability Surveys* **2**, 107–144.

[13] Breidt, F. J., Crato, N., and de Lima, P. (1998), The detection and estimation of long memory in stochastic volatility, *Journal of Econometrics* **83,** 325–348.

[14] Brillinger, D.R. (1981), *Time Series. Data Analysis and Theory*, 2'nd Ed. Oakland, CA: Holden-Day.

[15] Carrasco, M. and Chen X. (2002), Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models, *Econometric Theory* **18**, 17–39.

[16] Chung, K.L. (1974), *A Course In Probability Theory*, 2'nd Ed. New York: Academic Press.

[17] Clark, P.K., 1976, A subordinated Stochastic Process Model with finite variance for speculative prices, *Econometrica* 41, 135–155.

[18] Comte, F. and Renault, E. (1998), Long Memory in Continuous-Time Stochastic Volatility Models, *Mathematical Finance* **8**, 291–323.

[19] Daley, D.J., Rolski, T. and Vesilo, R. (2000), Long-Range Dependent Point Processes and their Palm-Khinchin Distributions, *Advances in Applied Probability* **32**, 1051–1063.

[20] Daley, D.J. and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes*, 2'nd Ed. New York: Springer.

[21] Deo, R.S. and Hurvich, C.M. (2001), On the Log Periodogram Regression Estimator of the Memory Parameter in Long Memory Stochastic Volatility Models, *Econometric Theory* **17**, 686–710.

[22] Deo, R. Hsieh, M. and Hurvich, C.M. (2007), Long Memory in Intertrade Durations, Counts and Realized Volatility of NYSE Stocks, *Preprint*.

[23] Deo, R., Hurvich, C. and Lu, Y. (2006), Forecasting Realized Volatility Using a Long Memory Stochastic Volatility Model: Estimation, Prediction and Seasonal Adjustment, *Journal of Econometrics* **131**, 29–58.

[24] Dittmann, I., and Granger, C.W.J. (2002), Properties of nonlinear transformations of fractionally integrated processes, *Journal of Econometrics* **110**, 113–133.

[25] Doukhan, P. (1994), *Mixing*, Lecture Notes in Statistics **85**, New York: Springer-Verlag.

[26] Engle, R. and Russell J. (1998), Autoregressive Conditional Duration: a New Model for Irregularly Spaced Transaction Data, *Econometrica* **66**, 1127–1162.

[27] Epps, T.W. and Epps, M.L. (1976), The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distribution Hypothesis, *Econometrica* **44**, 305–321.

[28] Hansen, P.R. and Lunde, A. (2006), Realized Variance and Market Microstructure Noise, *Journal of Business and Economic Statistics* **24**, 127–161.

[29] Harvey, A. C. (1998). Long memory in stochastic volatility. In: Knight, J., Satchell, S. (Eds.), *Forecasting volatility in financial markets*. Butterworth-Heinemann, London.

[30] Hurvich, C.M., Moulines, E. and Soulier, P. (2005), Estimating Long Memory in Volatility, *Econometrica* **73**, 1283–1328.

[31] Ibragimov, I.A. and Linnik, Yu.V. (1971), *Independent and Stationary Sequences of Random Variables*, Ed. J.F.C. Kingman, Groningen: Wolters-Noordhoff Publishing.

[32] Iglehart, D.L. and Whitt, W. (1971), The Equivalence of Functional Central Limit Theorems for Counting Processes and Associated Partial Sums, *Annals of Mathematical Statistics* **42**, 1372–1378.

[33] Isserlis, L. (1918), On a Formula for the Product-Moment Coefficient of any Order of a Normal Frequency Distribution in Any Number of Variables, *Biometrika* **12**, 134–139.

[34] Nelson, D. (1990), Stationarity and Persistence in the GARCH(1,1) Model, *Econometric Theory* **6**, 318–334.

[35] Nieuwenhuis, G. (1989), Equivalence of Functional Limit Theorems for Stationary Point Processes and Their Palm Distributions, *Probability Theory and Related Fields* **81**, 593–608.

[36] Oomen, R.C.A. (2006), Properties of Realized Variance under Alternative Sampling Schemes, *Journal of Business and Economic Statistics* **24**, 219 – 237.

[37] Press, S.J. (1967), A Compound Events Model for Security Prices, *Journal of Business* **40**, 317-335.

[38] Resnick, S.I. (1997), Heavy tail modeling and teletraffic data. With discussion and a rejoinder by the author, *Ann. Statist.* **25**, 1805–1869.

[39] Robinson, P.M. (1991), Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression, *Journal of Econometrics* **47**, 67–84.

[40] Robinson, P.M. and Henry, M. (1999), Long and Short Memory Conditional Heteroskedasticity in Estimating the Memory Parameter of Levels, *Econometric Theory* **15**, 299–336.

[41] Roll, R. (1984), A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market, *Journal of Finance* **39**, 1127–1139.

[42] Rydberg, T.H. and Shephard, N. (2003), Dynamics of Trade-by-Trade Price Movements: Decomposition and Models, *Journal of Financial Econometrics* **1**, 2–25.

[43] Surgailis, D. and Viano, M-C. (2002), Long Memory Properties and Covariance Structure of the EGARCH Model, *European Series in Applied and Industrial Mathematics: Probability and Statistics* **6**, 311–329.

[44] Tauchen, G. and Pitts, M. (1983), The Price Variability-Volume Relationship on Speculative Markets, *Econometrica* **51**, 485–505.

[45] Tsay, R. (2002), *Analysis of Financial Time Series*, New York: John Wiley & Sons.

[46] Viano, M.-C., Deniau, Cl. Oppenheim, G. (1995), Long-range dependence and mixing for discrete time fractional processes, *Journal of Time Series Analysis* **16**, 323–338.

[47] Yokoyama, R. (1980), Moment Bounds for Stationary Mixing Sequences, *Z. Wahrsch. Verw. Gebiete* **52**, 45–57.

[48] Zhang, L., Mykland, P.A., and Aït-Sahalia, Y. (2005), A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data, *Journal of the American Statistical Association* **100**, 1394–1411.

VI Appendix: Proofs

Let P denote the stationary distribution of the point process N on the real line, and let P^0 denote the corresponding Palm distribution. P determines and is completely determined by the stationary distribution P^0 of the doubly infinite sequence $\{\tau_k\}_{k=-\infty}^{\infty}$ of durations. Note that the counting process N is stationary under P, the durations are stationary under P^0, but in general there is no single distribution under which both the counting process and the durations are stationary. For more details on the correspondence between P and P^0, see Daley and Vere-Jones (2003), Baccelli and Brémaud (2003), or Nieuwenhuis (1989).

Following the standard notation for point processes on the real line (see, e.g., Nieuwenhuis 1989, p. 594), we assume that the event times $\{t_k\}_{k=-\infty}^{\infty}$ satisfy

$$\ldots < t_{-1} < t_0 \leq 0 < t_1 < t_2 < \ldots .$$

Let

$$u_k = \begin{cases} t_1 & if\ k = 1 \\ \tau_k & if\ k \geq 2 \end{cases} .$$

Here, the random variable $t_1 > 0$ is the time of occurrence of the first event following $t = 0$. For $t > 0$, define the count on the interval $(0, t]$, $N(t) := N(0, t]$, by

$$\begin{aligned} N(t) &= \max\{s : \sum_{i=1}^{s} u_i \leq t\}, \quad u_1 \leq t \\ &= 0, \quad\quad\quad\quad\quad\quad\quad\quad u_1 > t. \end{aligned}$$

Throughout the paper, the symbol $\Longrightarrow$ denotes weak convergence in the space $D[0, 1]$.

Proof of Theorem 1:

The proof proceeds by establishing the following facts.

A) Y_n converges weakly under P to Brownian motion ($d = 0$) or fractional Brownian motion ($d > 0$).

B) The standardized counting process converges weakly under P to a multiple of the same limit obtained in A).

The theorem then follows by applying the uniform integrability of the standardized counting process established in Lemma 2 below, in conjunction with Theorem 25.12 of Billingsley (1986, p. 348).

We start by proving A). By assumption *i*), $Y_n \Longrightarrow \sigma B_{1/2+d}$ under P^0, where $\sigma > 0$. First, we will apply Theorem 6.3 of Nieuwenhuis (1989) to the durations $\{\tau_k\}_{k=-\infty}^{\infty}$ to conclude that $Y_n \Longrightarrow \sigma B_{1/2+d}$ under P. Since the $\{\tau_k\}_{k=-\infty}^{\infty}$ are stationary under P^0 and are generated by the shift to the first event following time zero (see Nieuwenhuis 1989, p. 600), and since it follows from *ii*) and Remark 4 that P^0 is $\{\tau_k\}$-mixing, his Theorem 6.3 applies. It follows that $Y_n \Longrightarrow \sigma B_{1/2+d}$ under P.

We next establish B). Define

$$\tilde{Y}_n(s) = \frac{\sum_{k=1}^{\lfloor ns \rfloor}(u_k - \mu)}{n^{1/2+d}} \quad , \quad s \in [0,1].$$

Note that for all s, $\tilde{Y}_n(s) = Y_n(s) + n^{-(1/2+d)}(u_1 - \tau_1)$. From Baccelli and Brémaud (2003, Equation 1.4.2, page 33), for any measurable function h,

$$E[h(\tau_1)] = \lambda E^0[\tau_1 h(\tau_1)] \quad . \tag{5}$$

Since $u_1 \leq \tau_1$, and since assumption *iii*) implies that τ_1 has finite variance under P^0, using $h(x) = x$ in (5), it follows that $n^{-(1/2+d)}(u_1 - \tau_1)$ is $o_p(1)$ under P. Thus, $\tilde{Y}_n \Longrightarrow \sigma B_{1/2+d}$ under P.

Let

$$Z(t) = \frac{N(t) - t/\mu}{t^{1/2+d}} \quad . \tag{6}$$

By Iglehart and Whitt (1971, Theorem 1), it follows that $Z(t) \xrightarrow{\mathcal{L}} \tilde{C} B_{1/2+d}(1)$ under P as $t \to \infty$, where $\tilde{C} > 0$.

Finally, by Lemma 2, $Z^2(t)$ is uniformly integrable under P and hence by Theorem 25.12 of Billingsley (1986, p. 348), $\lim_t \mathrm{Var}[Z(t)] = \tilde{C}^2 \mathrm{Var}[B_{1/2+d}(1)]$. The theorem is proved. $\square$

Proof of Theorem 2:

We simply verify that the conditions of Theorem 1 hold for this process.

Assume first that $d > 0$. By definition $\{\tau_k\}$ is stationary under P^0 and by Lemma 3, P^0 is $\{\tau_k\}$ mixing. By Surgailis and Viano (2002), $Y_n \Longrightarrow \sigma B_{1/2+d}$ under P^0, where $\sigma > 0$ and by Proposition 1, $\sup_n E^0 \left| \frac{\sum_{k=1}^{n}(\tau_k - \mu)}{n^{1/2+d}} \right|^p < \infty$ for all p. Thus, the result is proved for $d > 0$. If $d = 0$, the proof follows along the same lines as the proof of Theorem 3, since in this case under our assumptions on the model for $d = 0$ the process is exponential α-mixing (see, e.g., Doukhan 1994, Corollary 1, Section 2.1, p. 58), and the weak convergence of Y_n follows again from Surgailis and Viano (2002) with $d = 0$. □

Proof of Theorem 3:

We simply verify that the conditions of Theorem 1 hold for this process.

By Lemma 3, $\{\tau_k\}$ is exponential α-mixing, and hence strong mixing and thus by Nieuwenhuis (1989), P^0 is $\{\tau_k\}$-mixing. Furthermore, since all moments of τ_k exist up to order $3 + \epsilon, \epsilon > 0$, we can apply results from Doukhan (1994, Theorem 1, Section 1.5, p. 46) to obtain

$$Y_n \Rightarrow CB, \tag{7}$$

if $\frac{1}{n}\mathrm{Var}(\sum_{k=1}^{n} \tau_k) \to C^2 > 0$, as $n \to \infty$.

It is well known that the GARCH(1,1) model can be represented as an ARMA(1,1) model, see Tsay (2002). Similarly, the ACD(1,1) model can also be re-formulated as an ARMA(1,1) model,

$$\tau_k = \omega + (\alpha + \beta)\tau_{k-1} + (\eta_k - \beta\eta_{k-1}) \tag{8}$$

where $\eta_k = \tau_k - \psi_k$ is white noise with finite variance since $E(\tau_k^{3+\epsilon}) < \infty$. The autoregressive and moving average parameters of the resulting ARMA(1,1) model are $(\alpha + \beta)$ and β, respectively.

It is also known that for any stationary invertible ARMA model $\{z_k\}$, $n\mathrm{Var}(\bar{z}) \to 2\pi f_z(0)$, where $f_z(0)$ is the spectral density of $\{z_k\}$ at zero frequency. For an ARMA(1,1) process, $f_z(0) > 0$ if the moving average coefficient is less than 1. Here, since $0 \leq \beta < 1$, we obtain $\frac{1}{n}\mathrm{Var}(\sum_{k=1}^{n} \tau_k) = n\mathrm{Var}(\bar{\tau}) \to 2\pi f_\tau(0) > 0$, as $n \to \infty$. Therefore (7) follows.

Define $y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \tau_k$. Since all moments of τ_k are bounded up to order $3 + \epsilon$, $(\epsilon > 0)$ under P^0, by Yokoyama (1980), we obtain

$$E^0\{|y_n - E^0(y_n)|^{3+\epsilon}\} \le K < \infty, \qquad \epsilon > 0 \tag{9}$$

uniformly in n, provided that $\{\tau_k\}$ is exponential α-mixing, which is proved in Lemma 3.

Therefore, we can apply Theorem 1 to the ACD(1,1) model and the result follows. □

Proof of Theorem 5: It follows from the proof of Lemma 2 that, for any fixed Δt and integer $m \ge 1$, under assumptions of Theorem 1,

$$E\{(\Delta N_s)^m\} = E\{[N(\Delta t)]^m\} = C < \infty \qquad . \tag{10}$$

Denote $RV = \sum_{s=1}^n r_s^2$. By the law of total variance,

$$\text{var}(RV) = E\{\text{var}[RV|N(\cdot)]\} + \text{var}\{E[RV|N(\cdot)]\}.$$

First, consider $\text{Var}\{E[RV|N(\cdot)]\}$. Since $\{\xi_i\}$ is *i.i.d.*, we have

$$E[r_s^2|N(\cdot)] = \mu_2 \Delta N_s$$

where $\mu_2 = E(\xi_i^2)$. Thus, by Theorem 1,

$$\text{var}\{E[RV|N(\cdot)]\} = \text{var}\{\sum_{s=1}^n E[r_s^2|N(\cdot)]\} = \text{var}\{\mu_2 N(n\Delta t)\} \sim Cn^{2d+1} \qquad . \tag{11}$$

Next, we consider $E\{\text{var}[RV|N(\cdot)]\}$. Since $E[r_s^4|N(\cdot)] \le \mu_4 \Delta N_s^4$, where $\mu_4 = E(\xi_i^4)$, we obtain

$$\text{var}[r_s^2|N(\cdot)] \quad = \quad E[r_s^4|N(\cdot)] - \{E[r_s^2|N(\cdot)]\}^2 \le \mu_4 \Delta N_s^4 - (\mu_2 \Delta N_s)^2 \qquad . \tag{12}$$

Since $\{\xi_i\}$ is *i.i.d.*, $\{r_s^2\}$ is serially independent, conditional on the counting process, $N(\cdot)$. By (10), (12) and the stationarity of $\{\Delta N_s\}$, we obtain

$$E\{\text{var}[RV|N(\cdot)]\} = E\{\sum_{s=1}^n \text{var}[r_s^2|N(\cdot)]\} = \sum_{s=1}^n E\{\text{var}[r_s^2|N(\cdot)]\} = nK \qquad , \tag{13}$$

where K is a positive constant.

The result follows from (11) and (13). $\square$

Proof of Proposition 1:

We present the proof for the case $0 < d < \frac{1}{2}$. The proof for the case $d = 0$ follows along similar lines. Also, we assume here that p is a positive even integer. The result for all positive odd integers follows by Hölder's inequality.

Let $\tilde{y}_n = y_n - E^0(y_n)$. Since $p \geq 2$ is even and $E^0(\tilde{y}_n)^p$ can be expressed as a linear combination of the products of the joint cumulants of $\tilde{y}_n$ of order $2, \ldots, p$, we have

$$\begin{aligned} 0 \leq E^0|\tilde{y}_n|^p = E^0(\tilde{y}_n^p) &= \sum_{\pi} \Big[c_\pi \prod_{j \in \pi} \text{cum}(\underbrace{\tilde{y}_n, \ldots, \tilde{y}_n}_{j \text{ terms}}) \Big] \\ &\leq \sum_{\pi} \Big[|c_\pi| \prod_{j \in \pi} |\text{cum}(\underbrace{\tilde{y}_n, \ldots, \tilde{y}_n}_{j \text{ terms}})| \Big] \end{aligned}$$

where π ranges over the additive partitions of n and c_π is a finite constant depending on π.

Since the first order cumulant of $\tilde{y}_n$ is zero and for all integers $m \geq 2$, the m-th order cumulant of $\tilde{y}_n$ is equal to that of y_n, it suffices to show that the absolute value of the m-th order cumulant of y_n is bounded uniformly in n under P^0, for all $m \in \{2, \ldots, p\}$.

We first consider the second and the third order cumulants.

For the second order cumulant ($m = 2$),

$$|\text{cum}(y_n, y_n)| = |\text{cum}(\frac{\sum_{k=1}^n \tau_k}{n^{d+\frac{1}{2}}}, \frac{\sum_{s=1}^n \tau_s}{n^{d+\frac{1}{2}}})| \leq \frac{1}{n^{2d+1}} \sum_{k=1}^n \sum_{s=1}^n |\text{cum}(\tau_k, \tau_s)| \quad .$$

To calculate the joint cumulant $\text{cum}(\tau_k, \tau_s)$, we briefly introduce some terminology, mainly from Brillinger (1981): consider a (not necessary rectangular) two-way table of indices,

$$\begin{matrix} (1,1) & \ldots & (1, J_1) \\ \vdots & \ldots & \vdots \\ (I,1) & \ldots & (I, J_I) \end{matrix}$$

and a partition $P_1 \cup P_2 \cup \ldots \cup P_M$ of its entries. We say sets $P_{m'}$, $P_{m''}$ of the partition **hook** if there exist $(i_1, j_1) \in P_{m'}$ and $(i_2, j_2) \in P_{m''}$ such that $i_1 = i_2$, i.e. at least one entry of $P_{m'}$ and one entry of $P_{m''}$ come from the same row in the two-way table. We say that sets $P_{m'}$ and $P_{m''}$ **communicate** if there exists a sequence of sets $P_{m_1} = P_{m'}, P_{m_2}, \ldots, P_{m_N} = P_{m''}$ such that P_{m_n} and $P_{m_{n+1}}$ hook for $n = 1, \ldots, N-1$. So $P_{m'}$ and $P_{m''}$ communicate as long as one can find an ordered sequence of sets such that all neighboring pairs hook, and this sequence links $P_{m'}$ and $P_{m''}$ together. Finally a partition is said to be **indecomposable** if all sets in the partition communicate.

By Brillinger (1981), Theorem 2.3.2, for a two-way array of random variables X_{ij}, $j = 1, \ldots, J_i$, $i = 1, \ldots, I$ (see the corresponding two-way table above), the joint cumulant of the I row products

$$Y_i = \prod_{j=1}^{J_i} X_{ij}, \qquad i = 1, \ldots, I$$

is given by,

$$\mathrm{cum}(Y_1, \ldots, Y_I) = \sum_{\nu} \mathrm{cum}(X_{ij}; ij \in \nu_1) \ldots \mathrm{cum}(X_{ij}; ij \in \nu_w)$$

where the summation is over all indecomposable partition $\nu = \nu_1 \cup \ldots \cup \nu_w$ of the two-way table of indices.

It is more convenient to write the partitions in terms of symbols representing the random variables, instead of the indices themselves. We will always use distinct symbols, so that there is a one-to-one correspondence between the indices and the symbols. Nevertheless, the random variables represented by distinct symbols need not be distinct. For example, e^{h_k} and e^{h_s} are distinct symbols, but if $k = s$, they are not different random variables. Ultimately, the cumulants are computed from the random variables.

To compute $\mathrm{cum}(\tau_k, \tau_s)$, we use the two-way table of indices (left) and the corresponding table of symbols (right),

$$\begin{matrix} (1,1) & (1,2) \\ (2,1) & (2,2) \end{matrix} \qquad , \qquad \begin{matrix} e^{h_k} & \epsilon_k \\ e^{h_s} & \epsilon_s \end{matrix}$$

with $I = 2, J_1 = 2$ and $J_2 = 2$.

From Brillinger (1981), Theorem 2.3.1, all joint cumulants corresponding to partitions with at least one of the symbols representing $\{e^{h_k}\}$ and at least one of the symbols representing $\{\epsilon_k\}$ in the same set, are zero because the corresponding random variable sequences are mutually independent. So for $m = 2$, excluding those with at least one of e^{h_k}, e^{h_s} and at least one of ϵ_k, ϵ_s in the same set, the only possible indecomposable partitions (here, the partition is given in terms of the symbols) are:

$$\begin{gathered}\{e^{h_k}, e^{h_s}\}, \{\epsilon_k, \epsilon_s\} \\ \{e^{h_k}, e^{h_s}\}, \{\epsilon_k\}, \{\epsilon_s\} \\ \{e^{h_k}\}, \{e^{h_s}\}, \{\epsilon_k, \epsilon_s\} \qquad .\end{gathered}$$

Thus, $|\mathrm{cum}(y_n, y_n)| \leq A + B + C$, where,

$$\begin{aligned} A &= \frac{1}{n^{2d+1}} \sum_{k=1}^{n} \sum_{s=1}^{n} |\mathrm{cum}(e^{h_k}, e^{h_s})||\mathrm{cum}(\epsilon_k, \epsilon_s)| \\ B &= \frac{1}{n^{2d+1}} \sum_{k=1}^{n} \sum_{s=1}^{n} |\mathrm{cum}(e^{h_k})\mathrm{cum}(e^{h_s})||\mathrm{cum}(\epsilon_k, \epsilon_s)| \\ C &= \frac{1}{n^{2d+1}} \sum_{k=1}^{n} \sum_{s=1}^{n} |\mathrm{cum}(e^{h_k}, e^{h_s})||\mathrm{cum}(\epsilon_k)||\mathrm{cum}(\epsilon_s)| \end{aligned}$$

Both A and B reduce to a single summation because of the serial independence of the $\{\epsilon_k\}$, so $A = O(n^{-2d})$ and $B = O(n^{-2d})$. For C, by Surgailis and Viano (2002), Corollary 5.3,

$$|\mathrm{cum}(e^{h_k}, e^{h_s})| = e^{\sigma_h^2} |e^{r_{|k-s|}} - 1|$$

where $r_{|k-s|} = \mathrm{cov}(h_k, h_s)$ and $\sigma_h^2 = \mathrm{Var}(h_k)$.

By the assumption on $\{b_j\}$ in the Theorem 2, it follows that $r_s \sim K s^{2d-1}$, as $s \to \infty$, where $K > 0$, so that

$$\begin{aligned} \sum_{k=1}^{n} \sum_{s=1}^{n} |e^{r_{|k-s|}} - 1| &\leq 2 \sum_{k=1}^{n} \sum_{s>k} |e^{r_{|k-s|}} - 1| + n|e^{r_0} - 1| \\ &\leq Kn \sum_{j=1}^{n} j^{2d-1} + n|e^{r_0} - 1| = O(n^{2d+1}). \end{aligned}$$

Thus term C is $O(1)$. Hence, $|\mathrm{cum}(y_n, y_n)|$ is $O(1)$.

Next, for the third order cumulant ($m = 3$), we have

$$|\mathrm{cum}(y_n, y_n, y_n)| = \frac{1}{n^{3d+\frac{3}{2}}} |\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} \mathrm{cum}(\tau_k, \tau_s, \tau_u)| \leq \frac{1}{n^{3d+\frac{3}{2}}} \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} |\mathrm{cum}(e^{h_k}\epsilon_k, e^{h_s}\epsilon_s, e^{h_u}\epsilon_u)| \quad .$$

We will use the following two-way table:

e^{h_k}	ϵ_k
e^{h_s}	ϵ_s
e^{h_u}	ϵ_u

For convenience, we group the indecomposable partitions according to how many sets ($L = 1, 2, 3$) the symbols $e^{h_k}, e^{h_s}, e^{h_u}$ are partitioned into.

We have three groups of indecomposable partitions, excluding those with at least one of $e^{h_k}, e^{h_s}, e^{h_u}$ and at least one of $\epsilon_k, \epsilon_s, \epsilon_u$ in the same set:

i) Group 1

$$\{e^{h_k}, e^{h_s}, e^{h_u}\}, \{\epsilon_k, \epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_s}, e^{h_u}\}, \{\epsilon_k, \epsilon_s\}, \{\epsilon_u\}$$

$$\{e^{h_k}, e^{h_s}, e^{h_u}\}, \{\epsilon_k, \epsilon_u\}, \{\epsilon_s\}$$

$$\{e^{h_k}, e^{h_s}, e^{h_u}\}, \{\epsilon_k\}, \{\epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_s}, e^{h_u}\}, \{\epsilon_k\}, \{\epsilon_s\}, \{\epsilon_u\}$$

ii) Group 2

$$\{e^{h_k}, e^{h_s}\}, \{e^{h_u}\}, \{\epsilon_k, \epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_s}\}, \{e^{h_u}\}, \{\epsilon_k\}, \{\epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_s}\}, \{e^{h_u}\}, \{\epsilon_s\}, \{\epsilon_k, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_u}\}, \{e^{h_s}\}, \{\epsilon_k, \epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_u}\}, \{e^{h_s}\}, \{\epsilon_k\}, \{\epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_u}\}, \{e^{h_s}\}, \{\epsilon_s\}, \{\epsilon_k, \epsilon_u\}$$

$$\{e^{h_u}, e^{h_s}\}, \{e^{h_k}\}, \{\epsilon_k, \epsilon_s, \epsilon_u\}$$

$$\{e^{h_u}, e^{h_s}\}, \{e^{h_k}\}, \{\epsilon_k\}, \{\epsilon_s, \epsilon_u\}$$

$$\{e^{h_u}, e^{h_s}\}, \{e^{h_k}\}, \{\epsilon_s\}, \{\epsilon_k, \epsilon_u\}$$

iii) Group 3

$$\{e^{h_k}\}, \{e^{h_s}\}, \{e^{h_u}\}, \{\epsilon_k, \epsilon_s, \epsilon_u\}.$$

We next study the order of the dominant contribution to $|\text{cum}(y_n, y_n, y_n)|$ corresponding to each group.

In Group 1, the dominant term arises from the last partition since it yields a triple summation,

$$\frac{1}{n^{3d+\frac{3}{2}}} \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} |\text{cum}(e^{h_k}, e^{h_s}, e^{h_u})||\text{cum}(\epsilon_k)||\text{cum}(\epsilon_s)||\text{cum}(\epsilon_u)| = \frac{\mu_\epsilon^3}{n^{3d+\frac{3}{2}}} \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} |\text{cum}(e^{h_k}, e^{h_s}, e^{h_u})|$$

where $\mu_\epsilon = E^0(\epsilon_1)$.

By Surgailis and Viano (2002), Corollary 5.3,

$$\begin{aligned}
& \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} |\text{cum}(e^{h_k}, e^{h_s}, e^{h_u})| \\
\leq \; & \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2} |e^{r_{|k-s|}} - 1||e^{r_{|k-u|}} - 1||e^{r_{|s-u|}} - 1| \\
+ \; & \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2} |e^{r_{|k-s|}} - 1||e^{r_{|k-u|}} - 1| + \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2} |e^{r_{|k-s|}} - 1||e^{r_{|s-u|}} - 1| \\
+ \; & \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2} |e^{r_{|k-u|}} - 1||e^{r_{|s-u|}} - 1|
\end{aligned}$$

The last three summations are actually the same due to symmetry: we can simply relabel the indices in the last summation by $s \leftrightarrow u$. As for the first summation, since $|r_{|k-u|}| = |\text{cov}(h_k, h_u)| \leq \sigma_h^2 = \text{Var}(h_k)$, we have $|e^{r_{|k-u|}}-1| \le (e^{\sigma_h^2}+1) < \infty$. So

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} |\text{cum}(e^{h_k}, e^{h_s}, e^{h_u})| &\le K\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2}|e^{r_{|k-s|}}-1||e^{r_{|s-u|}}-1| \\
&+ 3\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2}|e^{r_{|k-s|}}-1||e^{r_{|s-u|}}-1| \\
&\le K\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} |e^{r_{|k-s|}}-1||e^{r_{|s-u|}}-1| \qquad \text{(for some } K>0) \\
&= O(n^{4d+1})
\end{aligned}$$

The last step follows from Lemma 1. So $\frac{\mu_\epsilon^3}{n^{3d+\frac{3}{2}}}\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n}|\text{cum}(e^{h_k}, e^{h_s}, e^{h_u})|$ converges to zero because $(4d+1) < (3d+\frac{3}{2})$.

Similarly, the dominant contribution from Group 2 is of order

$$\frac{1}{n^{3d+\frac{3}{2}}}\sum_i\sum_j |\text{cum}(e^{h_i}, e^{h_j})||\text{cum}(e^{h_j})|$$

Note that in Group 2, all three of $e^{h_k}, e^{h_s}, e^{h_u}$ are partitioned into two sets. Therefore, partitions with all three of $\epsilon_k, \epsilon_s, \epsilon_u$ in different sets are not indecomposable, so the dominant contribution is a double sum,

$$\frac{1}{n^{3d+\frac{3}{2}}}\sum_i\sum_j |\text{cum}(e^{h_i}, e^{h_j})||\text{cum}(e^{h_j})| = \frac{\mu_{e^h}}{n^{3d+\frac{3}{2}}}\sum_i\sum_j |\text{cum}(e^{h_i}, e^{h_j})| \le Kn^{(2d+1)-(3d+\frac{3}{2})} = O(n^{-d-\frac{1}{2}})$$

where $\mu_{e^h} = E^0(e^{h_1})$.

So the dominant term in Group 2 also converges to zero.

For Group 3, all three of $e^{h_k}, e^{h_s}, e^{h_u}$ are partitioned into three different sets, so that the part of the partition involving $\epsilon_k, \epsilon_s, \epsilon_u$ must be $\{\epsilon_k, \epsilon_s, \epsilon_u\}$ in order to be indecomposable. The resulting summation now is only a single one of order $O(n^1)$. The dominant contribution again converges to zero.

Notice that the order of the dominant contribution from group 3 ($O(n^{-3d-\frac{1}{2}})$) is of smaller order than that from group 2 ($O(n^{-d-\frac{1}{2}})$), which is of smaller order of that from group 1 ($O(n^{d-\frac{1}{2}})$). This will be shown to hold in general for any m-th order joint cumulant.

Next, we prove that the m-th order joint cumulant, which satisfies

$$|\text{cum}(\underbrace{y_n, \dots, y_n}_{m \text{ terms}})| \leq \frac{1}{n^{m(d+\frac{1}{2})}} \sum_{k_1=1}^{n} \cdots \sum_{k_m=1}^{n} |\text{cum}(e^{h_{k_1}}\epsilon_{k_1}, \dots, e^{h_{k_m}}\epsilon_{k_m})| \tag{14}$$

converges to zero for all $m > 2$.

The indecomposable partitions of $(e^{h_{k_1}}\epsilon_{k_1}, \dots, e^{h_{k_m}}\epsilon_{k_m})$ are organized in a similar manner as before into m groups, where in Group L the symbols $e^{h_{k_1}}, \dots, e^{h_{k_m}}$ are divided into L sets $(L = 1, \dots, m)$.

a) First, consider Group 1. The dominant contribution to the righthand side of (14) corresponding to Group 1 must be the one from the partition in which all of the symbols $e^{h_{k_1}}, \dots, e^{h_{k_m}}$ are in one set and each of the symbols $\epsilon_{k_1}, \dots, \epsilon_{k_m}$ is in a set by itself. The resulting summation is an m-fold summation. By Corollary 5.3 of Surgailis and Viano (2002), the absolute value of the m-th joint cumulant, $|\text{cum}(e^{h_{k_1}}, \dots, e^{h_{k_m}})|$, is bounded by a summation taken over all connected graphs with m vertices. Each entry of the summation is a product of terms of the form $|e^{r_{|k_i-k_j|}} - 1|$ along the edges that connect vertices k_i and k_j of a connected m-vertex graph.

For a graph with m vertices, we need at least $(m-1)$ edges to connect them. It is known (see Andrasfai, 1977, Chapter 2) that any connected m-vertex graph with $(m-1)$ edges may be represented as a *tree*. Let $W_{\{k_i,\dots,k_j\}} < \infty$ be the total number of trees with vertices labeled by $k_i, k_{i+1}, \dots, k_j$.

If a connected m-vertex graph used in applying Corollary 5.3 of Surgailis and Viano (2002) has more than $(m-1)$ edges, it is not a tree, and there will be more than $(m-1)$ terms of the form $|e^{r_{|k_i-k_j|}} - 1|$ being multiplied together in the m-fold summation in (14). But, for all k_i, k_j, $|r_{|k_i-k_j|}| = |\text{cov}(h_{k_i}, h_{k_j})| \leq \sigma_h^2 = \text{Var}(h_{k_i})$, so $|e^{r_{|k_i-k_j|}} - 1| \leq (e^{\sigma_h^2} + 1) < \infty$, and for any connected m-vertex graph with more than $(m-1)$ edges, there exists an m-vertex subgraph that has a tree representation. So we can retain a product of $(m-1)$ terms of the form $|e^{r_{|k_i-k_j|}} - 1|$ in the m-fold summation in (14) and move remaining terms out of the summation, bounding each by $(e^{\sigma_h^2} + 1)$. The resulting product of $(m-1)$ terms of the form $|e^{r_{|k_i-k_j|}} - 1|$ is itself a product over the edges of an m-vertex tree.

In all, $|\text{cum}(e^{h_{k_1}}, \dots, e^{h_{k_m}})|$ is bounded by a constant times a summation over the set $G_{\{k_1,\dots,k_m\}}$ of all $W_{\{k_1,\ldots,k_m\}}$ trees. Each entry of the summation is a product of terms of the form $|e^{r|k_i-k_j|}-1|$ being multiplied over the $(m-1)$ edges of the tree. Thus, we have

$$\begin{aligned}\sum_{k_1=1}^{n}\cdots\sum_{k_m=1}^{n}|\mathrm{cum}(e^{h_{k_1}},\ldots,e^{h_{k_m}})| &\le K\sum_{k_1=1}^{n}\cdots\sum_{k_m=1}^{n}\Big\{\sum_{G_{\{k_1,\ldots,k_m\}}}\prod_{(k_i,k_j)\in\Omega(G_{\{k_1,\ldots,k_m\}})}|e^{r|k_i-k_j|}-1|\Big\},\quad (K>0)\\ &= K\sum_{G_{\{k_1,\ldots,k_m\}}}\sum_{k_1=1}^{n}\cdots\sum_{k_m=1}^{n}\Big\{\underbrace{\prod_{(k_i,k_j)\in\Omega(G_{\{k_1,\ldots,k_m\}})}|e^{r|k_i-k_j|}-1|}_{(m-1)\text{ terms}}\Big\}\end{aligned}$$

where $\Omega(G_{\{k_1,\ldots,k_m\}})$ is the set of edges of the graph indexed by $G_{\{k_1,\ldots,k_m\}}$.

By Lemma 1, each entry of the summation over $G_{\{k_1,\ldots,k_m\}}$ is of order $O(n^{2dm-2d+1})$. Also this summation is taken over a finite number of graphs ($W_{\{k_1,\ldots,k_m\}}<\infty$), therefore

$$\sum_{k_1=1}^{n}\cdots\sum_{k_m=1}^{n}|\mathrm{cum}(e^{h_{k_1}},\ldots,e^{h_{k_m}})| = O(n^{2dm-2d+1}).$$

Because the normalization term in (14) is of order $O(n^{m(d+\frac{1}{2})})$, the dominant contribution to $\mathrm{cum}(\underbrace{y_n,\ldots,y_n}_{m\text{ terms}})$ from Group 1 converges to zero, for any $m>2$.

b) For Group 2, the symbols $e^{h_{k_1}},\ldots,e^{h_{k_m}}$ are partitioned into two sets. Thus, the partitions with each of the m symbols $\epsilon_{k_1},\ldots,\epsilon_{k_m}$ in a set by itself are not indecomposable. Relabel the two sets as $\{e^{h_{g_1}},\ldots,e^{h_{g_q}}\}$, $\{e^{h_{g_{q+1}}},\ldots,e^{h_{g_m}}\}$. Since the partition must be indecomposable, there must be one $I\in(1,\ldots,q)$ and one $J\in(q+1,\ldots,m)$, such that $g_I=g_J$. The dominant contribution to (14) from Group 2 is therefore

$$\frac{1}{n^{m(d+\frac{1}{2})}}\sum_{g_1=1}^{n}\cdots\sum_{g_m=1}^{n}|\mathrm{cum}(e^{h_{g_1}},\ldots,e^{h_{g_q}})||\mathrm{cum}(e^{h_{g_{q+1}}},\ldots,e^{h_{g_m}})||\mathrm{cum}(\epsilon_{g_I},\epsilon_{g_J})| \qquad (15)$$

Similarly as above, after applying Corollary 5.3 of Surgailis and Viano (2002) and after bounding certain terms, we obtain

$$
\begin{aligned}
& \sum_{g_1=1}^{n} \cdots \sum_{g_m=1}^{n} |\mathrm{cum}(e^{h_{g_1}}, \ldots, e^{h_{g_q}})||\mathrm{cum}(e^{h_{g_{q+1}}}, \ldots, e^{h_{g_m}})||\mathrm{cum}(\epsilon_{g_I}, \epsilon_{g_J})| \\
\leq \; & K \sum_{g_1=1}^{n} \cdots \sum_{g_m=1}^{n} \Big\{ \sum_{G_{\{g_1,\ldots,g_q\}}} \underbrace{\prod_{(g_i,g_j)\in\Omega(G_{\{g_1,\ldots,g_q\}})} |e^{r|g_i-g_j|} - 1|}_{(q-1) \text{ terms}} \Big\} \\
& \cdot \Big\{ \sum_{G_{\{g_{q+1},\ldots,g_m\}}} \underbrace{\prod_{(g_i,g_j)\in\Omega(G_{\{g_{q+1},\ldots,g_m\}})} |e^{r|g_i-g_j|} - 1|}_{(m-q-1) \text{ terms}} \Big\} \Big\{ |\mathrm{cum}(\epsilon_{g_I}, \epsilon_{g_J})| \Big\} \\
= \; & K \sum_{G_{\{g_1,\ldots,g_q\}}} \sum_{G_{\{g_{q+1},\ldots,g_m\}}} \sum_{g_1=1}^{n} \cdots \sum_{g_m=1}^{n} \mathbf{1}_{\{g_I=g_J\}} \\
& \cdot \Big\{ \underbrace{\prod_{(g_i,g_j)\in\Omega(G_{\{g_1,\ldots,g_q\}})} |e^{r|g_i-g_j|} - 1| \prod_{(g_i,g_j)\in\Omega(G_{\{g_{q+1},\ldots,g_m\}})} |e^{r|g_i-g_j|} - 1|}_{(m-2) \text{ terms, denote as } \Gamma(g_1,\ldots,g_m:G_{\{g_1,\ldots,g_q\}},G_{\{g_{q+1},\ldots,g_m\}})} \Big\}.
\end{aligned}
$$

As mentioned before, any graph G_a in $G_{\{g_1,\ldots,g_q\}}$ and any graph G_b in $G_{\{g_{q+1},\ldots,g_m\}}$, can be represented by trees with q and $(m-q)$ vertices, respectively. Since for any two trees, the resulting structure obtained by merging one vertex from each tree is again a tree, under the constraint $g_I = g_J$, there exists a graph G_c in $G_{\{g_1,\ldots,g_{I-1},g_{I+1},\ldots,g_m\}}$, such that G_c is obtained by merging G_a and G_b together at the vertex $g_I = g_J$.

Therefore, the numerical value of the term Γ evaluated for graphs G_a and G_b and indices $\{g_1, \ldots, g_m\}$ with the constraint $g_I = g_J$ (which follows from the independence of the $\{\epsilon_{g_i}\}$) is equal to the value of the term Φ (defined below) evaluated using the graph G_c in $G_{\{g_1,\ldots,g_{I-1},g_{I+1},\ldots,g_m\}}$ and indices $\{g_1, \ldots, g_{I-1}, g_{I+1}, \ldots, g_m\}$ without any constraint on the values of these indices. After re-parameterizing $\{g_1, \ldots, g_{I-1}, g_{I+1}, \ldots, g_m\}$ by $\{l_1, \ldots, l_{m-1}\}$, we obtain

$$
\begin{aligned}
& \sum_{g_1=1}^{n} \cdots \sum_{g_m=1}^{n} |\mathrm{cum}(e^{h_{g_1}}, \ldots, e^{h_{g_q}})||\mathrm{cum}(e^{h_{g_{q+1}}}, \ldots, e^{h_{g_m}})||\mathrm{cum}(\epsilon_{g_I}, \epsilon_{g_J})| \\
\leq \; & K \sum_{G_{\{l_1,\ldots,l_{m-1}\}}} \sum_{l_1=1}^{n} \cdots \sum_{l_{m-1}=1}^{n} \underbrace{\prod_{(l_i,l_j)\in\Omega(G_{\{l_1,\ldots,l_{m-1}\}})} |e^{r|l_i-l_j|} - 1|}_{(m-2) \text{ terms, denote as } \Phi(l_1,\ldots,l_{m-1}:G_{\{l_1,\ldots,l_{m-1}\}})} \\
= \; & O(n^{2d(m-2)+1})
\end{aligned}
$$

where the final equality follows from Lemma 1.

The above $(m-1)$-fold summation for Group 2 is of smaller order than the m-fold summation from Group 1, which was $O(n^{2d(m-1)+1})$. Hence, the dominant contribution from Group 2 also converges to zero.

c) In general, for Group $L \in \{1,\ldots,m\}$, the symbols $e^{h_{k_1}},\ldots,e^{h_{k_m}}$ are partitioned into L sets. Relabel the L sets as $\{e^{h_{g_1}},\ldots,e^{h_{g_{q_1}}}\}$, $\{e^{h_{g_{q_1+1}}},\ldots,e^{h_{g_{q_2}}}\}$, $\ldots$, $\{e^{h_{g_{q_{L-1}+1}}},\ldots,e^{h_{g_m}}\}$. Since the partition must be indecomposable, there must be L indices $\{I,J,\ldots,Z\}$, where $I \in (1,\ldots,q_1)$, $J \in (q_1+1,\ldots,q_2),\ldots,Z \in (q_{L-1}+1,\ldots,m)$, such that $\underbrace{g_I = g_J = \ldots = g_Z}_{L \text{ terms}}$. The dominant contribution to (14) from Group L is then,

$$\frac{1}{n^{m(d+\frac{1}{2})}}\sum_{g_1=1}^{n}\cdots\sum_{g_m=1}^{n}\underbrace{|\text{cum}(e^{h_{g_1}},\ldots,e^{h_{g_{q_1}}})|\ldots|\text{cum}(e^{h_{g_{q_{L-1}+1}}},\ldots,e^{h_{g_m}})|}_{L-\text{terms}}|\text{cum}(\underbrace{\epsilon_{g_I},\epsilon_{g_J},\ldots,\epsilon_{g_Z}}_{L \text{ terms}})|. \quad (16)$$

Similarly as before, we obtain

$$\begin{aligned}
&\sum_{g_1=1}^{n}\cdots\sum_{g_m=1}^{n}\underbrace{|\text{cum}(e^{h_{g_1}},\ldots,e^{h_{g_{q_1}}})|\ldots|\text{cum}(e^{h_{g_{q_{L-1}+1}}},\ldots,e^{h_{g_m}})|}_{L-\text{terms}}|\text{cum}(\underbrace{\epsilon_{g_I},\epsilon_{g_J},\ldots,\epsilon_{g_Z}}_{L \text{ terms}})| \\
\leq\ & K\underbrace{\sum_{G_{\{g_1,\ldots,g_{q_1}\}}}\cdots\sum_{G_{\{g_{q_{L-1}+1},\ldots,g_m\}}}}_{L-\text{fold}}\sum_{g_1=1}^{n}\cdots\sum_{g_m=1}^{n}\mathbf{1}_{\{\underbrace{g_I=g_J=\ldots=g_Z}_{L \text{ terms}}\}} \\
&\cdot\underbrace{\Big\{\prod_{(g_i,g_j)\in\Omega(G_{\{g_1,\ldots,g_{q_1}\}})}|e^{r|g_i-g_j|}-1|\quad\ldots\prod_{(g_i,g_j)\in\Omega(G_{\{g_{q_{L-1}+1},\ldots,g_m\}})}|e^{r|g_i-g_j|}-1|\Big\}}_{(m-L) \text{ terms}} \\
\leq\ & K\sum_{G_{\{l_1,\ldots,l_{m-L+1}\}}}\sum_{l_1=1}^{n}\cdots\sum_{l_{m-L+1}=1}^{n}\underbrace{\prod_{(l_i,l_j)\in\Omega(G_{\{l_1,\ldots,l_{m-L+1}\}})}|e^{r|l_i-l_j|}-1|}_{(m-L) \text{ terms}} \\
=\ & O(n^{2d(m-L)+1}),
\end{aligned}$$

by Lemma 1.

The constraint $\underbrace{g_I = g_J = \ldots = g_Z}_{L \text{ terms}}$ allows the re-parameterization from $\{g_1,\ldots,g_m\}$ to $\{l_1,\ldots,l_{m-L+1}\}$ and reduces the m-fold summation in (16) to an $(m-L+1)$-fold summation in the last inequality. It was shown for Group 2 that the graph obtained by merging one vertex from each of any pair of trees is again a tree. By induction, we obtain a tree by merging one vertex from each of $L > 2$ trees, which allows us to apply Lemma 1 with $M = m - L + 1$ in the last step.

So, the dominant contribution from Group L is $O(n^{2d(m-L)+1-m(d+\frac{1}{2})})$, $(L = 1, \ldots, m)$. Since $d > 0$, the dominant contribution from *all* groups occurs for $L = 1$. Finally, the dominant contribution from Group 1 is $O(n^{2d(m-1)+1-m(d+\frac{1}{2})})$, which tends to zero for $m > 2$ since $d < \frac{1}{2}$. $\square$

Lemma 1 *For any $M > 2$ and $0 < d < \frac{1}{2}$,*

$$\underbrace{\sum_{k_1=1}^{n} \cdots \sum_{k_M=1}^{n}}_{M-fold} \underbrace{\left\{ \prod_{(k_i,k_j)\in\Omega(G)} |e^{r_{|k_i-k_j|}} - 1| \right\}}_{(M-1)\ terms} = O(n^{2d(M-1)+1}) \tag{17}$$

where $\Omega(G)$ is the set of edges of G, G is any connected M-vertex graph with vertices $\{k_1, \ldots, k_M\}$ and $(M-1)$ edges; $r_{|k_i-k_j|} = cov(h_{k_i}, h_{k_j}), 1 \leq i \leq M, 1 \leq j \leq M$, $\{h_{k_i}\}$ is a long memory process with memory parameter d.

Proof: Since G is a connected graph with M vertices and $(M-1)$ edges, it can be represented as a tree (see Andrasfai 1977, Chapter 2). The tree representation is not unique. Fix a particular representation. Then there is one vertex with no parent, called the root. A vertex with both a parent and a child is called a node. A vertex with no child is called a leaf.

We proceed iteratively. First, select any leaf vertex. By definition of a leaf, the corresponding index only appears once in the product, so the sum on this index can be evaluated for this term only, holding the other terms fixed. Since $r_s \sim Cs^{2d-1}$ as $s \to \infty$, we have for any fixed integer i with $1 \leq i \leq n$, $\sum_{j=1}^{n} |e^{r_{|i-j|}} - 1| = O(n^{2d})$.

It follows that the sum on the first index is $O(n^{2d})$. Next, delete the leaf just used from the tree. The resulting graph is again a tree. Repeat the process of selecting a leaf, performing the corresponding sum and deleting the leaf until only the root remains. The M-fold sum in (17) is now bounded by a constant times the sum of n terms each of which is $O(n^{2d(M-1)})$. Thus, the sum in (17) is $O(n^{2d(M-1)+1})$. $\square$

Lemma 2 *For durations $\{\tau_k\}$ satisfying the assumptions of Theorem 1, there exists $\delta > 0$ such that*

$$\sup_{t \geq 1} E[|Z(t)|^{2+\delta}] < \infty,$$

where $Z(t)$ is defined in Equation (6).

Proof: Select any $\delta \in (0, \epsilon)$. Let $\theta = 1/(2+\delta)$ and assume without loss of generality that $\mu = 1$, and hence that $\lambda = \frac{1}{\mu} = 1$. By Chung (1974, Theorem 3.2.1, page 42), $E[|Z(t)|^{2+\delta}] \leq 1 + \sum_{s=1}^{\infty} P[|Z(t)|^{2+\delta} \geq s]$. Thus, it suffices to show that

$$\sup_{t \geq 1} \sum_{s=1}^{\infty} P[|Z(t)|^{2+\delta} \geq s] < \infty. \tag{18}$$

Note that for any real k,

$$N(t) \geq k \Longleftrightarrow \sum_{i=1}^{\lfloor k \rfloor} u_i \leq t. \tag{19}$$

For any $s \geq 1$, we have

$$P[|Z(t)|^{2+\delta} \geq s] = P[Z(t) \leq -s^{\theta}] + P[Z(t) \geq s^{\theta}]. \tag{20}$$

Using the relationship (19), we obtain

$$\begin{aligned} P[Z(t) \leq -s^{\theta}] &= P[N(t) \leq t - s^{\theta} t^{1/2+d}] \\ &\leq P[N(t) < t - s^{\theta} t^{1/2+d} + 1] = \begin{cases} P(\sum_{i=1}^{\lfloor v(t,s) \rfloor} u_i > t), & s \leq a(t); \\ 0, & s > a(t). \end{cases} \end{aligned} \tag{21}$$

where $a(t) = t^{(1/2-d)(2+\delta)}$ and $v(t,s) = t - s^{\theta} t^{1/2+d} + 1$. Similarly,

$$P[Z(t) \geq s^{\theta}] = P[N(t) \geq t + s^{\theta} t^{1/2+d}] = P(\sum_{i=1}^{\lfloor g(t,s) \rfloor} u_i \leq t) \tag{22}$$

where $g(t,s) = t + s^{\theta} t^{1/2+d}$.

Next, we show that both (21) and (22) are summable in s, uniformly in t. We treat the cases $d \in (0, 1/2)$ and $d = 0$ separately, since in condition *iv)* of Theorem 1, we assume $p = 3 + \epsilon$ for $d = 0$, which is much weaker than what we assume for $d \in (0, 1/2)$.

Case I: $d \in (0, 1/2)$.

First, we consider (21). Suppose first that $s \leq a(t)$, so that $v(s,t) \geq 1$. Let

$$W = \frac{\sum_{i=1}^{\lfloor v(t,s) \rfloor} u_i - \lfloor v(t,s) \rfloor}{\lfloor v(t,s) \rfloor^{1/2+d}} .$$

Then, for $t \geq 4$, since $s^\theta t^{1/2+d} - 1 \geq \frac{1}{2} s^\theta t^{1/2+d}$ and $\lfloor v(t,s) \rfloor \leq v(t,s)$,

$$\begin{aligned} P(\sum_{i=1}^{\lfloor v(t,s) \rfloor} u_i > t) &= P\Big(\frac{\sum_{i=1}^{\lfloor v(t,s) \rfloor} u_i - \lfloor v(t,s) \rfloor}{\lfloor v(t,s) \rfloor^{1/2+d}} > \frac{t - \lfloor v(t,s) \rfloor}{\lfloor v(t,s) \rfloor^{1/2+d}} \Big) \leq P\Big(W > \frac{t - v(t,s)}{v(t,s)^{1/2+d}} \Big) \\ &\leq E(|W|^{2+\epsilon}) \cdot \frac{(t - s^\theta t^{1/2+d} + 1)^{(2+\epsilon)(1/2+d)}}{(s^\theta t^{1/2+d} - 1)^{2+\epsilon}} \\ &\leq E(|W|^{2+\epsilon}) \cdot \frac{(2t)^{(2+\epsilon)(1/2+d)}}{(\frac{1}{2} s^\theta t^{1/2+d})^{2+\epsilon}} = C \frac{E(|W|^{2+\epsilon})}{s^{(2+\epsilon)/(2+\delta)}} . \end{aligned} \quad (23)$$

where $C > 0$ is a constant.

For $s > a(t)$, $P[Z(t) \leq -s^{1/(2+\delta)}] = 0$.

Since $0 < \delta < \epsilon$, the righthand side of (23) is summable in s, uniformly in t, provided that $\sup_{t \geq 1, s \geq 1} E(|W|^{2+\epsilon}) < \infty$, which we show next.

Define

$$B_1 = \frac{u_1 - 1}{\lfloor v(t,s) \rfloor^{1/2+d}} \quad , \quad B_2 = \frac{\sum_{i=2}^{\lfloor v(t,s) \rfloor} (\tau_i - 1)}{\lfloor v(t,s) \rfloor^{1/2+d}},$$

so that $W = B_1 + B_2$. By Minkowski's Inequality,

$$E[|W|^{2+\epsilon}] \leq \Big[\big(E|B_1|^{2+\epsilon} \big)^{1/(2+\epsilon)} + \big(E|B_2|^{2+\epsilon} \big)^{1/(2+\epsilon)} \Big]^{2+\epsilon} .$$

Since $u_1 \leq \tau_1$, using $h(x) = (x+1)^{2+\epsilon}$ in $E[h(\tau_1)] = E^0[\tau_1 h(\tau_1)]$ (see Equation 1.4.2 on page 33 of Baccelli and Brémaud (2003)), and since our assumptions imply that $E^0[\tau_k^{3+\epsilon}] < \infty$, we obtain

$$\sup_{t \geq 1, s \geq 1} E|B_1|^{2+\epsilon} < \infty .$$

From Baccelli and Brémaud (2003, Equation 1.2.25, page 20), for any measurable function h,

$$E[h(\tau_2, \ldots, \tau_n)] = E^0[\tau_1 h(\tau_2, \ldots, \tau_n)] .$$

This, together with Holder's inequality, yields

$$E|B_2|^{2+\epsilon} = E^0(\tau_1|B_2|^{2+\epsilon}) \leq [E^0(\tau_1^{3+\epsilon})]^{1/(3+\epsilon)}\,[E^0|B_2|^{3+\epsilon}]^{(2+\epsilon)/(3+\epsilon)} \quad . \tag{24}$$

By assumption $iv)$ of Theorem 1, $\sup_{t\geq 1, s\geq 1} E^0|B_2|^{3+\epsilon} < \infty$, thus we have $\sup_{t\geq 1, s\geq 1} E|B_2|^{2+\epsilon} < \infty$, and

$$\sup_{t\geq 1, s\geq 1} E|W|^{2+\epsilon} = E\Big|\frac{\sum_{i=1}^{\lfloor v(t,s)\rfloor} u_i - \lfloor v(t,s)\rfloor}{\lfloor v(t,s)\rfloor^{1/2+d}}\Big|^{2+\epsilon} < \infty \quad . \tag{25}$$

Therefore,

$$\sup_{t\geq 1}\sum_{s=1}^{\infty} P[Z(t) \leq -s^{\theta}] = \sup_{t\geq 1}\sum_{s=1}^{\infty} P(\sum_{i=1}^{\lfloor v(t,s)\rfloor} u_i > t) < \infty. \tag{26}$$

Next, we consider (22). Defining

$$U = \frac{\sum_{i=1}^{\lfloor g(t,s)\rfloor} u_i - \lfloor g(t,s)\rfloor}{\lfloor g(t,s)\rfloor^{1/2+d}},$$

an argument similar to that in (23) gives

$$P(\sum_{i=1}^{\lfloor g(t,s)\rfloor} u_i \leq t) \quad \leq \quad E(|U|^m)\frac{(t+s^{\theta}t^{1/2+d})^{m(1/2+d)}}{(s^{\theta}t^{1/2+d}-1)^m}$$

for any $m > 0$.

For $t \geq 4$, since $s^{\theta}t^{1/2+d} - 1 \geq \frac{1}{2}s^{\theta}t^{1/2+d}$ and $t^{1/2+d} < t$, we obtain for all $m > 0$

$$P(\sum_{i=1}^{\lfloor g(t,s)\rfloor} u_i \leq t) \quad \leq \quad E(|U|^m)\frac{(t+s^{\theta}t^{1/2+d})^{m(1/2+d)}}{(\frac{1}{2}s^{\theta}t^{1/2+d})^m} \leq C\frac{E(|U|^m)}{s^{\theta m(1/2-d)}} \tag{27}$$

where C is a constant.

Since $d \in (0, 1/2)$, we can choose m sufficiently large so that (27) is summable in s. By a similar argument as in the proof of (25), we have for this same value of m that $\sup_{t\geq 1, s\geq 1} E(|U|^m) < \infty$. Therefore,

$$\sup_{t\geq 1}\sum_{s=1}^{\infty} P[Z(t) \geq s^{\theta}] = \sup_{t\geq 1}\sum_{s=1}^{\infty} P(\sum_{i=1}^{\lfloor g(t,s)\rfloor} u_i \leq t) < \infty \quad . \tag{28}$$

Case II: $d = 0$.

The bound for $P[Z(t) \leq -s^\theta]$ for case II follows along similar lines as in Case I, replacing d by 0 in the proof. Next, we obtain a bound for $P[Z(t) \geq s^\theta]$. Since $d = 0$, $g(t,s) = t + s^\theta t^{1/2}$.

Let integer $s_0 = \lfloor Kt^{(2+\delta)/2} \rfloor$ for some K to be chosen later. Consider

$$\sup_{t\geq 1} \sum_{s=1}^{\infty} P(\sum_{i=1}^{\lfloor g(t,s) \rfloor} u_i \leq t) \leq \sup_{t\geq 1} \sum_{s=1}^{s_0} P(\sum_{i=1}^{\lfloor g(t,s) \rfloor} u_i \leq t) + \sup_{t\geq 1} \sum_{s=s_0}^{\infty} P(\sum_{i=1}^{\lfloor g(t,s) \rfloor} u_i \leq t). \tag{29}$$

The first term on the righthand side of (29) involves summation in s from 1 to s_0. By (27) with $m = 2+\epsilon$, we obtain

$$P(\sum_{i=1}^{\lfloor g(t,s) \rfloor} u_i \leq t) \quad \leq \quad E(|U|^{2+\epsilon}) \frac{(t + s^\theta t^{1/2})^{(2+\epsilon)/2}}{(\frac{1}{2} s^\theta t^{1/2})^{2+\epsilon}} . \tag{30}$$

As in the proof of (25), we obtain $\sup_{t\geq 1, s\geq 1} E(|U|^{2+\epsilon}) < \infty$.

Since

$$(t + s^\theta t^{1/2})^{(2+\epsilon)/2} < C(t^{(2+\epsilon)/2} + s^{\theta(2+\epsilon)/2} t^{(2+\epsilon)/4}) \tag{31}$$

we obtain

$$P(\sum_{i=1}^{\lfloor g(t,s) \rfloor} u_i \leq t) \leq C\Big[\frac{1}{s^{\theta(2+\epsilon)}} + \frac{1}{s^{\theta(2+\epsilon)/2} t^{(2+\epsilon)/4}}\Big]. \tag{32}$$

The first term on the righthand side of (32) is summable in s since $\theta(2+\epsilon) = \frac{2+\epsilon}{2+\delta} > 1$. As for the second term, since $s_0 = \lfloor Kt^{(2+\delta)/2} \rfloor \leq Kt^{(2+\delta)/2}$, we obtain

$$\begin{aligned} \frac{1}{t^{(2+\epsilon)/4}} \sum_{s=1}^{Kt^{(2+\delta)/2}} \frac{1}{s^{\theta(2+\epsilon)/2}} \quad &\leq \quad \frac{C}{t^{(2+\epsilon)/4}} s^{-\theta(2+\epsilon)/2+1} \Big|_{s=\lfloor Kt^{(2+\delta)/2} \rfloor} \\ &\leq \quad \frac{C}{t^{(2+\epsilon)/4}} (Kt^{(2+\delta)/2})^{-\theta(2+\epsilon)/2+1} = Ct^{(\delta-\epsilon)/2} \end{aligned}$$

which is bounded uniformly in t. It follows that

$$\sum_{s=1}^{s_0} P(\sum_{i=1}^{\lfloor g(t,s) \rfloor} u_i \leq t) \leq C_1 + C_2$$

where the constants C_1 and C_2 are free of t. Hence

$$\sup_{t\geq 1} \sum_{s=1}^{s_0} P(\sum_{i=1}^{\lfloor g(t,s) \rfloor} u_i \leq t) < \infty \quad . \tag{33}$$

We now consider $P(\sum_{i=1}^{\lfloor g(t,s)\rfloor} u_i \le t)$ for $s \ge s_0$. By Equation 1.2.25 of Baccelli and Brémaud (2003)

$$P[Z(t) \ge s^\theta] = P(\sum_{i=1}^{\lfloor g(t,s)\rfloor} u_i \le t) = E^0\Big[\int_0^\infty I\{0 \le x \le \tau_1\} I\{\tau_1 - x + \sum_{i=2}^{\lfloor g(t,s)\rfloor} \tau_i \le t\} dx\Big]$$

$$\le \quad E^0\Big[\int_0^\infty I\{0 \le x \le \tau_1\} I\{\sum_{i=2}^{\lfloor g(t,s)\rfloor} \tau_i \le t\} dx\Big] = E^0 \tau_1 I\{\sum_{i=2}^{\lfloor g(t,s)\rfloor} \tau_i \le t\} \qquad . \tag{34}$$

We bound (34) by Holder's inequality:

$$E^0 \tau_1 I\{\sum_{i=1}^{\lfloor g(t,s)\rfloor} \tau_i \le t\} \le (E^0 \tau_1^\alpha)^{1/\alpha} \; [P^0(\sum_{i=1}^{\lfloor g(t,s)\rfloor} \tau_i \le t)]^{1/\beta}, \tag{35}$$

where $1/\alpha + 1/\beta = 1$, $\alpha > 0$, $\beta > 0$ and the values of α and β will be chosen later.

We now show that the term $[P^0(\sum_{i=1}^{\lfloor g(t,s)\rfloor} \tau_i \le t)]^{1/\beta}$ on the righthand side of (35) is bounded and summable in s. Note that, for all nonnegative integers i and all $a \ge 0$, $0 \le E^0(e^{-a\tau_i}) < \infty$, since by dominated convergence theorem, $\lim_{a \to 0} E^0(e^{-a\tau_i}) = 1$ and $\lim_{a\to\infty} E^0(e^{-a\tau_i}) = 0$. Hence

$$\exists\ 0 < a_0 < \infty, \text{ such that } E^0(e^{-a_0\tau_i}) = e^{-1},$$

where a_0 is free of i since $\{\tau_i\}$ are identically distributed under P^0.

Then,

$$P^0(\sum_{i=1}^{\lfloor g(t,s)\rfloor} \tau_i \le t) \quad = \quad P^0\Big(e^{-\frac{1}{\lfloor g(t,s)\rfloor}\sum_{i=1}^{\lfloor g(t,s)\rfloor} a_0\tau_i} \ge e^{-a_0 \frac{t}{\lfloor g(t,s)\rfloor}}\Big) \qquad . \tag{36}$$

Since by Jensen's inequality,

$$e^{-\frac{1}{\lfloor g(t,s)\rfloor}\sum_{i=1}^{\lfloor g(t,s)\rfloor} a_0\tau_i} \le \frac{1}{\lfloor g(t,s)\rfloor} \sum_{i=1}^{\lfloor g(t,s)\rfloor} e^{-a_0\tau_i}$$

we conclude from (36) that

$$P^0(\sum_{i=1}^{\lfloor g(t,s)\rfloor} \tau_i \le t) \quad \le \quad P^0\Big(\frac{1}{\lfloor g(t,s)\rfloor} \sum_{i=1}^{\lfloor g(t,s)\rfloor} e^{-a_0\tau_i} \ge e^{-a_0\frac{t}{\lfloor g(t,s)\rfloor}}\Big)$$

$$= \quad P^0\Big(\frac{1}{\lfloor g(t,s)\rfloor} \sum_{i=1}^{\lfloor g(t,s)\rfloor} x_i \ge e^{-1}(e^{1-a_0\frac{t}{\lfloor g(t,s)\rfloor}} - 1)\Big) \tag{37}$$

where $x_i = e^{-a_0\tau_i} - e^{-1}$.

Now, we choose $K > \max(2, 2a_0^{2+\delta})$. Since $s \geq s_0 = \lfloor Kt^{(2+\delta)/2} \rfloor$,

$$\begin{aligned} \lfloor g(t,s) \rfloor - a_0 t &\geq (t + s^\theta t^{1/2} - 1) - a_0 t \geq s^\theta t^{1/2} - a_0 t \geq (Kt^{(2+\delta)/2} - 1)^\theta t^{1/2} - a_0 t \\ &\geq (\frac{1}{2} K t^{(2+\delta)/2})^\theta t^{1/2} - a_0 t = [\Big(\frac{K}{2}\Big)^\theta - a_0] t \qquad \text{(since } t \geq 1) \\ &\geq \gamma > 0. \end{aligned}$$

Hence, for every $s \geq s_0$, $e^{1 - \frac{a_0 t}{\lfloor g(t,s) \rfloor}} - 1 > 0$.

Also, for fixed t, as $s \to \infty$, $e^{1 - \frac{a_0 t}{\lfloor g(t,s) \rfloor}}$ is monotonically nondecreasing. Thus,

$$\inf_{s \geq s_0} e^{1 - \frac{a_0 t}{\lfloor g(t,s) \rfloor}} - 1 \quad \geq \quad e^{1 - \frac{a_0 t}{\lfloor g(t,s_0) \rfloor}} - 1 \geq e^{1 - \frac{a_0 t}{g(t,s_0) - t}} - 1 \geq e^{1 - a_0 (\frac{2}{K})^\theta} - 1 = \xi,$$

using the fact that, for $s_0 = \lfloor Kt^{(2+\delta)/2} \rfloor \geq Kt^{(2+\delta)/2} - 1$,

$$g(t, s_0) - t \quad = \quad s_0^\theta t^{1/2} \geq (Kt^{(2+\delta)/2} - 1)^\theta t^{1/2} \geq (\frac{1}{2} K t^{(2+\delta)/2})^\theta t^{1/2} = \Big(\frac{K}{2}\Big)^\theta t$$

and for $K > 2a_0^{2+\delta}$, $0 < a_0 (\frac{2}{K})^\theta < 1$, so that $\xi > 0$.

Therefore, (37) becomes

$$\begin{aligned} P^0 (\sum_{i=1}^{\lfloor g(t,s) \rfloor} \tau_i \leq t) &\leq P^0 \Big(\frac{1}{\lfloor g(t,s) \rfloor} \sum_{i=1}^{\lfloor g(t,s) \rfloor} x_i \geq e^{-1} \xi \Big) \\ &\leq \frac{1}{\lfloor g(t,s) \rfloor^{m/2}} E^0 \Big| \frac{1}{\lfloor g(t,s) \rfloor^{1/2}} \sum_{i=1}^{\lfloor g(t,s) \rfloor} x_i \Big|^m \frac{1}{(e^{-1}\xi)^m}, \end{aligned} \tag{38}$$

for any $m > 0$.

Note that $E^0(x_i^k) < \infty$ for all positive integers k. Also $\{x_i\}$ is strong-mixing since $\{\tau_i\}$ is, hence by Yokoyama (1980), for any $m > 0$

$$E^0 \Big| \frac{1}{\lfloor g(t,s) \rfloor^{1/2}} \sum_{i=1}^{\lfloor g(t,s) \rfloor} x_i \Big|^m < C$$

and (38) yields,

$$P^0 (\sum_{i=1}^{\lfloor g(t,s) \rfloor} \tau_i \leq t) \leq \frac{C}{(t + s^\theta t^{1/2} - 1)^{m/2}} \leq \frac{C}{s^{m\theta}} \qquad \text{(since } t \geq 1).$$

Thus,

$$\left[P^0(\sum_{i=1}^{\lfloor g(t,s)\rfloor} \tau_i < t)\right]^{1/\beta} \leq \frac{C}{s^{m\theta/\beta}} \tag{39}$$

Now, in the righthand side of (35), we can choose $\alpha = 3 + \epsilon > 0$ and $\beta = \frac{\alpha}{\alpha-1} > 0$. Given this choice of α, β, we then choose m sufficiently large in (39) so that it is summable in s. Therefore, the righthand side of (35) is summable in s, uniformly in t.

This then implies,

$$\sup_{t\geq 1} \sum_{s=s_0}^{\infty} P(\sum_{i=1}^{\lfloor g(t,s)\rfloor} u_i < t) < \infty \quad . \tag{40}$$

In all, by (33) and (40), we obtain

$$\sup_{t\geq 1} \sum_{s=1}^{\infty} P[Z(t) \geq s^{\theta}] = \sup_{t\geq 1} \sum_{s=1}^{\infty} P(\sum_{i=1}^{\lfloor g(t,s)\rfloor} u_i \leq t) < \infty \quad . \tag{41}$$

□

Lemma 3 *Under the LMSD model described in Theorem 2 with memory parameter $d \in [0, \frac{1}{2})$, P^0 is $\{\tau_k\}$-mixing; The durations $\{\tau_k\}$ generated by the ACD(1,1) model described in Theorem 3 are exponential α-mixing.*

Proof: Under P^0, $\{h_k\}$ is a stationary Gaussian process with a log spectral density having an integral on $[-\pi, \pi]$ that is greater than $-\infty$, so that the innovation variance is positive. Since Gaussian processes are time reversible, it follows that we can represent $h_k = \sum_{j=0}^{\infty} a_j w_{k+j}$ where $\sum a_j^2 < \infty$ and $\{w_k\}$ is an *iid* Gaussian sequence. Arguing as in the proof of Theorem 17.3.1 of Ibragimov and Linnik (1971), *pp.* 311–312, replacing $\{\ldots w_{k-1}, w_k\}$ by $\{w_k, w_{k+1}, \ldots\}$, it follows that P^0 is $\{h_k\}$-mixing. Since the $\{\epsilon_k\}$ are *iid* it follows that P^0 is also $\{\epsilon_k\}$-mixing. Since for any process $\{\xi_k\}$, P^0 is $\{\xi_k\}$-mixing if and only if the future tail σ-field of $\{\xi_k\}$ is trivial (see, e.g., Nieuwenhuis (1989), Equation (3.3)), it follows from Lemma 4 that P^0 is $\{\tau_k\}$-mixing, where $\tau_k = e^{h_k}\epsilon_k$.

For the ACD(1,1) model, by Proposition 17 of Carrasco and Chen (2002), $\{\tau_k\}$ is exponential β-mixing (or also called absolutely regular) if $\{\tau_0, \psi_0\}$ are initialized from the stationary distribution. Their result still holds for a doubly infinite sequence $\{\tau_k\}, k \in (-\infty, \infty)$. It is well known that β-mixing implies α-mixing (or strong mixing), (see Bradley (2005), Section 2.1). Therefore, $\{\tau_k\}$ is also exponential α-mixing, which further implies $\{\tau_k\}$-mixing of P^0 for the ACD(1,1) model, see Nieuwenhuis (1989), Equation (3.5). $\square$

Lemma 4 *Let $\{\xi_s\}$ and $\{\zeta_s\}$ be two* independent *processes whose future tail σ-fields are trivial. Then the future tail σ-field of the process $\{\xi_s, \zeta_s\}$ is trivial.*

Proof: Define $\mathcal{S}_t = \sigma(\xi_s, s \geq t)$, $\mathcal{T}_t = \sigma(\zeta_s, s \geq t)$ and $\mathcal{U}_t = \sigma(\xi_s, \zeta_s, s \geq t)$. As pointed out by Ibragimov and Linnik (1971, p. 303) (for regularity), to prove that $\mathcal{U}_\infty$ is trivial, it suffices to prove that for all $\mathcal{U}_0$-measurable zero mean random variables η such that $\mathbf{E}[\eta^2] \leq 1$, $\mathbf{E}[\eta \mid \mathcal{U}_t]$ converges to 0 in quadratic mean. By standard arguments, it suffices to prove this for a random variable η that can be expressed as $\eta = \eta_1\eta_2$ with η_1 $\mathcal{S}_0$-measurable and η_2 $\mathcal{T}_0$-measurable and, without loss of generality, both with zero mean. Then, by independence of $\{\xi_s\}$ and $\{\zeta_s\}$,

$$\mathbf{E}[\eta \mid \mathcal{U}_t] = \mathbf{E}[\eta_1 \mid \mathcal{S}_t] \times \mathbf{E}[\eta_2 \mid \mathcal{T}_t]\ .$$

Since $\mathcal{S}_\infty$ and $\mathcal{T}_\infty$ are trivial, both terms in the right hand side above tend to 0 in q.m. By independence, their product also tends to 0 in q.m. $\square$

Proof of Theorem 6: We will separately consider the cases $d \in (0, 1/2)$ and $d = 0$.

<u>Case I:</u> $d \in (0, 1/2)$.

By the law of total variance

$$\mathrm{var}(\widetilde{RV}_T) = E\{\mathrm{var}[\widetilde{RV}_T|N(\cdot)]\} + \mathrm{var}\{E[\widetilde{RV}_T|N(\cdot)]\}$$

First, consider $\mathrm{Var}\{E[\widetilde{RV}_T|N(\cdot)]\}$. We have

$$E[\widetilde{RV}_T|N(\cdot)] = E[\sum_{j=1}^{\lfloor N(T)/K \rfloor} \tilde{r}_j^2 | N(\cdot)] = \sigma_{\tilde{r}}^2 \Big\lfloor \frac{N(T)}{K} \Big\rfloor, \tag{42}$$

where $\sigma_{\tilde{r}}^2 = E(\tilde{r}_j^2)$.

By Theorem 1, we know that $\text{Var}\left[\frac{N(T)}{K}\right] \sim CT^{2d+1}$ which goes to infinity as T increases. Since

$$\left\lfloor \frac{N(T)}{K} \right\rfloor = \frac{N(T)}{K} - A \qquad ,$$

where the random variable A can take only K finite values in $[0, 1)$ hence $\text{var}(A)$ must be finite, by the Cauchy-Schwartz inequality, we obtain

$$\text{var}\left\lfloor \frac{N(T)}{K} \right\rfloor \sim CT^{2d+1} \qquad . \tag{43}$$

Using (43) in (42), it follows that

$$\text{var}\{E[\widetilde{RV}_T|N(\cdot)]\} \sim CT^{2d+1} \qquad . \tag{44}$$

Next, conditionally on $N(\cdot)$, since $\{\tilde{r}_j\}$ is at most q-dependent, by the Cauchy-Schwartz inequality, we have

$$0 \le \text{var}[\widetilde{RV}_T|N(\cdot)] = \text{var}[\sum_{j=1}^{\lfloor N(T)/K \rfloor} \tilde{r}_j^2|N(\cdot)] \le \sigma_{\tilde{r}^2}^2 \cdot (1+2q) \cdot \left\lfloor \frac{N(T)}{K} \right\rfloor$$

where $\sigma_{\tilde{r}^2}^2 = \text{var}(\tilde{r}_j^2)$. Since $E\left\lfloor \frac{N(T)}{K} \right\rfloor = O(T)$,

$$E\{\text{var}[\widetilde{RV}_T|N(\cdot)]\} \le CE\left\lfloor \frac{N(T)}{K} \right\rfloor = O(T), \tag{45}$$

where $C > 0$.

Finally, by (44) and (45), we obtain $\text{var}(\widetilde{RV}_T) \sim CT^{2d+1}$.

<u>Case II:</u> $d = 0$.

Equation (44) still holds when $d = 0$ since Theorem 1 includes the $d = 0$ case.

Next, we consider $E\{\text{var}[\widetilde{RV}_T|N(\cdot)]\}$. Denote the lag-$k$ autocorrelation of $\{\tilde{r}_j^2\}$ by γ_k. Since $\{\tilde{r}_j\}$ is at most q-dependent, we have $\gamma_k = 0$ for all $k > q$. Conditionally on $N(\cdot)$, if $\left\lfloor \frac{N(T)}{K} \right\rfloor \geq q$, we have

$$\begin{aligned}
\operatorname{var}[\widetilde{RV}_T | N(\cdot)] &= \operatorname{var}\Big\{ \sum_{j=1}^{\lfloor N(T)/K \rfloor} \tilde{r}_j^2 | N(\cdot) \Big\} \\
&= \Big\lfloor \frac{N(T)}{K} \Big\rfloor \sigma^2_{\tilde{r}^2} + 2(\Big\lfloor \frac{N(T)}{K} \Big\rfloor - 1)\sigma^2_{\tilde{r}^2}\gamma_1 + \ldots + 2(\Big\lfloor \frac{N(T)}{K} \Big\rfloor - q)\sigma^2_{\tilde{r}^2}\gamma_q \\
&= \underbrace{(1 + 2\gamma_1 + \ldots + 2\gamma_q)\sigma^2_{\tilde{r}^2}}_{K_{1,q}} \Big\lfloor \frac{N(T)}{K} \Big\rfloor \underbrace{-2(\gamma_1 + \ldots + q\gamma_q)\sigma^2_{\tilde{r}^2}}_{K_{2,q}} \\
&= K_{1,q} \Big\lfloor \frac{N(T)}{K} \Big\rfloor + K_{2,q}. \qquad (46)
\end{aligned}$$

Since $(1 + 2\gamma_1 + \ldots + 2\gamma_q)$ is equal to the spectral density of $\{\tilde{r}_j^2\}$ at zero frequency, which we assume to be positive, $K_{1,q} > 0$.

Similarly, conditionally on $N(\cdot)$, if $\left\lfloor \frac{N(T)}{K} \right\rfloor = k$ where $k = 0, \ldots, q-1$, we have

$$\operatorname{var}[\widetilde{RV}_T | N(\cdot)] = K_{1,k} \Big\lfloor \frac{N(T)}{K} \Big\rfloor + K_{2,k} \qquad (47)$$

for some constants $K_{1,k}$ and $K_{2,k}$.

Overall, conditionally on $N(\cdot)$, by (46) and (47),

$$\begin{aligned}
&\operatorname{var}[\widetilde{RV}_T | N(\cdot)] \\
= \;& \Big\{ K_{1,q} \Big\lfloor \frac{N(T)}{K} \Big\rfloor + K_{2,q} \Big\} I\{\Big\lfloor \frac{N(T)}{K} \Big\rfloor \geq q\} + \Big\{ K_{1,q-1} \Big\lfloor \frac{N(T)}{K} \Big\rfloor + K_{2,q-1} \Big\} I\{N(T) = q-1\} \\
& + \ldots + \Big\{ K_{1,0} \Big\lfloor \frac{N(T)}{K} \Big\rfloor + K_{2,0} \Big\} I\{N(T) = 0\} \\
= \;& \Big\{ K_{1,q} \Big\lfloor \frac{N(T)}{K} \Big\rfloor + K_{2,q} \Big\} - \underbrace{\Big\{ K_{1,q} \Big\lfloor \frac{N(T)}{K} \Big\rfloor + K_{2,q} \Big\} I\{\Big\lfloor \frac{N(T)}{K} \Big\rfloor < q\}}_{\Gamma_q} \\
& + \underbrace{\Big\{ K_{1,q-1} \Big\lfloor \frac{N(T)}{K} \Big\rfloor + K_{2,q-1} \Big\} I\{\Big\lfloor \frac{N(T)}{K} \Big\rfloor = q-1\}}_{\Gamma_{q-1}} + \ldots + \underbrace{\Big\{ K_{1,0} \Big\lfloor \frac{N(T)}{K} \Big\rfloor + K_{2,0} \Big\} I\{\Big\lfloor \frac{N(T)}{K} \Big\rfloor = 0\}}_{\Gamma_0}
\end{aligned}$$

Since $K_{1,q} > 0$ and $E\left\lfloor \frac{N(T)}{K} \right\rfloor = O(n)$, we have

$$E\Big\{ K_{1,q} \Big\lfloor \frac{N(T)}{K} \Big\rfloor + K_{2,q} \Big\} \sim CT$$

for some $C > 0$. Therefore, to prove the theorem for the $d = 0$ case, it is enough to show that $E(\Gamma_0), \ldots, E(\Gamma_q)$ are all $o(T)$.

Consider Γ_q. Since

$$\Big|E\Big\{\big[K_{1,q}\Big\lfloor\frac{N(T)}{K}\Big\rfloor + K_{2,q}\big]I\{\Big\lfloor\frac{N(T)}{K}\Big\rfloor < q\}\Big\}\Big| \leq |K_{1,q}|q + |K_{2,q}| = O(1)$$

we have $E(\Gamma_q) = o(T)$. Similarly, $E(\Gamma_k) = o(T)$ for $k = 0, \ldots, q-1$. $\square$