The Periodogram of Spurious Long-Memory Processes[*]

Christian Leschinski and Philipp Sibbertsen[1]

Institute of Statistics, Faculty of Economics and Management,
Leibniz University Hannover, Germany

Abstract

We derive the properties of the periodogram local to the zero frequency for a large class of spurious long-memory processes. The periodogram is of crucial importance in this context, since it forms the basis for most commonly used estimation methods for the memory parameter. The class considered nests a wide range of processes such as deterministic or stochastic structural breaks and smooth trends as special cases. Several previous results on these special cases are generalized and extended. All of the spurious long-memory processes considered share the property that their impact on the periodogram at the Fourier frequencies local to the origin is different than that of true long-memory processes. Both types of processes therefore exhibit clearly distinct empirical features.

JEL-Numbers: C18, C32
Keywords: Long Memory · Spurious Long Memory · Structural Change

[*] An earlier version of this paper was published as a discussion paper under the title "Origins of Spurious Long Memory" as Hannover Economic Papers 595.

[1] Corresponding author:
Phone: +49-511-762-3783
E-Mail: sibbertsen@statistik.uni-hannover.de

1 Introduction

Long memory is defined by a hyperbolic decay of the autocorrelation function or equivalently by the order of a pole at the origin of the spectral density of a time series. This behavior is commonly found in data sets from a wide range of subject areas such as economics, finance, hydrology, or climatology. However, the applicability of long-memory time-series models is severely hampered by the fact that a variety of structural-change and trend processes is known to generate similar empirical behavior such as a pole in the spectrum at the origin and significant autocorrelations at large lags. This phenomenon is widely studied and referred to as spurious long memory. However, previous contributions typically focus on specific data generating processes that do not allow to draw general conclusions.

In this paper, we study the periodogram of a large class of structural-change and trend processes that can generate spurious long memory. The generality of these results allows conclusions about the nature of spurious long memory and the different impact of smooth and abrupt level changes, rare and more frequent level shifts, and accumulative or non-accumulative level-shift processes.

As examples for the application of these results, we consider a number of spurious long-memory processes that have been discussed in the previous literature. In particular, we derive the properties of the periodogram for stationary random level-shift processes, Markov-switching-mean models, the STOPBREAK process, and quadratic trends. Furthermore, we are able to recover several previous findings on the periodogram of non-stationary random level-shift processes, deterministic breaks, and fractional and linear trends that are extended and generalized. For example, we allow for more complex dependence structures in subsequent mean changes, and we can provide exact rates instead of upper bounds for a number of deterministic trends.

All of these processes are treated within a joint framework that allows for both – abrupt level shifts and smooth deterministic trends. It is shown how different assumptions on the data generating process can lead to different growth rates of the j-th periodogram ordinate if j is fixed and the sample size T increases, and to different rates of decay, as j increases and T is fixed.

The literature on spurious long memory has produced a diverse range of results. Smooth trends have been studied since Bhattacharya et al. (1983), who show that the fractional trend model generates a bias in the R/S statistic so that it indicates the presence of long memory. Further results on the impact of trends (including trends with structural breaks) on R/S-type statistics are obtained by Giraitis et al. (2001). Künsch (1986) establishes an upper bound on the impact of monotonous trends on the periodogram at low frequencies. A similar result is obtained by Iacone (2010), and Qu (2011) shows that the same upper bound holds for Lipschitz continuous trends.

With respect to abrupt level changes, Lobato and Savin (1998) show that the empirical autocorrelation function of a structural break process converges to non-zero constants at all lags. Granger and Teräsvirta (1999) demonstrate that the log-periodogram (GPH) estimator of Geweke and Porter-Hudak (1983) is biased for a Markov-switching-mean model with low switching probabilities. An analytical expression for the bias of the GPH estimator in presence of mean changes is derived by Smith (2005).

Similar to the results of Lobato and Savin (1998) for a deterministic break, Gourieroux and Jasiak (2001) study the empirical autocovariance function for a class of stationary random level-shift processes, and Granger and Hyung (2004) show that the autocorrelations of a random level-shift process converge to a constant as the sample size goes to infinity.

Diebold and Inoue (2001) study the growth rate of the partial sum of a number of processes including random level shifts, the STOPBREAK process of Engle and Smith (1999), and a class of Markov-switching processes and show that it can be the same as that of an $I(d)$ process given the right conditions.

There are also some contributions that consider the relationship between long memory and the concept of fractional integration. For example Granger and Ding (1996) discuss several non-linear processes that have long memory and Haldrup and Valdés (2017) show that a long-memory process generated by aggregation is not fractionally integrated. Miller and Park (2010) consider nonlinear transformations of random walks with heavy tailed innovations. They show that these processes can empirically replicate characteristics of long-memory processes.

A number of contributions also consider duration-driven long range dependence. This includes Taqqu et al. (1997), Parke (1999), Liu (2000), Gourieroux and Jasiak (2001), Davidson and Sibbertsen (2005), and Hsieh et al. (2007). These processes have stationary random level shifts and heavy tailed regime lengths so that their autocorrelation function fulfills the long memory definition. Similar results are obtained by Leipus and Surgailis (2003) for a class of regime switching random coefficient autoregressive processes. Unlike fractionally integrated processes, however, a number of contributions including Mikosch et al. (2002), Leipus et al. (2005), and Davidson and Sibbertsen (2005) show that the partial sums of these processes converge to stable Levy processes with independent increments.

Most closely related to our paper are those of Mikosch and Stărică (2004), Qu and Perron (2007), and McCloskey and Perron (2013), who consider the order of the periodogram of processes with abrupt level changes. Mikosch and Stărică (2004) derive the expectation of the periodogram of a time series consisting of disjoint subsamples generated by distinct short memory processes and show that it is $O(Tj^{-2})$ at the j-th Fourier frequency local to the origin, where T is the sample size. Similar results are obtained by Qu and Perron (2007) for simple non-stationary random level-shift process. McCloskey and Perron (2013) show that this result also applies to deterministic structural breaks. The focus on the properties of the periodogram is particularly useful, since this is the basis for most popular estimation procedures. Based on these findings, Smith (2005), Iacone (2010), Qu (2011), McCloskey and Perron (2013), Hou and Perron (2014), Christensen and Varneskov (2017), McCloskey and Hill (2017), and Sibbertsen et al. (2018)

develop a number of estimation and testing procedures that allow to distinguish true and spurious long memory.

One important contribution of this paper is to show that the rate $O_P(j^{-2})$ for fixed T is an upper bound for the periodogram of a much wider class of processes, but the scaling with T is specific to processes with rare shifts. This finding can be expected to be useful for the development of future robust methods.

While it is sometimes argued that long-memory and structural-change models can be used interchangeably, since they model the same data feature, our findings show for all of the processes discussed above that spurious long memory has a different effect at the origin of the periodogram than true long memory and is thus empirically distinct. In general, the effect of structural change tends to be restricted to a smaller neighborhood of the origin than that of long memory. The only exception are non-accumulative and relatively frequent level shift processes that affect a larger number of Fourier frequencies.

The rest of the paper is organized as follows. Section 2 gives a short discussion of the orders of poles in the spectrum for some typical processes. Section 3 discusses our structural-change model and gives some first results on its discrete Fourier transform (DFT) and the properties of its components. In Section 4, we use these results to establish the order of the periodogram for smooth trends, and abrupt level changes are treated in Section 5. The relationship of these results to other findings in the literature and their application to several examples is discussed in Section 6. Finally, Section 7 concludes.

2 Long Memory and the Behavior of the Periodogram

It is said that a stochastic process exhibits stationary long memory if the spectral density local to the origin behaves as

$$\lim_{\lambda\to 0} f(\lambda) \sim G|\lambda|^{-2d}, \tag{1}$$

where $0 < d < 1/2$ is the memory parameter. There are a number of competing definitions in the literature that are not in all cases equivalent to the one adopted here. For example, long memory is often defined through a hyperbolic decay of the autocorrelation function, the rate of the partial sum, self similarity in the covariance sense, the summability of the autocovariance function, the boundedness of the spectral density, or the limit process of the partial sums. An extensive review of these competing definitions is provided by Guégan (2005) and a recent discussion can be found in Haldrup and Valdés (2017).

Nevertheless, we argue that the definition in (1) is the most prevalent and the one which is of the highest practical relevance, since it is the basis for most modern estimation procedures such as the log-periodogram regression of Geweke and Porter-Hudak (1983) and Robinson (1995a), the local Whittle estimator of Künsch (1987) and Robinson (1995b), and the Whittle-ML estimator of Fox and Taqqu (1986) and others.

Furthermore, it is one of the main points in the seminal paper of Diebold and Inoue

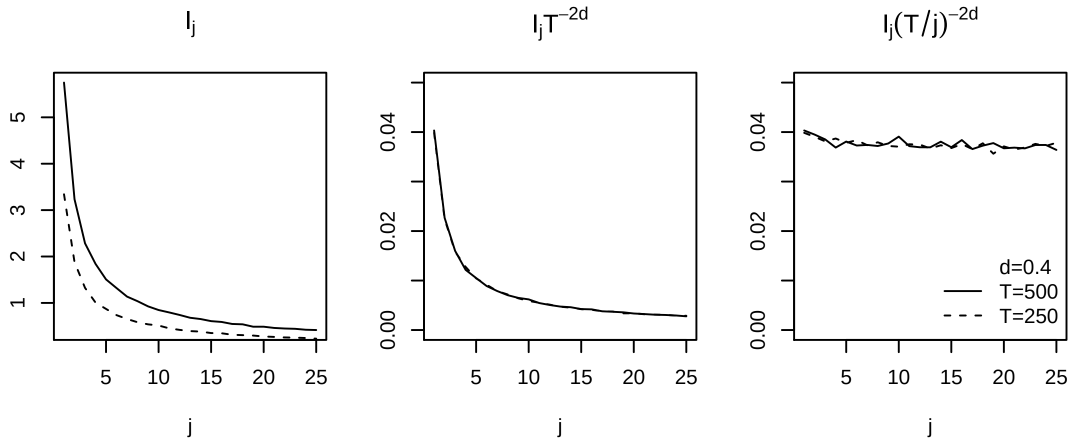

Figure 1: Average (rescaled) periodogram of a fractionally integrated Gaussian noise with $d = 0.4$ local to the origin. The index j indicates the Fourier frequency.

(2001) that (1) should be used to define long memory, since the definition in terms of the partial sum is also fulfilled by a range of other processes that should not be considered to be processes with long-range dependence.

Spurious long memory, on the other hand, is not well defined and we use the term loosely for the collection of behaviors that are discussed in the related literature. These typically exhibit long memory according to one of the less restrictive definitions, have a behavior that is empirically hard to distinguish from that of processes that fulfill one of the definitions, or cause a positive bias in one of the popular estimators.

The defining property of a long-memory process according to (1) is a pole in the spectrum that has a specific shape depending on the memory parameter d. Since the spectral density is not directly observable in practice, it can only be determined from the periodogram whether a time series has true or spurious long memory. This is the quantity that is used to draw conclusions about $f(\lambda)$. We therefore focus on the behavior of the periodogram at the Fourier frequencies $\lambda_j = (2\pi j)/T$ local to the origin as $\lambda_j \to 0_+$, or equivalently $j/T \to 0$. Denote the Fourier transform of the series z_t by $w_z(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} z_t e^{i\lambda t}$. Then the periodogram is given by $I_{zz}(\lambda) = w_z(\lambda) w_z^*(\lambda)$, where the asterisk denotes the complex conjugate.

From the definition in (1), the periodogram of a true long-memory process is

$$I_{zz}(\lambda_j) = O_P\left(\lambda_j^{-2d}\right) = O_P\left(\left(\frac{2\pi j}{T}\right)^{-2d}\right) = O_P\left(\left(\frac{T}{j}\right)^{2d}\right). \tag{2}$$

The long-memory definition therefore has two implications for the expected periodogram ordinates local to zero. First, for fixed j, the periodogram of a true long-memory process grows with rate T^{2d} as the sample size increases. And second, the magnitude of the pole shrinks with rate j^{-2d}, as the distance from the origin increases.

This is illustrated in Figure 1 that shows the effect of different standardizations on the average periodogram ordinates local to the origin for different sample sizes. The definition of long memory therefore determines the impact of low frequencies and the relative impact of low frequencies and intermediate frequencies on the variance of the process.

In contrast to the long memory case, the random level-shift process of Qu and Perron (2007) and deterministic structural breaks in McCloskey and Perron (2013) are found to generate periodograms that are of order

$$O_P(T/j^2), \tag{3}$$

which means that the peak at the origin grows with T like that of a long-memory process with $d = 1/2$, but the rate of decay for increasing j is like that of a long-memory process with $d = 1$.

In the following, we consider a component model of the form

$$y_t = \mu_t + x_t, \tag{4}$$

where x_t is a zero-mean time-series process that possibly has long memory and μ_t is a time-varying mean that can be stochastic or deterministic and $t = 1,...,T$. It is easy to show that

$$I_{yy}(\lambda) = I_{\mu\mu}(\lambda) + I_{xx}(\lambda) + I_{\mu x}(\lambda) + I_{x\mu}(\lambda),$$

so that

$$E[I_{yy}(\lambda)] = E[I_{\mu\mu}(\lambda)] + E[I_{xx}(\lambda)],$$

if x_t and μ_t are assumed to be independent. Since it is known for stationary processes that $E[I_{xx}(\lambda)] = f_x(\lambda)$, the remainder of this paper focuses on the behavior of $I_{\mu\mu}(\lambda) = I(\lambda)$, where we drop the subscripts for notational convenience.

For $j = \lfloor T^{\epsilon} \rfloor$, with $\epsilon \in (0,1)$ and from (2) and (3) it is clear to see that it depends on the position parameter ϵ, whether the periodogram of (4) is dominated by the structural-change component that is $O_P(T^{1-2\epsilon})$ or the long-memory component that is $O_P(T^{2d(1-\epsilon)})$. This is the property used by the procedures of Qu and Perron (2007), Qu (2011), McCloskey and Perron (2013), Hou and Perron (2014), McCloskey and Hill (2017), and Sibbertsen et al. (2018) to discriminate between true long memory and random level shifts. Therefore, the applicability of these procedures depends on whether the order in (3) is specific to the data generating process (DGP) considered or whether it generalizes to other structural-change processes.

3 A General Model of Structural Change and its Fourier Transform

To study the behavior of the periodogram of potential spurious long-memory processes, we require a versatile representation of the structural-change process μ_t. Therefore, we consider structural-change processes that can be expressed as

$$\mu_t = \mu_0 + \sum_{k=1}^{K} \Delta\mu_k \mathbb{I}(t \geqslant T_k), \tag{5}$$

where μ_0 is the initial offset of the first mean from the sample mean $\mu_0 = \mu_1 - \mu$ and $\mu = T^{-1} \sum_{t=1}^{T} \mu_t$. The variable K is the number of observations for which there is a structural change in the mean relative to the previous observation, and T_k is the point in time at which the k-th change occurs. If we set $T_0 = 0$, then we can also express μ as $\mu = \sum_{k=0}^{K} \mu_k (T_{k+1} - T_k)/T$.

The representation in (5) is extremely versatile. For deterministic structural breaks this representation is obvious and commonly used, albeit μ_0 usually represents the mean of the initial observations and not the offset from the time mean.

If the changes $\Delta\mu_k$ and the breakpoints T_k are stochastic, then the process nests a random level-shift process. Even a random walk is nested in (5) if $K = T$, $T_k = t$ for all $k = t$, and the $\Delta\mu_k$ are a martingale difference sequence.

Similarly, deterministic trends can be represented as in (5). Deterministic trends are usually modeled as functions $h(s)$ on $[0,1]$. Let $s_t = t/T$ for $t = 1,...,T$, then we have $h(s_t) = h(0) + \sum_{k=1}^{t} [h(k/T) - h((k-1)/T)]$ and obviously $\mu_t = h(s_t)$ if $\Delta\mu_k = [h(k/T) - h((k-1)/T)]$, $K = T$, and $T_k = 1,2,...,T$. Restrictions on $h(s)$ – such as Lipschitz-continuity – correspond to restrictions on the $\Delta\mu_k$.

It is important to note that the nature of structural change in (5) is accumulative in the sense that the mean at time t depends on all shifts that occurred before time t. While this seems appropriate for a wide range of commonly used structural-change models such as deterministic breaks, random level shifts, and the STOPBREAK process, there is a number of processes such as stationary random level shifts, Markov-switching mean models, or duration-driven long-range-dependent processes, for which it is more convenient to consider the following non-cumulative model of structural change

$$\mu_t = \mu_0 + \sum_{k=0}^{K} \mu_k \mathbb{I}(T_{k-1} \leqslant t < T_k), \tag{6}$$

where μ_0 is the unconditional expectation of μ_t, $T_{-1} = 1$, and $\mathbb{I}(\cdot)$ is the indicator function that takes the value one if its argument is true. Here, the mean in each segment between two level changes is determined by μ_k, and the dependence between the means in neighboring segments is directly determined by the dependence structure in the μ_k. Since the mean in each segment between breaks is re-centered around the mean, the variance of the process in (6) is independent of t, whereas that of (5) is not.

For the structural-change processes in (5) and (6) we have the following result.

Lemma 1. *The discrete Fourier transform (DFT) of the process in (5) can be represented as*

$$w_{\mu}(\lambda_j) = -\frac{1}{\sqrt{2\pi T}} \sum_{k=1}^{K} \Delta\mu_k D_{T_k}(\lambda_j),$$

and that of the process in (6) can be represented as

$$w_{\mu}(\lambda_j) = -\frac{1}{\sqrt{2\pi T}} \sum_{k=0}^{K} \mu_k \left(D_{T_k - 1}(\lambda_j) - D_{T_{k-1}}(\lambda_j) \right),$$

where $D_{T_k}(\lambda) = \sum_{t=1}^{T_k} e^{i\lambda t}$ *is a version of the Dirichlet kernel.*

The proof of this and all following results is given in the appendix. Note that Lemma 1 is completely algebraic and we do not impose any conditions on the $\Delta\mu_k$, μ_k, or T_k.

It is well known that the DFT of a constant is zero at the Fourier frequencies λ_j. This is the mechanism behind the self-centering property of the periodogram and the last step in the proof of Lemma 1. However, the DFT of an indicator function is non-zero at the Fourier frequencies and so is that of a structural-change process that can be expressed in terms of indicator functions as in (5) or (6).

Lemma 1 shows that the DFT of an accumulative structural-change process can be represented as a sum of Dirichlet kernels that depends on the location of the mean changes – each weighted by the magnitude of the corresponding mean change. Similarly, the DFT of the non-accumulative structural-change process can be represented as a sum of differences between Dirichlet kernels – each weighted by the magnitude of the corresponding mean.

Since Lemma 1 implies that the properties of the DFT and thus the periodogram of a structural-change process are directly related to those of the Dirichlet kernel, the following lemma provides an approximation for the Dirichlet kernel at frequencies local to zero.

Lemma 2. *We have for* $T_k/T = \delta_k \in [0,1]$ *and* $j/T \to 0$,

$$D_{T_k}(\lambda_j) = Tj^{-1}\frac{\sin(2\delta_k \pi j)}{2\pi} + \sin^2(\pi j \delta_k) + i\left[Tj^{-1}\frac{\sin^2(\pi\delta_k j)}{\pi} - \frac{1}{2}\sin(2\pi\delta_k j)\right] + O_P(jT^{-1}).$$

Clearly, from Lemma 2, both the real and the imaginary part of the Dirichlet kernel are $O(Tj^{-1})$ for deterministic δ_k and $O_p(Tj^{-1})$ if any of the δ_k is stochastic. Furthermore, the order is exact. Again, this is an approximation based on a Laurent expansion that holds irrespective of the stochastic properties of the T_k.

For deterministic structural breaks the result that $I(\lambda) = O(T/j^2)$ from McCloskey and Perron (2013) is obtained as a direct consequence of Lemmas 1 and 2. The same holds true for the result that the random level-shift process has a periodogram of order $O_p(T/j^2)$, as derived in Qu and Perron (2007). The full range of admissible processes in (5) and (6) will be covered in Section 5. Before that, the next section provides detailed results on the behavior of smooth trends.

4 The Periodogram of Smooth Deterministic Trends

Some of the earliest discussions of spurious long memory in the literature consider the effect of deterministic trends. Bhattacharya et al. (1983), for example, derive the limit of the rescaled range statistic for a class of trend processes that is referred to as fractional trends. For monotonous trends Künsch (1986) showed that $I(\lambda) = O(T/j^2)$ – similar results are later derived by Qu (2011) and Iacone (2010). Based on the difference between this rate and that of the long-memory process in (2), Künsch (1986) argues that asymptotically consistent estimates of d in presence of monotonous trends can be obtained by trimmed log-periodogram regressions, as shown by McCloskey and Perron (2013).

Based on the representation result from Lemma 1 and the results on the order of the Dirichlet kernel in Lemma 2, we can now derive the exact order of the periodogram of a deterministic trend.

Consider a trend function $h(s,T): [0,1] \times \mathbb{N} \rightarrow \mathbb{R}$, such that $\mu_t = h(t/T,T)$. Note that the trend is allowed to depend on both, the sample fraction s and the sample size T. This means we allow for a triangular array structure, which is necessary to include examples such as the typical linear trend model $\mu_t = \beta_0 + \beta_1 t$, which is obtained as a special case for $h(s,T) = \beta_0 + \beta_1 sT$.

To establish the behavior of the periodogram of a deterministic trend model, the trend is required to be bounded, continuous, and sufficiently smooth so that the first derivative is finite.

Assumption 1. $|h(s,T)|$, $\left|\frac{\partial h(s,T)}{\partial s}\right| < \infty$, *for* $s \in [0,1]$.

For finite K it is obvious from Lemma 1 that the DFT will have the order of the Dirichlet kernel when there is a $\Delta\mu_k > 0$. However, in the case of a Lipschitz continuous trend, we have $|h(r) - h(s)| \leqslant \tilde{K}|r-s|$ for a finite constant $\tilde{K}$. Therefore, at successive observations t and $t-1$, we have $\Delta\mu_t \leqslant \tilde{K}/T \rightarrow 0$ so that the behavior of the DFT is not obvious. This situation is covered in the following theorem. Here and in the following, $A \sim B$ denotes that the ratio of the real and the imaginary parts of the left and right-hand side converges to one as $T \rightarrow \infty$.

Theorem 1. *If* $\mu_t = h(t/T,T)$, *under Assumptions 1 we have*

$$I(\lambda_j) \sim \frac{T}{8\pi^3 j^2} \left\{ \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} \sin(2\pi js) ds \right]^2 + \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} (1 - \cos(2\pi js)) ds \right]^2 \right\}.$$

Similar to Lemmas 1 and 2, Theorem 1 is an asymptotic approximation, which is required to replace the sums from the DFT by integrals and the $\Delta\mu_k$ by $\partial h(s,T)/\partial s$, but no statistical limit theory is required, since the periodogram of a deterministic trend is itself a deterministic function.

Since the integrals in Theorem 1 are functions of j (and possibly T), it can be seen that the exact rate of the periodogram depends on the derivative of the trend function. Therefore, if the trend function is known and the integrals have a closed form solution, it is possible to determine the exact order. If this is not the case, we can still recover the upper bound on the rate of decay for increasing j that was established by Künsch (1986), Qu (2011), and Iacone (2010). To see this, note that $\sin(2\pi js) \leqslant 1$ and $1 - cos(2\pi js) \leqslant 2$ for all j and s. It therefore follows immediately for $\mu_t = h(s_t, T) = h(s_t)$ that the periodogram is $I(\lambda_j) = O(Tj^{-2})$.

5 The Periodogram of Processes with Abrupt Level Shifts

We now turn to the behavior of the periodogram of abrupt level-shift processes. To simplify the exposition, let ζ_k denote either $\Delta\mu_k$ or μ_k, depending on whether the accumulative structural-change model (5) or the non-accumulative model (6) is considered.

To characterize the behavior of different groups of processes, we require different groups of assumptions. First, in the case of deterministic structural breaks, we assume:

Assumption 2. *$|\zeta_k| < \infty$ and the $\delta_k = T_k/T$ are deterministic with $0 < \delta_k < 1$, for $k = 1,...,K < \infty$.*

For stochastic level shifts we require the following assumptions.

Assumption 3. *$E[\zeta_k] = 0$ and $Var[\zeta_k] = \sigma^2_\Delta T^{-\beta}$, for some $0 \leqslant \beta \leqslant 1$ and $0 < \sigma^2_\Delta < \infty$.*

Assumption 4. *$P(t \in \{T_1,...,T_K\}) = p_t$, where $0 \leqslant p_t \leqslant 1$, and $E[p_t] = \tilde{p}T^{-\alpha}$, for some $0 \leqslant \alpha \leqslant 1$. Furthermore, the dependence in p_t is limited such that $E[K] = \tilde{p}T^{1-\alpha}$, $E\left[((T_k - T_{k-1})/T)^2\right] = \frac{2\tilde{D}}{\tilde{p}^2}T^{2(\alpha-1)})$, and $E\left[((T_k - T_{k-1})/T)^4\right] = O(T^{4(\alpha-1)})$, for some $0 < \tilde{p}, \tilde{D} < \infty$.*

Assumption 5. *p_t is independent of ζ_k for all $k = 1,...,K$ and $t = 1,...,T$. Additionally, $\sum_{\tau=1}^{\infty} \left|E[\zeta_k\zeta_{k-\tau}]\right| = Var[\zeta_k]\tilde{C}$, for $k = 1,2,...,$ and $0 \leqslant \tilde{C} < \infty$.*

The rate $T^{-\beta}$ in Assumption 3 is required to nest a number of mean-change processes from the literature, such as the STOPBREAK process of Engle and Smith (1999). For other processes setting $\beta = 0$ gives the familiar setup with non-degenerate breaks.

Assumption 4 imposes structure on the nature of the mean change process. Assuming $P(t \in \{T_1,...,T_K\}) = p_t$ means that the regime change process is a generalized Bernoulli process with time-varying success probability p_t. The nature of the dependence in p_t is restricted by the additional requirement that the expected squared length of the kth regime expressed as a fraction of the sample is $\frac{2\tilde{D}}{\tilde{p}^2}T^{2(\alpha-1)}$, which means that the second moment of the regime lengths is still of the same order as that of a geometric distribution.

In this context, the constant $\tilde{D}$ depends on the dependence in p_t, and it is equal to one, if $p_t = p$ for all $t = 1, ..., T$.

Since there are T observations in the sample, the expected number of mean shifts in the series is $E[K] = \tilde{p}T^{1-\alpha}$. The parameter α controls the asymptotic frequency of level changes. The expected number of shifts remains constant for $\alpha = 1$, whereas it goes to infinity for $\alpha < 1$. The first case ($\alpha = 1$) is referred to as *rare shifts* asymptotics or *low frequency contaminations*. We refer to the second case ($\alpha < 1$) as *intermediate frequency contaminations*. Here, we have $K \to \infty$ but $K/T \to 0$, as $T \to \infty$. That means we asymptotically have an infinite number of shifts, but also an infinite number of observations between shifts. Finally, for $\alpha > 1$ shifts are so rare that we will no longer observe any in a sample, asymptotically. The empirically relevant parameter range is therefore $0 < \alpha \leqslant 1$.

Even though it may seem unusual to tie the properties of the process to the sample size, this is a common approach in the related literature. Guégan (2005) refers to this practice as a *thought experiment*. The validity of this approach depends on the purpose of the analysis. Obviously, it is unreasonable to assume that structural changes will become less common in the future if the objective is to forecast a time series. On the other hand, if the objective is statistical inference based on a given sample, we argue that assuming that the frequency of structural change is tied to the sample size T can be thought of as an asymptotic framework that is better suited to approximate the statistical properties of the quantities of interest than keeping p fixed. The latter would imply, for example, that level changes are so frequent that the mean between two shifts cannot be estimated consistently.

Finally, we require some bound on the degree of dependence between the means or mean changes ζ_k in consecutive segments. This is imposed by Assumption 5 according to which the autocovariance function of the ζ_k has to be absolutely summable. We then obtain the following result.

Theorem 2. *For $j/T \to 0$ and level-shift processes characterized by (5),*

i.) $I(\lambda_j) \sim \frac{T}{4\pi^3 j^2} O(1)$*, under Assumption 2.*

ii.) $I(\lambda_j) \sim \frac{\sigma_\Delta^2 \tilde{p} T^{1-\beta}}{4\pi^3 j^2} O_P(1)$*, for $\alpha = 1$, and under Assumptions 3 and 4.*

iii.) $E\left[I(\lambda_j)\right] \sim \frac{\sigma_\Delta^2 \tilde{p} T^{2-\alpha-\beta}}{4\pi^3 j^2} \left(1 + O(1)\tilde{C}\right)$*, for $\alpha < 1$, and under Assumptions 3, 4, and 5.*

Theorem 2 establishes the properties of the periodogram of the accumulative mean-change process in (5). The first case i.) derives the growth rate of the peak near the origin and the rate of decay for frequencies further away from the zero frequency for a deterministic structural break process. This order was previously established by McCloskey and Perron (2013). Interesting is the contrast to case ii.), where rare random level shifts are considered. In contrast to i.) the periodogram becomes stochastic instead of deterministic. Furthermore, the scaling factor $T^{-\beta}$ of the $\Delta\mu_k$ influences the scaling of

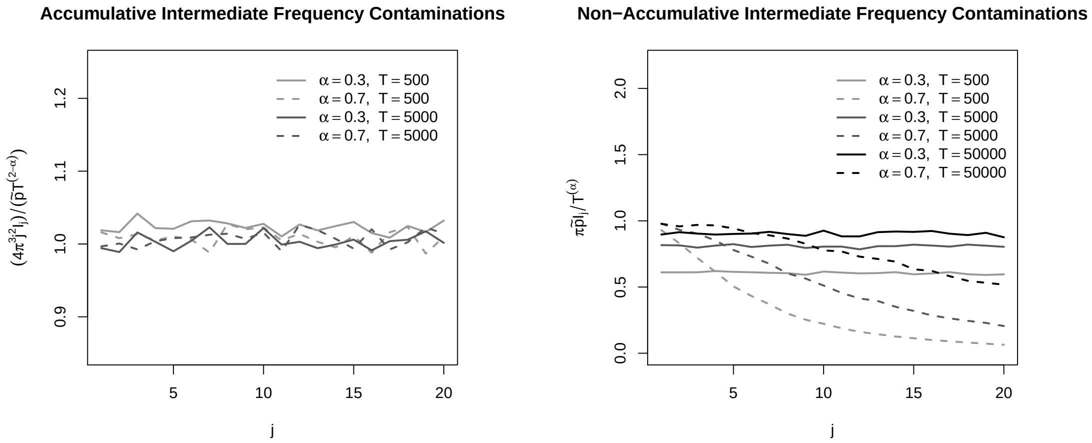

Figure 2: Average rescaled periodogram for accumulative and non-accumulative processes with intermediate frequency contaminations. In the left plot the simulated DGP is $\mu_t = \mu_{t-1} + \pi_t \varepsilon_t$ and in the right plot it is $\mu_t = (1-\pi_t)\mu_{t-1} + \pi_t \varepsilon_t$. In both cases $\pi_t \sim B(p)$, $p = 5/T^{\alpha}$, and $\varepsilon_t \sim N(0,1)$.

the peak local to the origin, which is of order $T^{1-\beta}$ instead of T. An important distinction between the first two cases and iii.) lies in the fact that iii.) makes a statement about the expectation of the periodogram. In i.) the periodogram is a deterministic function and in ii.) there is a well defined expectation, but the process is not ergodic since with $\alpha = 1$, the expected number of shifts in the sample is always given by $E[K] = \tilde{p}$. Case iii.), on the other hand, covers intermediate frequency contaminations, where $\alpha < 1$ so that the expected number of shifts is $E[K] = \tilde{p}T^{1-\alpha}$ and the process is ergodic. In this situation, the scaling of the peak near the origin is determined by both $-\alpha$ and β. Since $\alpha < 1$, the growth rate is always faster than that in cases i.) and ii.). The rate of decay for increasing j, however, is the same for all three types of processes.

Similar results to these can be obtained for the non-accumulative mean-change process in (6).

Theorem 3. *For $j/T \to 0$ and level-shift processes characterized by (6),*

i.) $I(\lambda_j) \sim \frac{T}{2\pi^3 j^2} O(1)$, *under Assumption 2.*

ii.) $I(\lambda_j) \sim \frac{\sigma_\Delta^2 \tilde{p} T^{1-\beta}}{2\pi^3 j^2} O_P(1)$, *for $\alpha = 1$, and under Assumptions 3 and 4.*

iii.) $E\left[I(\lambda_j)\right] \sim \frac{\sigma_\Delta^2 \tilde{D}}{\pi \tilde{p}} T^{\alpha-\beta}\left(1 + O(1)\tilde{C}\right)$, *for $\alpha < 1$, and under Assumptions 3, 4, and 5.*

As one can see, the orders for cases i.) and ii.) in Theorem 3 are identical to those in Theorem 2. That means that accumulative and non-accumulative structural change have the same impact on the periodogram local to zero, as long as the mean changes are deterministic or rare. In contrast to that, case iii.) is remarkably different. In presence of intermediate frequency contaminations, when the process becomes ergodic, the order of the peak is reduced to $T^{\alpha-\beta}$, instead of $T^{2-\alpha-\beta}$. Furthermore, the peak local to zero no longer decays for increasing j.

Important special cases of both the accumulative and the non-accumulative process are obtained for $\alpha = 0$. In this case the accumulative process boils down to a unit root process and the non-accumulative process becomes a simple stationary short memory process. In this situation, case iii.) in Theorems 2 and 3 reduces to the well known result that the periodogram of the unit root process local to the origin is of order $O_P(T^2/j^2)$ and that of the short memory process is $O_P(1)$.

The precision of the statements in case iii.) of Theorems 2 and 3 in finite samples is investigated in a small Monte Carlo study. The results are shown in Figure 2. As examples for accumulative and non-accumulative structural-change processes with intermediate frequency contaminations, we simulate the random level-shift process

$$\mu_t = \mu_{t-1} + \pi_t \varepsilon_t$$

and its stationary counterpart

$$\mu_t = (1 - \pi_t)\mu_{t-1} + \pi_t \varepsilon_t.$$

In both cases $\pi_t \sim B(p)$, $p = 5/T^{\alpha}$, and $\varepsilon_t \overset{iid}{\sim} N(0,1)$.

To analyze the accuracy of the asymptotic approximations in Theorems 2 and 3, we calculate the average periodogram from 5000 realizations of the respective process for different sample sizes and standardize it with the rate implied by the theorems. The resulting rescaled averaged periodogram is expected to be flat, if the theorem applies. Since the results are obtained under the assumption that $j/T \to 0$, it can be expected that the accuracy of the approximation is decreasing in j. On the left-hand side of Figure 2, it can be seen that the asymptotic approximation for accumulative structural-change processes in Theorem 2 iii.) is precise for small as well as for larger samples. On the other hand, the plot on the right-hand side of Figure 2 shows that the results in Theorem 3 for non-accumulative structural-change processes hold up well in finite samples for $\alpha = 0.3$, but the approximation seems to be much more imprecise for $\alpha = 0.7$.

The explanation for this effect can be seen from (22) in the proof of Theorem 3, where the first term of order $O(T^{\alpha-\beta})$ is the dominating one that drives the asymptotic result, but there is an approximation error of order $O(T^{-2+3\alpha-\beta})$, whose impact vanishes only slowly if α is relatively large.

Nevertheless, it can be seen that the average rescaled periodogram converges to its predicted value as T increases.

6 Examples and Relationship to Other Results

To illustrate the generality and usefulness of the results in Theorems 1 to 3, we discuss their implications for a number of important spurious long-memory processes. Some of these processes have already been studied in the literature. Where this is the case, we discuss how our results extend previous findings. Furthermore, we provide some more simulation evidence on the accuracy of our asymptotic approximations in finite samples.

6.1 Examples of Smooth Trends

First consider a linear trend of the form

$$h(s) = s. \tag{7}$$

Obviously, $\frac{\partial h(s)}{\partial s} = 1$, so that from Theorem 1

$$I(\lambda_j) \sim \frac{T}{8\pi^3 j^2}\left\{\left[\frac{\sin^2(\pi j)}{\pi j}\right]^2 + \left[1 - \frac{\sin(2\pi j)}{2\pi j}\right]^2\right\} = \frac{T}{8\pi^3 j^2}, \tag{8}$$

since $\sin(\pi j) = \sin(2\pi j) = 0$ for $j \in \mathbb{N}$. The linear trend is monotonous and generates a peak that is of the same order as that of discontinuous level changes such as deterministic structural breaks and random level-shift processes. The typical linear trend model $\mu_t = \beta_0 + \beta_1 t$ is obtained as a special case of the degenerate trend model for $h(t/T,T) = \beta_0 + \beta_1(t/T)T$. This leads to $I(\lambda_j) = (T^2\beta_1)/(8\pi^3 j^2)$, so that the periodogram is of a considerably larger magnitude than that in (8), for large T.

Non-monotonic trends can generate peaks that decay faster than those of monotonous trends as j increases. As an example, consider the quadratic trend

$$h(s) = 1/2 - (2s-1)^2, \tag{9}$$

with derivative $\frac{\partial h(s)}{\partial s} = 4 - 8s$. Then

$$\begin{aligned} I(\lambda_j) &\sim \frac{T}{8\pi^3 j^2}\left\{\left[\int_0^1 (4-8s)\sin(2\pi js)ds\right]^2 + \left[\int_0^1 (4-8s)(1-\cos(2\pi js))ds\right]^2\right\} \\ &= \frac{T}{8\pi^3 j^2}\left\{\left[\frac{4\cos(\pi j)(\pi j\cos(\pi j) - \sin(\pi j))}{\pi^2 j^2}\right]^2 + \left[\frac{2(\pi j\sin(2\pi j) + \cos(2\pi j) - 1)}{\pi^2 j^2}\right]^2\right\} \\ &= \frac{T}{8\pi^3 j^2}\left[\frac{4\cos^2(\pi j)}{\pi j}\right]^2 = \frac{2T}{\pi^5 j^4}, \end{aligned} \tag{10}$$

since $\sin(\pi j) = \sin(2\pi j) = 0$, $\cos(2\pi j) = 1$, and $\cos(\pi j) = -1$, for $j \in \mathbb{N}$. Obviously, in this case the number of Fourier frequencies that are meaningfully different from zero is much lower than in the monotonic case.

Both the linear trend as well as the quadratic trend are shown on the left-hand side of Figure 3. In analogy to Figure 2, the left-hand side of Figure 3 shows the accuracy of the result for different sample sizes T. It can be seen, that the accuracy decays slightly for increasing j if $T = 500$, but this is on a very small scale and the effect disappears as the sample size increases. We therefore find that the result is a reliable approximation in finite samples.

Simple sinusoidal trends do not cause spurious long memory if there is an integer

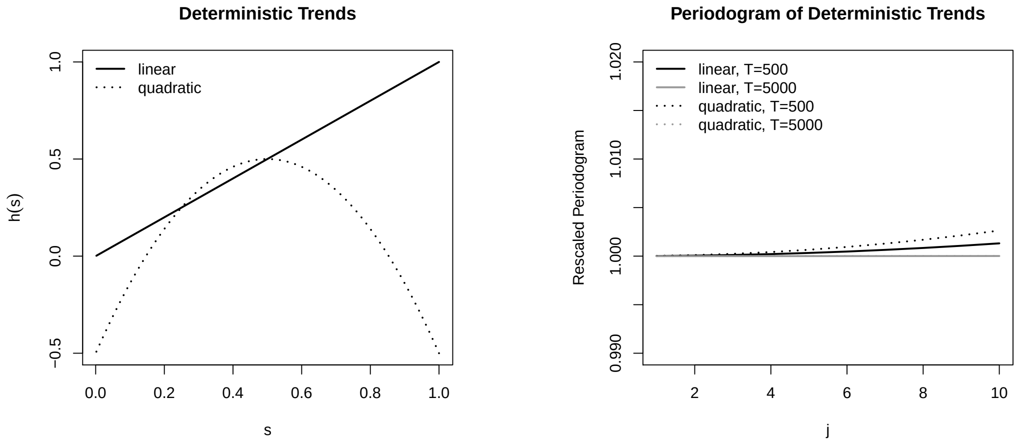

Figure 3: Linear and quadratic trend model as in (7) and (9) (left) and periodogram of the trends rescaled according to the results in (8) and (10) (right), at the j-th Fourier frequency.

number of cycles in the sample. As an example, we consider

$$h(s) = \sin(4\pi s),$$

as in the simulation study of Qu (2011). Then

$$\begin{aligned} I(\lambda_j) &\sim \frac{T}{8\pi^3 j^2} \left\{ \left[\int_0^1 4\pi \cos(4\pi s) \sin(2\pi js) ds \right]^2 + \left[\int_0^1 4\pi \cos(4\pi s)(1 - \cos(2\pi js)) ds \right]^2 \right\} \\ &= \frac{T}{8\pi j^2} \left\{ \left[\frac{\cos(2\pi j) - 1}{2\pi(j+1)} - \frac{\cos(2\pi j) - 1}{2\pi(j-1)} \right]^2 + \left[\frac{\sin(2\pi j)}{2\pi(j-1)} + \frac{\sin(2\pi j)}{2\pi(j+1)} \right]^2 \right\} = 0, \end{aligned}$$

for $j \neq 1$, since $\sin(2\pi j) = \cos(2\pi j) - 1 = 0$, for $j \in \mathbb{N}$, and by the product-to-sum identities. The periodic trend does only generate a single peak at the first Fourier frequency. The spectrum at all other frequencies local to zero is flat. Intuitively, this is obvious from the regression interpretation of the periodogram, since each periodogram ordinate can be interpreted as the sum of the squared coefficients on the sine and cosine term with the respective frequency in the Fourier approximation of the process. This implies exactly the behavior derived above, since these coefficients are zero for all but one frequency in the case of the sine trend.

The regression interpretation of the periodogram also leads to an intuitive explanation of the origin of spurious long memory. Since the periodogram ordinates are directly related to the Fourier coefficients, the impact of the low frequencies (other than zero) reflects the approximation error between the deterministic trend and the Fourier series.

A case that has received special attention in the literature is the fractional trend model of Bhattacharya et al. (1983) that is given by

$$h(s, T) = \frac{c_1}{(c_2 + sT)^\gamma}, \tag{11}$$

where $c_1, c_2 > 0$ and $\gamma \in [0, 1/2]$. For $\gamma \in [0, 1/2]$, Bhattacharya et al. (1983) show, that according to the rescaled range statistic $d = 1/2 - \gamma$. This monotonous trend model is degenerate in the sense that asymptotically all of the mean change of the process is concentrated in the immediate neighborhood of the sample origin.

It is obvious that $\frac{\partial h(s,T)}{\partial s} = -\frac{c_1 \gamma T}{(c_2 + sT)^{\gamma+1}}$. Therefore, we have from Theorem 1 that

$$I(\lambda_j) \sim \frac{T}{8\pi^3 j^2} \left\{ \left[\int_0^1 \frac{c_1 \gamma T}{(c_2 + sT)^{\gamma+1}} \sin(2\pi js) ds \right]^2 + \left[\int_0^1 \frac{c_1 \gamma T}{(c_2 + sT)^{\gamma+1}} (1 - \cos(2\pi js)) ds \right]^2 \right\}. \tag{12}$$

Unfortunately, the expressions in (12) are not readily integrable. However, as discussed in Section 4, we can still recover the result in Künsch (1986) and show that $\lim_{T\to\infty} I(\lambda_j) T^{-1} = O(j^{-2})$. If we set $\sin(2\pi js) < 1$ and $1 - \cos(2\pi js) < 2$, we obtain

$$\lim_{T\to\infty} \frac{I(\lambda_j)}{\gamma^2 T^3 \left[\int_0^1 (c_2 + sT)^{-\gamma-1} ds \right]^2} \leqslant \frac{5c_1^2}{8\pi^3 j^2}.$$

Furthermore, solving the integral in the denominator gives

$$\lim_{T\to\infty} \frac{I(\lambda_j)}{\gamma^2 T^3 \left[\frac{c_2^{-\gamma} - (c_2+T)^{-\gamma}}{T\gamma} \right]^2} = \lim_{T\to\infty} \frac{I(\lambda_j)}{T \left[c_2^{-\gamma} - (c_2 + T)^{-\gamma} \right]^2} = \lim_{T\to\infty} I(\lambda_j) T^{-1} c_2^{2\gamma},$$

so that

$$\lim_{T\to\infty} I(\lambda_j) T^{-1} \leqslant \frac{5c_1^2}{8\pi^3 j^2 c_2^{2\gamma}}.$$

Consequently, the effect of the fractional trend on the periodogram is at most as large as that of a random or deterministic level-shift process and therefore concentrated more heavily at the origin than that of a true long-memory process.

6.2 Examples of Abrupt Level Changes

Like Theorem 1, Theorems 2 and 3 cover many prominent special cases from the previous literature. As discussed above, the order of deterministic structural breaks is directly covered by part i.) of both theorems. Theorem 3 underlines that this result is not dependent on the exact formulation of the deterministic structural break process so that it applies to the non-accumulative case as well.

The non-stationary (rare) random level-shift process given by $\mu_t = \mu_{t-1} + \pi_t \varepsilon_t$, with $\varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$, $\pi_t \overset{iid}{\sim} B(1, p)$, and $p = O(T^{-1})$ as studied in Perron and Qu (2010), Diebold and Inoue (2001), and Qu (2011), among others, is directly covered by case ii.) of

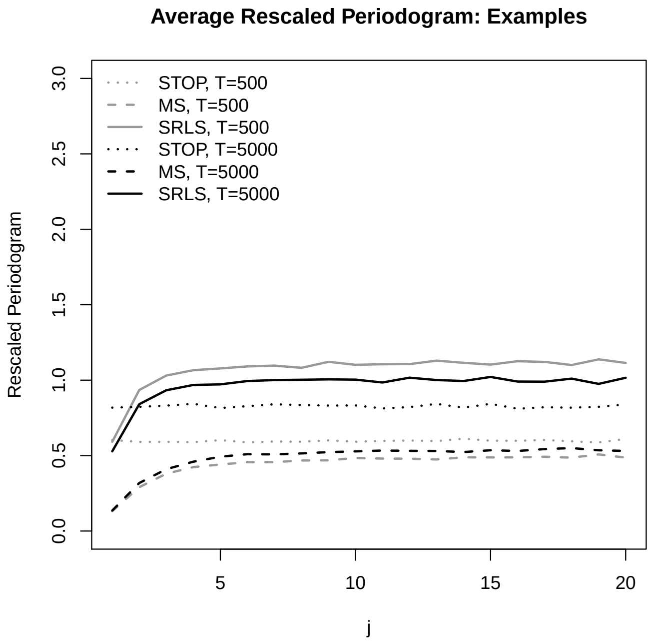

Figure 4: Average rescaled periodogram for examples from the spurious long-memory literature. STOP stands for the STOPBREAK process of Engle and Smith (1999), MS for the Markov-switching process as specified in Diebold and Inoue (2001), and SRLS stands for the stationary random level-shift process as considered in the simulation studies of Qu (2011).

Theorem 2, so that

$$I(\lambda_j) \sim \frac{O_P(T)}{4\pi^3 j^2}.$$

Different from Breidt and Hsu (2002), we obtain from Theorem 3 for the stationary random level-shift process given by $\mu_t = \mu_{t-1}(1-\pi_t) + \pi_t \varepsilon_t$ that it is of the same order as the non-stationary case for $p = \tilde{p}T^{-1}$, but it is $O_P\left(\frac{\sigma_\Delta^2}{\tilde{p}\pi} T^{\alpha-\beta}\right)$, for $p = O(T^{-\alpha})$ and $0 \leqslant \alpha < 1$.

Another example is the STOPBREAK process proposed by Engle and Smith (1999) that is given by $\mu_t = \mu_{t-1} + \frac{\varepsilon_{t-1}^2}{\gamma_T + \varepsilon_{t-1}^2} \varepsilon_{t-1}$. Here, the scaling factor γ_T determines the variance of the mean changes so that $\delta/2$ corresponds to β in Assumption 3.

For this process Diebold and Inoue (2001) show that $\gamma_T = O(T^\delta)$ implies $I(1-\delta)$-behavior of the partial sums. Assuming Gaussian innovations, we obtain

$$I(\lambda) \sim \frac{15\sigma_\varepsilon^6}{4\pi^3 j^2} T^{2(1-\delta)},$$

directly from case (ii.) of Theorem 2 with $\alpha = 0$.

Guégan (2005) mentions that for this process and all the other processes in Diebold and Inoue (2001) so far there has only been empirical evidence that they are not long memory in a spectral density sense. Our results now give a proof for this claim.

The accuracy of our results for these example processes is demonstrated in Figure 4 that shows the average rescaled periodograms for different sample sizes.

Finally, Markov-switching mean models have been considered as examples of spurious long-memory processes by Diebold and Inoue (2001) and Gourieroux and Jasiak (2001). Gourieroux and Jasiak (2001) focus on the autocovariance function of the Markov-switching models with rare shifts where $\alpha = 1$. Diebold and Inoue (2001) allow for more flexible dynamics and show for a process where the shift probability is of order $T^{-\alpha}$ that the variance of the partial sum process behaves like that of a process that is $I(\alpha/2)$. Obviously, the mean-change process in the Markov-switching model is non-accumulative so that the special cases of Gourieroux and Jasiak (2001) and Diebold and Inoue (2001) are covered by cases ii.) and iii.) in Theorem 3. For $\alpha = 1$, the Markov switching mean model therefore causes the same behavior of the periodogram local to the origin as stationary or non-stationary random level-shift processes with rare shifts. For $\alpha < 1$, the peak local to the origin is of order $O_p(T^{\alpha})$, but flat – and not decaying as j increases. Granger and Teräsvirta (1999) consider Markov-switching mean models with small but fixed switching probabilities. Technically, these processes fall into case (ii.) of Theorem 3, with $\alpha = 0$. However, in finite samples and when the probabilities for a regime change are very small, assuming $\alpha > 0$, or even $\alpha = 1$ might give a better approximation for the behavior of the periodogram.

7 Conclusion

We provide results on the behavior of a large class of spurious long-memory processes that nests a wide range of special cases previously discussed in the literature. In particular, we conduct the first comprehensive analysis of the different impact of low and intermediate frequency contaminations and accumulative and non-accumulative structural-change processes.

As can be seen from Theorems 1 to 3, with exception of case iii.) in Theorem 3, the impact of these processes on the periodogram is concentrated more heavily around the origin compared to true long-memory processes. This is the case as long as shifts are sufficiently rare or their effect is accumulative. While the peak generated by abrupt level changes decays with rate j^{-2} as the distance $2\pi j/T$ from the pole increases, the effect of trends can be even more local to the zero frequency, as can be seen for the example of the quadratic trend in the previous section, that is of order $O(j^{-4})$.

It is true that many of the spurious long-memory processes generate patterns such as non-zero autocorrelations at large lags or poles in the spectrum that are similar to true long-memory processes according to the definition in (1). However, we show that both types of processes have empirical features that are clearly distinct.

Furthermore, there are important differences in the effect of more or less frequent level changes. The rate of structural change modulates the effect of the sample size T on the scale of the peak in the periodogram and if structural changes are non-accumulative and of an intermediate frequency so that $0 < \alpha < 1$, then the effect on the periodogram is less localized to zero than that of true long-memory processes.

The differences between the behavior of accumulative rare level changes and true long-memory processes was previously derived by Perron and Qu (2010), among others, and it is the basis for many robust estimation procedures such as those of Iacone (2010), McCloskey and Perron (2013), Hou and Perron (2014), Christensen and Varneskov (2017), and McCloskey and Hill (2017). These methods use trimming or an augmented pseudo-spectral density that assumes the situation in case ii.) of Theorem 2. As our results show, the necessary amount of trimming, or the form of correction needed, depends on the structural-change model so that the domain on which these methods are valid could be extended.

Furthermore, these results show that the test for the null hypothesis of true long memory according to (1) proposed by Qu (2011) and extended by Sibbertsen et al. (2018) is applicable against the full range of structural-change processes nested in the class considered here.

We therefore believe that our results contribute to the development of a deeper understanding of the relationship between true and spurious long memory and to the applicability of robust methods for the estimation of the memory parameter in presence of structural change.

Appendix

Proof of Lemma 1

Proof. From (5), we have

$$\begin{aligned} w_\mu(\lambda) &= \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \mu_t e^{i\lambda t} \\ &= \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \left\{ \mu_0 + \sum_{k=1}^{K} \mathbb{I}(t \geqslant T_k) \Delta\mu_k \right\} e^{i\lambda t} \\ &= \frac{1}{\sqrt{2\pi T}} \left\{ \mu_0 \sum_{t=1}^{T} e^{i\lambda t} + \sum_{t=1}^{T} \sum_{k=1}^{K} \mathbb{I}(t \geqslant T_k) \Delta\mu_k e^{i\lambda t} \right\} \\ &= \frac{1}{\sqrt{2\pi T}} \left\{ \mu_0 \sum_{t=1}^{T} e^{i\lambda t} + \sum_{k=1}^{K} \Delta\mu_k \sum_{t=1}^{T} \mathbb{I}(t \geqslant T_k) e^{i\lambda t} \right\}. \end{aligned}$$

Here,

$$\sum_{t=1}^{T} \mathbb{I}(t \geqslant T_k) e^{i\lambda t} = \sum_{t=T_k}^{T} e^{i\lambda t} = \sum_{t=1}^{T} e^{i\lambda t} - \sum_{t=1}^{T_k} e^{i\lambda t} = D_T(\lambda) - D_{T_k}(\lambda).$$

Therefore,

$$\begin{aligned} w_\mu(\lambda) &= \frac{1}{\sqrt{2\pi T}} \left\{ D_T(\lambda)\mu_0 + \sum_{k=1}^{K} \Delta\mu_k \left[D_T(\lambda) - D_{T_k}(\lambda) \right] \right\} \\ &= \frac{1}{\sqrt{2\pi T}} \left\{ D_T(\lambda)\mu_0 + \sum_{k=1}^{K} \Delta\mu_k D_T(\lambda) - \sum_{k=1}^{K} \Delta\mu_k D_{T_k}(\lambda) \right\} \\ &= \frac{1}{\sqrt{2\pi T}} \left\{ \left[\mu_0 + \sum_{k=1}^{K} \Delta\mu_k \right] D_T(\lambda) - \sum_{k=1}^{K} \Delta\mu_k D_{T_k}(\lambda) \right\}. \end{aligned}$$

Furthermore, we have

$$D_T(\lambda_j) = \frac{e^{i(T+1)\lambda_j} - e^{i\lambda_j}}{e^{i\lambda_j} - 1} = e^{i(T-1)\lambda_j/2} \frac{\sin(T\lambda_j/2)}{\sin(\lambda_j/2)}, \tag{13}$$

cf. Beran et al. (2013), p. 327. Note that $\lambda_j T = 2\pi j$, $e^{i(T+1)\lambda_j} = e^{i\lambda_j T} e^{i\lambda_j}$, and $e^{i2\pi j} = \cos(2\pi j) + i\sin(2\pi j) = \cos(2\pi) + i\sin(2\pi) = 1$. Therefore, $D_T(\lambda_j) = \frac{e^{i\lambda_j} - e^{i\lambda_j}}{e^{i\lambda_j} - 1} = 0$, which proves the first part of the theorem.

Similarly, for the second part of the theorem, we have from (6) that

$$\begin{aligned} w_\mu(\lambda) &= \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \left\{ \mu_0 + \sum_{k=0}^{K} \mu_k \mathbb{I}(T_{k-1} \leqslant t < T_k) \right\} e^{i\lambda t} \\ &= \frac{1}{\sqrt{2\pi T}} \left\{ \mu_0 \sum_{t=1}^{T} e^{i\lambda t} + \sum_{k=0}^{K} \mu_k \sum_{t=1}^{T} \mathbb{I}(T_{k-1} \leqslant t < T_k) e^{i\lambda t} \right\}. \end{aligned}$$

Here,

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{I}(T_{k-1} \leqslant t < T_k) e^{i\lambda t} &= \sum_{t=1}^{T} \{\mathbb{I}(T_k > t) - \mathbb{I}(T_{k-1} \geqslant t)\} e^{i\lambda t} \\
&= \sum_{t=1}^{T} \mathbb{I}(T_k > t) e^{i\lambda t} - \sum_{t=1}^{T} \mathbb{I}(T_{k-1} \geqslant t) e^{i\lambda t} \\
&= \sum_{t=1}^{T_k - 1} e^{i\lambda t} - \sum_{t=1}^{T_{k-1}} e^{i\lambda t} \\
&= D_{T_k - 1}(\lambda) - D_{T_{k-1}}(\lambda).
\end{aligned}$$

Therefore, since $D_T(\lambda_j) = 0$, we have

$$w_\mu(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{k=0}^{K} \mu_k \left\{ D_{T_k - 1}(\lambda_j) - D_{T_{k-1}}(\lambda_j) \right\}.$$

□

Proof of Lemma 2

Proof. From the second expression in (13) in the proof of Lemma 1 we can decompose the real and the imaginary parts of the DFT at the Fourier frequencies $\lambda_j = 2\pi j/T$ as follows

$$\begin{aligned}
D_{T_k}(\lambda_j) &= \frac{e^{i(T_k - 1)\lambda_j/2} \sin(T_k \lambda_j/2)}{\sin(\lambda_j/2)} \\
&= \frac{\left[\cos((T_k - 1)\lambda_j/2) + i \sin((T_k - 1)\lambda_j/2)\right] \sin(T_k \lambda_j/2)}{\sin(\lambda_j/2)} \\
&= \frac{\cos((T_k - 1)\lambda_j/2) \sin(T_k \lambda_j/2)}{\sin(\lambda_j/2)} + i \frac{\sin((T_k - 1)\lambda_j/2) \sin(T_k \lambda_j/2)}{\sin(\lambda_j/2)} \\
&= \frac{\cos(\frac{T_k - 1}{T}\pi j) \sin(\frac{T_k}{T}\pi j)}{\sin(\frac{\pi j}{T})} + i \frac{\sin(\frac{T_k - 1}{T}\pi j) \sin(\frac{T_k}{T}\pi j)}{\sin(\frac{\pi j}{T})}.
\end{aligned}$$

It follows by the sum-to-product identities that

$$\begin{aligned}
D_{T_k}(\lambda_j) &= \frac{\sin(\frac{T_k - 1}{T}\pi j + \frac{T_k}{T}\pi j) - \sin(\frac{T_k - 1}{T}\pi j - \frac{T_k}{T}\pi j)}{2 \sin(\frac{\pi j}{T})} + i \frac{\cos(\frac{T_k - 1}{T}\pi j - \frac{T_k}{T}\pi j) - \cos(\frac{T_k - 1}{T}\pi j + \frac{T_k}{T}\pi j)}{2 \sin(\frac{\pi j}{T})} \\
&= \frac{\sin(2\delta_k \pi j - \frac{\pi j}{T}) + \sin(\frac{\pi j}{T})}{2 \sin(\frac{\pi j}{T})} + i \frac{\cos(\frac{\pi j}{T}) - \cos(2\delta_k \pi j - \frac{\pi j}{T})}{2 \sin(\frac{\pi j}{T})}.
\end{aligned}$$

By a Laurent series approximation around $\lambda_j = 0$, we obtain

$$\begin{aligned}
D_{T_k}(\lambda_j) = T j^{-1} &\frac{\sin(2\delta_k \pi j)}{2\pi} + \sin^2(\pi j \delta_k) + O_P(jT^{-1}) \\
&+ i \left[T j^{-1} \frac{\sin^2(\pi \delta_k j)}{\pi} - \frac{1}{2} \sin(2\pi \delta_k j) + O_P(jT^{-1}) \right],
\end{aligned}$$

where the Laurent series is obtained from separate Taylor approximations for each of the trigonometric functions. This proves the lemma. □

Proof of Theorem 1

Proof. For $\mu_t = h(t/T,T)$ by Lemma 1 we have $w_\mu(\lambda_j) = \frac{-1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\Delta\mu_t D_t(\lambda_j)$. Furthermore, from Lemma 2, we have

$$D_{T_k}(\lambda_j) = Tj^{-1}\frac{\sin(2\delta_k\pi j)}{2\pi} + \sin^2(\pi j\delta_k) + O_P(jT^{-1}) + i\left[Tj^{-1}\frac{\sin^2(\pi\delta_k j)}{\pi} - \frac{1}{2}\sin(2\pi\delta_k j) + O_P(jT^{-1})\right].$$

Therefore,

$$\begin{aligned}
I(\lambda_j) &= \left|-\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\Delta\mu_t D_t(\lambda_j)\right|^2 \\
&= (2\pi T)^{-1}\left|\sum_{t=1}^{T}\Delta\mu_t\left\{Tj^{-1}\frac{\sin(2\pi jt/T)}{2\pi} + \sin^2(\pi jt/T) + O_P(jT^{-1})\right.\right. \\
&\left.\left.+i\left[Tj^{-1}\frac{\sin^2(\pi jt/T)}{\pi} - 1/2\sin(2\pi jt/T) + O_P(jT^{-1})\right]\right\}\right|^2
\end{aligned}$$

$$\begin{aligned}
&= (2\pi T)^{-1}\left\{\left[\sum_{t=1}^{T}\Delta\mu_t\left(Tj^{-1}\frac{\sin(2\pi jt/T)}{2\pi} + \sin^2(\pi jt/T) + O_P(jT^{-1})\right)\right]^2\right. \\
&\left.+\left[\sum_{t=1}^{T}\Delta\mu_t\left(Tj^{-1}\frac{\sin^2(\pi jt/T)}{2\pi}\right) - 1/2\sin(2\pi jt/T) + O_P(jT^{-1})\right]^2\right\}
\end{aligned}$$

$$\begin{aligned}
&= (2\pi T)^{-1}\left\{\left[T\sum_{t=1}^{T}\frac{\Delta\mu_t\sin(2\pi jt/T)}{2\pi j} + \sum_{t=1}^{T}\Delta\mu_t\sin^2(\pi jt/T) + T^{-1}\sum_{t=1}^{T}\Delta\mu_t O_P(j)\right]^2\right. \\
&\left.+\left[T\sum_{t=1}^{T}\Delta\mu_t\frac{\sin^2(\pi jt/T)}{\pi j} - 1/2\sum_{t=1}^{T}\Delta\mu_t\sin(2\pi jt/T) + T^{-1}\sum_{t=1}^{T}\Delta\mu_t O_P(j)\right]^2\right\}
\end{aligned}$$

Factoring out T from the square brackets gives

$$\begin{aligned}
2\pi I(\lambda_j)T^{-1} = &\left\{\left[\sum_{t=1}^{T}\frac{\Delta\mu_t\sin(2\pi jt/T)}{2\pi j} + T^{-1}\sum_{t=1}^{T}\Delta\mu_t\sin^2(\pi jt/T) + T^{-2}\sum_{t=1}^{T}\Delta\mu_t O_P(j)\right]^2\right. \\
&\left.+\left[\sum_{t=1}^{T}\Delta\mu_t\frac{\sin^2(\pi jt/T)}{\pi j} - \frac{1}{2T}\sum_{t=1}^{T}\Delta\mu_t\sin(2\pi jt/T) + T^{-2}\sum_{t=1}^{T}\Delta\mu_t O_P(j)\right]^2\right\}.
\end{aligned}$$

Now, using $\Delta\mu_t = h(t/T,T) - h((t-1)/T,T)$, we have

$$\lim_{T\to\infty}\Delta\mu_t T = \frac{\partial h(t/T,T)}{\partial(t/T)},$$

so that

$$2\pi I(\lambda_j)T^{-1} \sim \left\{ \left[\frac{1}{2\pi jT} \sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} \sin(2\pi jt/T) + \sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} \frac{1}{T^2} \sin^2(\pi jt/T) + \frac{1}{T^3} \sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} O_P(j) \right]^2 \right.$$
$$\left. + \left[\sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} \frac{\sin^2(\pi jt/T)}{\pi jT} - \frac{1}{2} \sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} \frac{1}{T^2} \sin(2\pi jt/T) + \frac{1}{T^3} \sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} O_P(j) \right]^2 \right\},$$

where $a \sim b$ means that the ratio of a and b converge to 1, as $T \to \infty$. Here and in the following this shorthand notation is used to improve the readability of the proof.
By the definition of a Riemann integral

$$2\pi I(\lambda_j)T^{-1} \sim \left\{ \left[\frac{1}{2\pi j} \int_0^1 \frac{\partial h(s,T)}{\partial s} \sin(2\pi js)ds + \frac{1}{T} \int_0^1 \frac{\partial h(s,T)}{\partial s} \sin^2(\pi js)ds + T^{-2} \int_0^1 \frac{\partial h(s,T)}{\partial s} O_P(j)ds \right]^2 \right.$$
$$\left. + \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} \frac{\sin^2(\pi js)}{\pi j} ds - \frac{1}{2T} \int_0^1 \frac{\partial h(s,T)}{\partial s} \sin(2\pi js)ds + T^{-2} \int_0^1 \frac{\partial h(s,T)}{\partial s} O_P(j)ds \right]^2 \right\}.$$

Clearly, both parts of this expression are dominated by the first term in the respective square bracket, such that

$$I(\lambda_j) \sim \frac{T}{8\pi^3 j^2} \left\{ \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} \sin(2\pi js)ds \right]^2 + \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} (1 - \cos(2\pi js))ds \right]^2 \right\},$$

which finishes our proof. □

Proof of Theorem 2

Proof. First, by (13) in the proof of Lemma 1 we have

$$\begin{aligned}
D_{T_k}(\lambda_j)D^*_{T_l}(\lambda_j) &= e^{i(T_k-1)\lambda_j/2} \frac{\sin(T_k\lambda_j/2)}{\sin(\lambda_j/2)} e^{-i(T_l-1)\lambda_j/2} \frac{\sin(T_l\lambda_j/2)}{\sin(\lambda_j/2)} \\
&= e^{i(T_k-T_l)\lambda_j/2} \frac{\sin(T_k\lambda_j/2)\sin(T_l\lambda_j/2)}{\sin^2(\lambda_j/2)} \\
&= 2\frac{e^{i\lambda_j(T_k-T_l)/2}\sin(T_k\lambda_j/2)\sin(T_l\lambda_j/2)}{1-\cos(\lambda_j)} \\
&= 2e^{i\pi j(\delta_k-\delta_l)} \frac{\sin(\delta_k\pi j)\sin(\delta_l\pi j)}{1-\cos(\lambda_j)}.
\end{aligned}$$

By a Laurent expansion around $\lambda_j = 0$, we have

$$\begin{aligned}
D_{T_k}(\lambda_j)D^*_{T_l}(\lambda_j) &= 2e^{i\pi j(\delta_k-\delta_l)} \frac{\sin(\delta_k\pi j)\sin(\delta_l\pi j)}{1-[1-2\pi^2(j/T)^2+O((j/T)^4)]} \\
&= 2e^{i\pi j(\delta_k-\delta_l)} \frac{\sin(\delta_k\pi j)\sin(\delta_l\pi j)}{2\pi^2(j/T)^2+O((j/T)^4)} \\
&= \frac{T^2}{\pi^2 j^2} e^{i\pi j(\delta_k-\delta_l)} \sin(\delta_k\pi j)\sin(\delta_l\pi j) + O_P(1). \qquad (14)
\end{aligned}$$

In particular, for the case when $T_k = T_l = t$, we obtain

$$D_t(\lambda_j)D_t^*(\lambda_j) = \frac{T^2}{2\pi^2 j^2}(1-\cos(2\pi t j/T)) + O(1). \tag{15}$$

Furthermore, we have

$$\sum_{k=1}^{K} \Delta\mu_k D_{T_k}(\lambda_j) = \sum_{t=1}^{T} \Delta\mu_t D_t(\lambda_j), \quad \text{where} \quad \Delta\mu_t = \begin{cases} \Delta\mu_k, & \text{if } t = T_k \\ 0, & \text{otherwise.} \end{cases}$$

In addition to that,

$$\begin{aligned} I(\lambda) &= \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T} \Delta\mu_t \Delta\mu_s D_t(\lambda) D_s^*(\lambda) \\ &= \frac{1}{2\pi T}\sum_{t=1}^{T}(\Delta\mu_t)^2 D_t(\lambda)D_t^*(\lambda) + \frac{1}{2\pi T}\sum_{t\neq s}\Delta\mu_t\Delta\mu_s D_t(\lambda)D_s^*(\lambda) \\ &= A + B. \end{aligned}$$

Consequently, we have for term A and from (15) above

$$\begin{aligned} A &= \frac{1}{2\pi T}\sum_{t=1}^{T}(\Delta\mu_t)^2 D_t(\lambda_j)D_t^*(\lambda_j) \\ &= \frac{T}{4\pi^3 j^2}\sum_{t=1}^{T}(\Delta\mu_t)^2(1-\cos(2\pi jt/T)) + \frac{O(1)}{2\pi T}\sum_{t=1}^{T}(\Delta\mu_t)^2 \\ &= \frac{T}{4\pi^3 j^2}\left\{\sum_{t=1}^{T}(\Delta\mu_t)^2 - \sum_{t=1}^{T}(\Delta\mu_t)^2\cos(2\pi jt/T)\right\} + O_P(1). &(16) \\ &= \frac{T}{4\pi^3 j^2}\left\{\sum_{k=1}^{K}(\Delta\mu_k)^2 - \sum_{k=1}^{K}(\Delta\mu_k)^2\cos(2\pi j\delta_k)\right\} + O_P(1). &(17) \end{aligned}$$

To deal with term B, we revert back to the original representation in which the sum is random and write

$$\begin{aligned} B &= \frac{1}{2\pi T}\sum_{t\neq s}\Delta\mu_t\Delta\mu_s D_t(\lambda_j)D_s^*(\lambda_j) \\ &= \frac{1}{2\pi T}\sum_{k\neq l}\Delta\mu_k\Delta\mu_l D_{T_k}(\lambda_j)D_{T_l}^*(\lambda_j), \end{aligned}$$

where $k,l = 1,\ldots,K$. Similar to the approach above, we have from (14)

$$B = \frac{T}{2\pi^3 j^2}\sum_{k\neq l}\Delta\mu_k\Delta\mu_l e^{i\pi j(\delta_k-\delta_l)}\sin(\delta_k\pi j)\sin(\delta_l\pi j) + \frac{1}{2\pi T}\sum_{k\neq l}\Delta\mu_k\Delta\mu_l O_P(1) \tag{18}$$

The first part of the theorem i.) follows immediately from (17) and (18). Similarly, ii.) is also a direct consequence of (16) and (18), since for $\alpha = 1$, from Assumption 4 we have $E[K] = \tilde{p}$, so that the process is not ergodic and the sums in (17) and (18) are finite in probability.

For the third part of the theorem, from Assumption 4 we have $E[K] = E[p_t T] = \tilde{p}T^{1-\alpha}$ and, from

(16) and (17)

$$
\begin{aligned}
E[A] &= \frac{T}{4\pi^3 j^2}\left\{E\left[\sum_{k=1}^{K}(\Delta\mu_k)^2\right] - E[\Delta\mu_t^2]\sum_{t=1}^{T}\cos(2\pi jt/T)\right\} + O(1) \\
&= \frac{T}{4\pi^3 j^2}E\left[\sum_{k=1}^{K}(\Delta\mu_k)^2\right] + O(1),
\end{aligned}
$$

since $\sum_{t=1}^{T}\cos(2\pi jt/T) = 0$.

Now, from Assumption 3 we have $E[(\Delta\mu_k)^2] = \sigma_\Delta^2 T^{-\beta}$, so that by the generalized Wald identity of Brown (1974) and Assumption 4

$$
E[A] = \frac{T}{4\pi^3 j^2}E[K]E[(\Delta\mu_k)^2] = \frac{\tilde{p}\sigma_\Delta^2 T^{2-\alpha-\beta}}{4\pi^3 j^2} + o(1).
$$

Similarly, from (18)

$$
E[B] = \frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq l}\Delta\mu_k\Delta\mu_l e^{i\pi j(\delta_k-\delta_l)}\sin(\delta_k\pi j)\sin(\delta_l\pi j)\right] + \frac{O(1)}{2\pi T}E\left[\sum_{k\neq l}\Delta\mu_k\Delta\mu_l\right]
$$

and by the generalized Wald identity of Brown (1974) in conjunction with Assumption 5

$$
E[B] = \frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq l}E\left[\Delta\mu_k\Delta\mu_l\right]E\left[e^{i\pi j(\delta_k-\delta_l)}\sin(\delta_k\pi j)\sin(\delta_l\pi j)\right]\right] + \frac{O(1)}{2\pi T}E\left[\sum_{k\neq l}E\left[\Delta\mu_k\Delta\mu_l\right]\right].
$$

Therefore,

$$
\begin{aligned}
\left|E[B]\right| &\leqslant \frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq l}\left|E\left[\Delta\mu_k\Delta\mu_l\right]\right|\left|E\left[e^{i\pi j(\delta_k-\delta_l)}\sin(\delta_k\pi j)\sin(\delta_l\pi j)\right]\right|\right] + \left|\frac{O(1)}{2\pi T}E\left[\sum_{k\neq l}E\left[\Delta\mu_k\Delta\mu_l\right]\right]\right| \\
&\leqslant \frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq l}\left|E\left[\Delta\mu_k\Delta\mu_l\right]\right|\right] + \frac{|O(1)|}{2\pi T}E\left[\sum_{k\neq l}\left|E\left[\Delta\mu_k\Delta\mu_l\right]\right|\right].
\end{aligned}
$$

Assumption 5 combined with Assumptions 3 and 4 implies that

$$
\begin{aligned}
E\left[\sum_{k\neq l}\left|E[\Delta\mu_k\Delta\mu_l]\right|\right] &= 2E\left[\sum_{k=2}^{K}\sum_{\tau=1}^{k-1}\left|E[\Delta\mu_k\Delta\mu_{k-\tau}]\right|\right] \\
&\leqslant 2E\left[K\right]Var(\Delta\mu_k)\tilde{C} \\
&= 2\tilde{p}\tilde{C}\sigma_\Delta^2 T^{1-\alpha-\beta},
\end{aligned}
$$

so that

$$
\left|E[B]\right| \leqslant \frac{\tilde{p}\sigma_\Delta^2\tilde{C}}{\pi^3 j^2}T^{2-\alpha-\beta} + \left|O(T^{-\alpha-\beta})\right|.
$$

□

Proof of Theorem 3

Proof. From $w_\mu(\lambda) = -\frac{1}{\sqrt{2\pi T}}\sum_{k=0}^{K}\mu_k[D_{T_k-1}(\lambda) - D_{T_{k-1}}(\lambda)]$, as shown in Lemma 1, we have

$$\begin{aligned}
I(\lambda) &= \frac{1}{2\pi T}\sum_{k=0}^{K}\sum_{l=0}^{K}\mu_k\mu_l[D_{T_k-1}(\lambda) - D_{T_{k-1}}(\lambda)][D^*_{T_l-1}(\lambda) - D^*_{T_{l-1}}(\lambda)] \\
&= \frac{1}{2\pi T}\sum_{k=0}^{K}\sum_{l=0}^{K}\mu_k\mu_l[D_{T_k-1}(\lambda)D^*_{T_l-1}(\lambda) - D_{T_k-1}(\lambda)D^*_{T_{l-1}}(\lambda) - D_{T_{k-1}}(\lambda)D^*_{T_l-1}(\lambda) + D_{T_{k-1}}(\lambda)D^*_{T_{l-1}}(\lambda)] \\
&= \frac{1}{2\pi T}\sum_{k=0}^{K}\mu_k^2[D_{T_k-1}(\lambda)D^*_{T_k-1}(\lambda) - D_{T_k-1}(\lambda)D^*_{T_{k-1}}(\lambda) - D_{T_{k-1}}(\lambda)D^*_{T_k-1}(\lambda) + D_{T_{k-1}}(\lambda)D^*_{T_{k-1}}(\lambda)] \\
&+ \frac{1}{2\pi T}\sum_{k\neq l}\mu_k\mu_l[D_{T_k-1}(\lambda)D^*_{T_l-1}(\lambda) - D_{T_k-1}(\lambda)D^*_{T_{l-1}}(\lambda) - D_{T_{k-1}}(\lambda)D^*_{T_l-1}(\lambda) + D_{T_{k-1}}(\lambda)D^*_{T_{l-1}}(\lambda)] \\
&= \tilde{A} + \tilde{B}. \qquad (19)
\end{aligned}$$

Denoting $(T_k - 1)/T = \tilde{\delta}_k$, we have from (14) for the term in square brackets in $\tilde{A}$

$$\begin{aligned}
\tilde{a}_k =& [D_{T_k-1}(\lambda)D^*_{T_k-1}(\lambda) - D_{T_k-1}(\lambda)D^*_{T_{k-1}}(\lambda) - D_{T_{k-1}}(\lambda)D^*_{T_k-1}(\lambda) + D_{T_{k-1}}(\lambda)D^*_{T_{k-1}}(\lambda)] \\
=& \frac{T^2}{\pi^2 j^2}[\sin^2(\tilde{\delta}_k\pi j) + \sin^2(\delta_{k-1}\pi j) - e^{i\pi j(\tilde{\delta}_k - \delta_{k-1})}\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j) - e^{i\pi j(\delta_{k-1} - \tilde{\delta}_k)}\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)] + O_P(1) \\
=& \frac{T^2}{\pi^2 j^2}[1 - \frac{1}{2}\cos(2\tilde{\delta}_k\pi j) - \frac{1}{2}\cos(2\delta_{k-1}\pi j) - 2\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))] + O_P(1) \\
=& \frac{T^2}{\pi^2 j^2}[1 - \frac{1}{2}\left[\cos(2\tilde{\delta}_k\pi j) + \cos(2\delta_{k-1}\pi j)\right] - 2\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))] + O_P(1),
\end{aligned}$$

from Euler's formula. As in the proof of Theorem 1, the notation $A \sim B$ is used as a shorthand for $\lim_{T\to\infty} A/B = 1$. By the sum-to-product identity for the cosine, it follows

$$\begin{aligned}
\tilde{a}_k &= \frac{T^2}{\pi^2 j^2}\left[1 - \frac{1}{2}\left[2\cos(\pi j(\tilde{\delta}_k + \delta_{k-1}))\cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))\right] - 2\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))\right] + O_P(1) \\
&= \frac{T^2}{\pi^2 j^2}\left[1 - \cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))\left[\cos(\pi j(\tilde{\delta}_k + \delta_{k-1})) + 2\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\right]\right] + O_P(1).
\end{aligned}$$

Now, by the product-to-sum identity of the sine

$$\begin{aligned}
\tilde{a}_k &= \frac{T^2}{\pi^2 j^2}\left[1 - \cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))\left[\cos(\pi j(\tilde{\delta}_k + \delta_{k-1})) + \cos(\pi j(\tilde{\delta}_k - \delta_{k-1})) - \cos(\pi j(\tilde{\delta}_k + \delta_{k-1}))\right]\right] + O_P(1) \\
&= \frac{T^2}{\pi^2 j^2}\left[1 - \cos^2(\pi j(\tilde{\delta}_k - \delta_{k-1}))\right] + O_P(1).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\tilde{A} &= \frac{1}{2\pi T}\sum_{k=0}^{K}\mu_k^2\tilde{a}_k \\
&= \frac{T}{2\pi^3 j^2}\sum_{k=0}^{K}\mu_k^2[1 - \cos^2(\pi j(\tilde{\delta}_k - \delta_{k-1}))] + \frac{(K+1)O_P(1)}{2\pi T} \qquad (20)
\end{aligned}$$

For $\alpha < 1$, by a Taylor expansion of the squared cosine at zero

$$\begin{aligned}\tilde{a}_k &= \frac{T^2}{\pi^2 j^2}\left[1-\left[1-\pi^2 j^2(\tilde{\delta}_k-\delta_{k-1})^2+\pi^4 j^4 O((\tilde{\delta}_k-\delta_{k-1})^4)\right]\right]+O_P(1)\\ &= T^2\left[(\tilde{\delta}_k-\delta_{k-1})^2-\pi^2 j^2 O((\tilde{\delta}_k-\delta_{k-1})^4)\right]+O_P(1).\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}\tilde{A} &= \frac{1}{2\pi T}\sum_{k=0}^{K}\mu_k^2\left\{T^2\left[(\tilde{\delta}_k-\delta_{k-1})^2-\pi^2 j^2 O((\tilde{\delta}_k-\delta_{k-1})^4)\right]+O_P(1)\right\}\\ &= \frac{T}{2\pi}\sum_{k=0}^{K}\left\{\mu_k^2(\tilde{\delta}_k-\delta_{k-1})^2\right\}-\frac{T\pi j^2}{4}\sum_{k=0}^{K}\mu_k^2 O\left((\tilde{\delta}_k-\delta_{k-1})^4\right)+\frac{O_P(1)}{2\pi T}\sum_{k=0}^{K}\mu_k^2. \qquad (21)\end{aligned}$$

By applying the Wald identity for dependent random sums of Brown (1974) and then using Assumption 5, we obtain

$$\begin{aligned}E[\tilde{A}] &= E[K+1]E[\mu_k^2]\left\{\frac{T}{2\pi}E[(\tilde{\delta}_k-\delta_{k-1})^2]-\frac{T\pi j^2}{4}E[O((\tilde{\delta}_k-\delta_{k-1})^4)]+\frac{O(1)}{2\pi T}\right\}\\ &= (\tilde{p}T^{1-\alpha}+1)\sigma_\Delta^2 T^{-\beta}\left\{\frac{T}{\pi}\frac{\tilde{D}}{\tilde{p}^2}T^{2(\alpha-1)}-\frac{T\pi j^2}{2}O(T^{4(\alpha-1)})+O(T^{-1})\right\}\\ &= \frac{\sigma_\Delta^2\tilde{D}}{\pi\tilde{p}}T^{\alpha-\beta}-\frac{\sigma_\Delta^2\tilde{p}\pi j^2}{2}O(T^{-2+3\alpha-\beta})+\sigma_\Delta^2\tilde{p}O(T^{-\alpha-\beta})\\ &+\frac{\sigma_\Delta^2\tilde{D}}{\pi\tilde{p}}T^{-1+2\alpha-\beta}-\pi j^2 O(T^{-3+4\alpha-\beta})+O(T^{-1-\beta}). \qquad (22)\end{aligned}$$

For $\tilde{B}$, we have from (14),

$$\begin{aligned}\tilde{B} &= \frac{T}{2\pi^2 j^2}\sum_{k\neq l}\mu_k\mu_l\left[e^{i\pi j(\delta_k-\delta_l)}\sin(\delta_k\pi j)\sin(\delta_l\pi j)-e^{i\pi j(\delta_k-\delta_{l-1})}\sin(\delta_k\pi j)\sin(\delta_{l-1}\pi j)\right.\\ &\left.-e^{i\pi j(\delta_{k-1}-\delta_l)}\sin(\delta_{k-1}\pi j)\sin(\delta_l\pi j)+e^{i\pi j(\delta_{k-1}-\delta_{l-1})}\sin(\delta_{k-1}\pi j)\sin(\delta_{l-1}\pi j)\right]\\ &+\frac{O_P(1)}{2\pi T}\sum_{k\neq l}\mu_k\mu_l. \qquad (23)\end{aligned}$$

Denote the term in the square bracket by $\tilde{b}$, and let $\tilde{b}_1$ denote the first two summands and $\tilde{b}_2$ the last two summands, so that $\tilde{b}=\tilde{b}_1+\tilde{b}_2$. We have

$$\begin{aligned}\tilde{b}_1 &= e^{i\pi j(\delta_k-\delta_l)}\sin(\delta_k\pi j)\sin(\delta_l\pi j)-e^{i\pi j(\delta_k-\delta_{l-1})}\sin(\delta_k\pi j)\sin(\delta_{l-1}\pi j)\\ &= \sin(\delta_k\pi j)\left[e^{i\pi j(\delta_k-\delta_l)}\sin(\delta_l\pi j)-e^{i\pi j(\delta_k-\delta_{l-1})}\sin(\delta_{l-1}\pi j)\right]\\ &= \sin(\delta_k\pi j)\left[\cos(\pi j(\delta_k-\delta_l))\sin(\delta_l\pi j)-\cos(\pi j(\delta_k-\delta_{l-1}))\sin(\delta_{l-1}\pi j)\right.\\ &\left.+i\left\{\sin(\pi j(\delta_k-\delta_l))\sin(\pi j\delta_l)-\sin(\pi j(\delta_k-\delta_{l-1}))\sin(\pi j\delta_{l-1})\right\}\right].\end{aligned}$$

Now, let $\gamma_l=\delta_l-\delta_{l-1}$. Then, by a Taylor approximation at $\gamma_l=0$

$$\begin{aligned}\tilde{b}_1 &= \sin(\delta_k\pi j)\left[-\pi j\gamma_l\cos(\pi j(\delta_k-2\delta_l))+O_P(\gamma_l^2)+i\left\{-\pi j\gamma_l\sin(\pi j(\delta_k-2\delta_l))+O_P(\gamma_l^2)\right\}\right]\\ &= \pi j\gamma_l\sin(\delta_k\pi j)e^{i\pi j(\delta_k-2\delta_l)}+O_P(\gamma_l^2).\end{aligned}$$

Similarly, we have for the third and fourth term in the square bracket

$$\begin{aligned}\tilde{b}_2 &= -e^{i\pi j(\delta_{k-1}-\delta_l)}\sin(\delta_{k-1}\pi j)\sin(\delta_l\pi j)+e^{i\pi j(\delta_{k-1}-\delta_{l-1})}\sin(\delta_{k-1}\pi j)\sin(\delta_{l-1}\pi j)\\ &= -\sin(\delta_{k-1}\pi j)\left[e^{i\pi j(\delta_{k-1}-\delta_l)}\sin(\delta_l\pi j)-e^{ipij(\delta_{k-1}-\delta_{l-1})}\sin(\delta_{l-1}\pi j)\right],\end{aligned}$$

and by a Taylor approximation at $\gamma_l = 0$,

$$\tilde{b}_2 = -\pi j\gamma_l\sin(\delta_{k-1}\pi j)e^{i\pi j(\delta_{k-1}-2\delta_l)}+O_P(\gamma_l^2).$$

Therefore, we have

$$\tilde{b} = -\pi j\gamma_l\left[\sin(\delta_k\pi j)e^{i\pi j(\delta_k-2\delta_l)}-\sin(\delta_{k-1}\pi j)e^{i\pi j(\delta_{k-1}-2\delta_l)}\right].$$

Defining $\gamma_k = \delta_k - \delta_{k-1}$, and approximating at $\gamma_k = 0$, we obtain

$$\tilde{b} = \pi^2 j^2\gamma_l\gamma_k e^{2i\pi j(\delta_k-\delta_l)}+O_P(\gamma_l^2)+O_P(\gamma_k^2),$$

so that

$$\begin{aligned}\tilde{B} &= \frac{T}{2\pi^2 j^2}\sum_{k\neq l}\mu_k\mu_l\left[\pi^2 j^2\gamma_l\gamma_k e^{2i\pi j(\delta_k-\delta_l)}+O_P(\gamma_l^2)+O_P(\gamma_k^2)\right]+\frac{O_P(1)}{2\pi T}\sum_{k\neq l}\mu_k\mu_l\\ &= \frac{T}{2}\sum_{k\neq l}\mu_k\mu_l\gamma_l\gamma_k e^{2i\pi j(\delta_k-\delta_l)}+\frac{T}{2\pi^2 j^2}\sum_{k\neq l}\mu_k\mu_l O_P(\gamma_l^2)+\frac{T}{2\pi^2 j^2}\sum_{k\neq l}\mu_k\mu_l O_P(\gamma_k^2)+\frac{O_P(1)}{2\pi T}\sum_{k\neq l}\mu_k\mu_l. \quad (24)\end{aligned}$$

Similar to the proof of Theorem 2, we have from the Wald identity of Brown (1974) and Assumption 5

$$\begin{aligned}E[\tilde{B}] &= \frac{T}{2}E\left[\sum_{k\neq l}E\left[\mu_k\mu_l\right]E\left[\gamma_l\gamma_k e^{2i\pi j(\delta_k-\delta_l)}\right]\right]+\frac{T}{2\pi^2 j^2}E\left[\sum_{k\neq l}E\left[\mu_k\mu_l\right]E[O_P(\gamma_k^2)]\right]+\frac{O(1)}{2\pi T}E\left[\sum_{k\neq l}E[\mu_k\mu_l]\right]\\ &= \tilde{B}_1+\tilde{B}_2+\tilde{B}_3.\end{aligned}$$

For the first term

$$\begin{aligned}\left|E[\tilde{B}_1]\right| &\leqslant \frac{T}{2}E\left[\sum_{k\neq l}\left|E\left[\mu_k\mu_l\right]E\left[\gamma_l\gamma_k e^{2i\pi j(\delta_k-\delta_l)}\right]\right|\right]\\ &\leqslant \frac{T}{2}E\left[\sum_{k\neq l}\left|E\left[\mu_k\mu_l\right]\right|\left|E\left[\gamma_l\gamma_k e^{2i\pi j(\delta_k-\delta_l)}\right]\right|\right]\\ &\leqslant \frac{T}{2}E\left[\sum_{k\neq l}\left|E\left[\mu_k\mu_l\right]\right|\left|E\left[\gamma_l\gamma_k\right]\right|\right]\\ &= TE\left[\sum_{k=1}^{K}\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]\right|\left|E\left[\gamma_k\gamma_{k-\tau}\right]\right|\right]\\ &\leqslant TE\left[\sum_{k=1}^{K}\left|E\left[\gamma_k^2\right]\right|\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]\right|\right].\end{aligned}$$

Therefore, by Assumptions 4 and 5

$$
\begin{aligned}
\left|E[\tilde{B}_1]\right| &\leqslant TE\left[\sum_{k=1}^{K}\left|E\left[\gamma_k^2\right]\right| Var(\mu_k)\tilde{C}\right] \\
&= TE[K]E\left[\gamma_k^2\right] Var(\mu_k)\tilde{C} \\
&= \tilde{C}\sigma_\Delta^2 T^{1-\beta}\left\{\frac{2\tilde{D}}{\tilde{p}^2}T^{2(\alpha-1)}+O(T^{\alpha-2})\right\}E[K] \\
&= \frac{2\tilde{C}\tilde{D}\sigma_\Delta^2}{\tilde{p}}T^{\alpha-\beta}+\tilde{C}\left[O(T^{2\alpha-1-\beta})+O(T^{-\beta})+O(T^{\alpha-\beta-1})\right]. \qquad (25)
\end{aligned}
$$

Similarly, for the second term

$$
\begin{aligned}
\left|E[\tilde{B}_2]\right| &\leqslant \frac{T}{\pi^2 j^2}E\left[\sum_{k\neq l}\left|E\left[\mu_k\mu_l\right]E\left[O_P(\gamma_k^2)\right]\right|\right] \\
&= \frac{2T}{\pi^2 j^2}E\left[\sum_{k=1}^{K}\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]E\left[O_P(\gamma_k^2)\right]\right|\right] \\
&\leqslant \frac{2T}{\pi^2 j^2}E\left[\sum_{k=1}^{K}E\left[O_P(\gamma_k^2)\right]\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]\right|\right] \\
&\leqslant \frac{2\tilde{C}\sigma_\Delta^2}{\pi^2 j^2}T^{1-\beta}E\left[\sum_{k=1}^{K}O_P(\gamma_k^2)\right] \\
&= \frac{2\tilde{C}\sigma_\Delta^2\tilde{p}}{\pi^2 j^2}T^{2-\alpha-\beta}\left[O(T^{2(\alpha-1)})+O(T^{\alpha-2})\right] \\
&= \frac{2\tilde{C}\sigma_\Delta^2\tilde{p}}{\pi^2 j^2}\left[O(T^{\alpha-\beta})+O(T^{-\beta})\right]. \qquad (26)
\end{aligned}
$$

The term $\tilde{B}_3$ is of order $O(T^{-\alpha-\beta})$ by the same arguments as in the proof of Theorem 2. Consequently, parts i.) and ii.) of the Theorem follow directly from Equations (20) and (23). Similarly, part iii.) is the direct consequence of Equation (22) and Equations (25) and (26). □

Acknowledgements

Financial support of the Deutsche Forschungsgemeinschaft (DFG, grant number SI 745/9-2) is gratefully acknowledged. We further thank the participants of the CFE 2017, the econometrics study group of the "Verein für Socialpolitik", the Statistical Week 2017, the Brunel Conference on Financial and Macro Economics and Econometrics 2018, Martin Wagner, Rolf Tschernig, Bent Nielsen, Fabrizio Iacone, Niels Haldrup, Tommaso Proietti, Remigijus Leipus, Kai Wenger, Michelle Voges, Simon Wingert, and Michael Massmann for their helpful remarks and comments.

References

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long memory processes: Probabilistic properties and statistical methods*. Springer London, Limited.

Bhattacharya, R. N., Gupta, V. K., and Waymire, E. (1983). The hurst effect under trends. *Journal of Applied Probability*, 20(03):649–662.

Breidt, F. J. and Hsu, N.-J. (2002). A class of nearly long-memory time series models. *International Journal of Forecasting*, 18(2):265–281.

Brown, B. (1974). Generalized wald equations in discrete time. *Stochastic Processes and their Applications*, 2(4):349–357.

Christensen, B. J. and Varneskov, R. T. (2017). Medium band least squares estimation of fractional cointegration in the presence of low-frequency contamination. *Journal of Econometrics*, 197(2):218–244.

Davidson, J. and Sibbertsen, P. (2005). Generating schemes for long memory processes: regimes, aggregation and linearity. *Journal of Econometrics*, 128(2):253–282.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1):131–159.

Engle, R. F. and Smith, A. D. (1999). Stochastic permanent breaks. *Review of Economics and Statistics*, 81(4):553–574.

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

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

Giraitis, L., Kokoszka, P., and Leipus, R. (2001). Testing for long memory in the presence of a general trend. *Journal of Applied Probability*, 38(4):1033–1054.

Gourieroux, C. and Jasiak, J. (2001). Memory and infrequent breaks. *Economics Letters*, 70(1):29–41.

Granger, C. W. and Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, 73(1):61–77.

Granger, C. W. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11(3):399–421.

Granger, C. W. and Teräsvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economics Letters*, 62(2):161–165.

Guégan, D. (2005). How can we define the concept of long memory? An econometric survey. *Econometric Reviews*, 24(2):113–149.

Haldrup, N. and Valdés, J. E. V. (2017). Long memory, fractional integration, and cross-sectional aggregation. *Journal of Econometrics*, 199(1):1–11.

Hou, J. and Perron, P. (2014). Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations. *Journal of Econometrics*, 182(2):309–328.

Hsieh, M.-C., Hurvich, C. M., and Soulier, P. (2007). Asymptotics for duration-driven long range dependent processes. *Journal of Econometrics*, 141(2):913–949.

Iacone, F. (2010). Local whittle estimation of the memory parameter in presence of deterministic components. *Journal of Time Series Analysis*, 31(1):37–49.

Künsch, H. R. (1986). Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability*, 23(4):1025–1030.

Künsch, H. R. (1987). Statistical aspects of self-similar processes. *Proceedings of the First World Congress of the Bernoulli Society*, 1:67–74.

Leipus, R., Paulauskas, V., and Surgailis, D. (2005). Renewal regime switching and stable limit laws. *Journal of Econometrics*, 129(1-2):299–327.

Leipus, R. and Surgailis, D. (2003). Random coefficient autoregression, regime switching and long memory. *Advances in Applied Probability*, 35(3):737–754.

Liu, M. (2000). Modeling long memory in stock market volatility. *Journal of Econometrics*, 99(1):139–171.

Lobato, I. N. and Savin, N. E. (1998). Real and spurious long-memory properties of stock-market data. *Journal of Business & Economic Statistics*, 16(3):261–268.

McCloskey, A. and Hill, J. B. (2017). Parameter estimation robust to low-frequency contamination. *Journal of Business & Economic Statistics*, 35(4):598–610.

McCloskey, A. and Perron, P. (2013). Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory*, 29(06):1196–1237.

Mikosch, T., Resnick, S., Rootzén, H., and Stegeman, A. (2002). Is network traffic approximated by stable lévy motion or fractional brownian motion? *Annals of Applied Probability*, pages 23–68.

Mikosch, T. and Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86(1):378–390.

Miller, J. I. and Park, J. Y. (2010). Nonlinearity, nonstationarity, and thick tails: How they interact to generate persistence in memory. *Journal of Econometrics*, 155(1):83–89.

Parke, W. R. (1999). What is fractional integration? *Review of Economics and Statistics*, 81(4):632–638.

Perron, P. and Qu, Z. (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economic Statistics*, 28(2):275–290.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business & Economic Statistics*, 29(3):423–438.

Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502.

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

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23(3):1048–1072.

Sibbertsen, P., Leschinski, C., and Busch, M. (2018). A multivariate test against spurious long memory. *Journal of Econometrics*, 203(1):33 – 49.

Smith, A. (2005). Level shifts and the illusion of long memory in economic time series. *Journal of Business & Economic Statistics*, 23(3):321–335.

Taqqu, M. S., Willinger, W., and Sherman, R. (1997). Proof of a fundamental result in self-similar traffic modeling. *ACM SIGCOMM Computer Communication Review*, 27(2):5–23.