The Annals of Applied Probability
2024, Vol. 34, No. 3, 3227–3250
https://doi.org/10.1214/23-AAP2037

CLUSTERING OF LARGE DEVIATIONS IN MOVING AVERAGE PROCESSES: THE SHORT MEMORY REGIME

BY ARIJIT CHAKRABARTY[1,a] AND GENNADY SAMORODNITSKY[2,b]

[1]*Theoretical Statistics and Mathematics Unit, Indian Statistical Institute,* [a]*arijit.isi@gmail.com*
[2]*School of Operations Research and Information Engineering, Cornell University,* [b]*gs18@cornell.edu*

We describe the cluster of large deviations events that arise when one such large deviations event occurs. We work in the framework of an infinite moving average process with a noise that has finite exponential moments.

1. Introduction. Equity premium is a common notion in finance. It is the difference between long-term cumulative returns on the stock and the safe interest rate returns (see, e.g., Goyal and Welch (2003)); it enters as input into stock options issued on a regular basis. Long-term persistence in climate refers to the phenomenon where climatologically relevant parameters (e.g., the temperature) take higher than usual mean values over multiple time long periods (see e.g., Rybski et al. (2006)). What these diverse themes and many others like them have in common is that in the critical cases one observes a time series over long intervals where it averages out to "unexpected" values. These unexpected values cluster. What type of probabilistic analysis should address such phenomena?

A very crude classification of how we analyse random systems might split the work into distributional analysis and large deviations analysis. The distributional analysis deals with the "usual" deviations of a system from its "average" state, while the large deviations analysis deals the "unusually large" deviations, that are, by necessity, rare (but may have a major impact). The idea of clustering is a major idea in how we look at random systems. Clustering typically means that certain related events occur "in proximity to each other" and, when it happens, the impact of the events may be magnified, as in the situations we started with. Clustering is also interesting in its own right because it may shed light on certain structural elements in a random system. Clustering is most frequently studied in distributional analysis; an important example is clustering of extreme values; see, for example, Embrechts, Klüppelberg and Mikosch (2003). In the above examples, however, one is interested in clustering of "unexpected" values over long time periods. The point of view of large deviations is, therefore, called for.

In this work we discuss clustering of large deviations events. From a different point of view, we would like to understand whether or not a (rare) large deviations event is likely to cause a cascade of additional large deviations events and, if so, what does this cascade look like. Literature on large deviations analysis is vast, however to the best of our knowledge, clustering has not been considered even in the i.i.d. case; the applications we have in mind call for models with dependence. The nature of large deviations is known to be different in stochastic systems with "light tails" and with "heavy tails". The texts such as Dembo and Zeitouni (1998) or Deuschel and Stroock (1989) describe large deviations of light-tailed systems, while Mikosch and Nagaev (1998) will give the reader an idea of how large deviations occur in heavy-tailed systems. Large deviations are affected not only by the "tails" in a random system, but also the "memory" in that system, in particular by whether the memory is "short" or "long". The change from short to long memory in a system can be viewed as a

Received July 2023; revised November 2023.
MSC2020 subject classifications. 60F10.
Key words and phrases. Large deviations, clustering, infinite moving average.

phase transition (see Samorodnitsky (2016)), and it affects large deviations as well. In this work we study clustering of large deviations in a light-tailed system, but we will consider both short memory and long memory situations.

Let us now be more specific about the class of stochastic models we will consider. We will consider centered infinite moving average processes

$$X_n = \sum_{i=-\infty}^{\infty} a_i Z_{n-i}, \quad n \in \mathbb{Z}, \tag{1.1}$$

where $(Z_n : n \in \mathbb{Z})$ is a collection of i.i.d. nondegenerate random variables (the noise) with distribution F_Z satisfying

$$\int_{\mathbb{R}} e^{tz} F_Z(dz) < \infty \quad \text{for all } t \in \mathbb{R}, \tag{1.2}$$

and

$$\int_{\mathbb{R}} z F_Z(dz) = 0. \tag{1.3}$$

For future reference we denote

$$\sigma_Z^2 = \int_{\mathbb{R}} z^2 F_Z(dz). \tag{1.4}$$

Let $\dots, a_{-1}, a_0, a_1, a_2 \dots$ be real numbers satisfying

$$\sum_{j=-\infty}^{\infty} a_j^2 < \infty. \tag{1.5}$$

Since the assumption (1.2) implies that the noise variables have a finite second moment, the zero mean property assumed in (1.3) and the square integrability of the coefficients (1.5) imply that the infinite sum in the right hand side of (1.1) converges in L^2 and a.s. and defines a zero mean stationary ergodic process. Therefore, for $\varepsilon > 0$ the event

$$E_0(n, \varepsilon) = \left\{ \frac{1}{n} \sum_{i=0}^{n-1} X_i \ge \varepsilon \right\}$$

is, for large n, a rare, large deviations, event. We would like to understand whether occurrence of this event may cause a cascade of related events. Specifically, for $j \ge 0$ we denote

$$E_j(n, \varepsilon) = \left\{ \frac{1}{n} \sum_{i=j}^{n+j-1} X_i \ge \varepsilon \right\}, \tag{1.6}$$

so that each event $E_j(n, \varepsilon)$ is equally rare, and we would like to know how many of the events for j "reasonably close to $j = 0$" occur if $E_0(n, \varepsilon)$ occurs (the reason for the qualifier "reasonably close to $j = 0$" is that by ergodicity, the events $E_j(n, \varepsilon)$ will keep recurring eventually, regardless of the structure of the system).

The difference between short memory infinite moving average processes and long memory infinite moving average processes lies in the rate the coefficients (a_n) converge to zero (subject to the square summability, of course). This will lead to markedly different cascading of the events $E_j(n, \varepsilon)$, conditionally on the event $E_0(n, \varepsilon)$ occurring. In this paper we consider the short memory processes; the long memory case is studied in another paper—Chakrabarty and Samorodnitsky (2023). Specifically, we describe the limiting distribution of the large deviation cluster caused by the rare event $E_0(n, \varepsilon)$ as well as the behaviour of the size of that cluster as the overshoot ε becomes small. It turns out that for such ε the size of the cluster is of the order ε^{-2}.

Our main results on the cluster of large deviations for short memory infinite moving average processes are in Section 2. They are presented in a somewhat more general way than discussed so far. Even though the "causal" point of view encourages us to consider the "forward" large deviation events $E_j(n,\varepsilon)$ in (1.6) for $j \geq 0$, in certain situations it is also interesting to look at the "backward" large deviation events $E_j(n,\varepsilon)$ with $j < 0$, and this is included in Section 2. The concluding Section 3 contains a discussion of the results we obtain and connects them to what one may expect when the memory becomes long.

2. Short memory moving average processes. We follow the common terminology and say that the infinite moving average process (1.1) has short memory if

$$\sum_{n=-\infty}^{\infty} |a_n| < \infty \quad \text{and} \quad \sum_{n=-\infty}^{\infty} a_n \neq 0. \tag{2.1}$$

Large and moderate deviations for such processes have been studied in, for example, Jiang, Wang and Rao (1992) and Djellout and Guillin (2001).

We investigate the clustering of the rare events $(E_j(n,\varepsilon))$ in the following way. We will show that the conditional law of the process of occurrences of the large deviation events,

$$\big(1\big(E_j(n,\varepsilon), j \in \mathbb{Z}\big)\big), \tag{2.2}$$

given $E_0(n,\varepsilon)$ has a nondegenerate weak limit and describe that limit. This will show, in particular, that for fixed $K_-, K_+ \in \mathbb{Z}$ the joint conditional law of the total numbers of occurrences among the last K_- of the events $(E_j(n,\varepsilon))$ before 0 and the first K_+ of the events $(E_j(n,\varepsilon))$ after 0,

$$\nu_n(K_-, K_+, \varepsilon)(\cdot) = P\left[\left(\sum_{j=-K_-}^{-1} 1\big(E_j(n,\varepsilon)\big), \sum_{j=1}^{K_+} 1\big(E_j(n,\varepsilon)\big)\right) \in \cdot \,|E_0(n,\varepsilon)\right] \tag{2.3}$$

has a weak limit. That weak limit itself converges weakly, as $K_-, K_+ \to \infty$, to an a.s. finite random variable that we interpret as the size of the cluster of large deviation events containing a large deviation event at time zero. An interesting regime is that of a small $\varepsilon > 0$, and we show that a properly normalized size of the cluster of large deviation events converges weakly, as $\varepsilon \to 0$, to an (interesting) limit. As we have explained above, the limits should be taken in this specific order.

To state the main results of this section we need to introduce some notation first. Denote

$$A = \sum_{n=-\infty}^{\infty} a_n = \sum_{n=-\infty}^{-1} a_n + \sum_{n=0}^{\infty} a_n := A^- + A^+. \tag{2.4}$$

In the sequel we will assume that $A > 0$. Note that, in view of (2.1), this introduces no real loss of generality because, if the sum is negative, we simply multiply both (Z_n) and (a_n) by -1 and reduce the situation to the case $A > 0$ we are considering. Further, for $n = 0, 1, 2, \dots$ we write

$$A_n^- = \sum_{j=1}^{n} a_{-j}, \qquad A_n^+ = \sum_{j=0}^{n} a_j. \tag{2.5}$$

Next, we let

$$\varphi_Z(t) = \log\left(\int_{\mathbb{R}} e^{tz} F_Z(dz)\right), \quad t \in \mathbb{R} \tag{2.6}$$

be the log-Laplace transform of a noise variable. For $\theta \in \mathbb{R}$ we denote by G_θ the probability measure on $\mathbb{R}$ obtained by exponentially tilting F_Z as follows:

$$G_\theta(dx) = e^{-\varphi_Z(\theta)+\theta x} F_Z(dx), \quad x \in \mathbb{R}. \tag{2.7}$$

Further, let

$$s_0 = \sup\{x \in \mathbb{R} : F_Z(x) < 1\} \in (0, \infty]$$

be the right endpoint of the support of a noise variable. We will consider the events $(E_j(n, \varepsilon))$ for ε satisfying

$$0 < \varepsilon/A < s_0. \tag{2.8}$$

The function φ_Z is infinitely differentiable, and its first derivative φ_Z' strictly increases from 0 at $t = 0$ to s_0 as $t \to \infty$. Therefore, for ε satisfying (2.8), we can unambiguously define $\tau(\varepsilon) > 0$ by

$$\varphi_Z'\big(\tau(\varepsilon)\big) = \varepsilon/A. \tag{2.9}$$

We now introduce a collection $\{Z_j^u : j \in \mathbb{Z}, u = + \text{ or } -\}$ of independent random variables with the following laws:

$$\begin{aligned} Z_{-j}^- &\sim G_{\tau(\varepsilon)(A^+ - A_{j-1}^+)/A}, \quad j \geq 1, \\ Z_j^- &\sim G_{\tau(\varepsilon)(A^+ + A_j^-)/A}, \quad j \geq 0, \\ Z_{-j}^+ &\sim G_{\tau(\varepsilon)(A_{j-1}^+ + A^-)/A}, \quad j \geq 1, \\ Z_j^+ &\sim G_{\tau(\varepsilon)(A^- - A_j^-)/A}, \quad j \geq 0. \end{aligned} \tag{2.10}$$

Finally, let T^* be an exponential random variable with parameter $\tau(\varepsilon)/A$, independent of the family (2.10).

It is elementary to check that for any $j \in \mathbb{Z}$ and $u = +$ or $-$,

$$E\big(|Z_j^u|\big) \leq \int_{\mathbb{R}} |x| G_{A^{-1}\bar{A}\tau(\varepsilon)}(dx) + \int_{\mathbb{R}} |x| G_{-A^{-1}\bar{A}\tau(\varepsilon)}(dx) < \infty,$$

where

$$\bar{A} = \sum_{n=-\infty}^{\infty} |a_n|.$$

Therefore, the infinite series

$$U_n^- = \sum_{i=-\infty}^{\infty} a_i Z_{n-i}^-, \quad n \in \mathbb{Z}, \tag{2.11}$$

$$U_n^+ = \sum_{i=-\infty}^{\infty} a_i Z_{n-i}^+, \quad n \in \mathbb{Z} \tag{2.12}$$

converge in L^1 and a.s. We define

$$V_j(\varepsilon) = \begin{cases} \mathbf{1}\left(T^* \geq \sum_{i=0}^{j-1} (U_i^- - U_i^+)\right), & j \geq 1, \\ \mathbf{1}\left(T^* \geq \sum_{i=j}^{-1} (U_i^+ - U_i^-)\right), & j < 0. \end{cases} \tag{2.13}$$

REMARK 2.1. It is instructive to see what the above random objects become in the i.i.d. case $a_0 = 1$, $a_i = 0$ for all $i \neq 0$. Then $U_{-j}^- = Z_{-j}^- \sim G$, $j \geq 1$, $U_j^- = Z_j^- \sim G_{\tau(\varepsilon)}$, $j \geq 0$, $U_{-j}^+ = Z_{-j}^+ \sim G_{\tau(\varepsilon)}$, $j \geq 1$, $U_j^+ = Z_j^+ \sim G$, $j \geq 0$. The processes $\sum_{i=0}^{j-1}(U_i^- - U_i^+)$, $j \geq 1$ and $\sum_{i=j}^{-1}(U_i^+ - U_i^-)$, $j < 0$ become independent random walks with the same step distribution, that of the difference $A - B$ of independent random variables, $A \sim G_{\tau(\varepsilon)}$, $B \sim G$. That is, $V_j(\varepsilon)$ for both positive and negative j are equal to 1 as long as these random walks stay below level T_*, and then they vanish.

We are now ready to state the main theorems of this section. They rely on a technical assumption, excluding the case of a lattice-valued noise. We assume that

$$\left|\int_{\mathbb{R}} e^{itz} F_Z(dz)\right| < 1 \quad \text{for any } t \neq 0 \text{ where } i = \sqrt{-1}. \tag{2.14}$$

Our first result describes the behaviour of the sequence of conditional laws of the process of the overshoots of the level $n\varepsilon$ by the partial sums of length n. This leads to the limiting behaviour of the sequence of conditional laws of the process of occurrences of the large deviations events (2.2) and of the sequence (2.3) of the total number of occurrences among the first K of the events $(E_j(n,\varepsilon))$. The weak convergence of sequences of random variables stated in this theorem occurs in the usual topology of finite-dimensional convergence in $\mathbb{R}^\infty$ (or its restriction to $\{0,1\}^\infty$).

THEOREM 1. *Assume that* (2.1) *holds and* $A > 0$ *in* (2.4). *Assume, further, that the characteristic function of the noise variables satisfies* (2.14). *Let* ε *be as in* (2.8). *Then, as* $n \to \infty$,

$$\begin{aligned} &P\left(\left(\sum_{i=j}^{j+n-1} X_i - n\varepsilon,\ j \in \mathbb{Z}\right) \in \cdot \,|\, E_0(n,\varepsilon)\right) \\ &\quad \Rightarrow \left(T^* + \sum_{i=j}^{-1} U_i^- - \sum_{i=j}^{-1} U_i^+,\ j = \cdots, -2, -1,\right. \\ &\qquad \left. T^* - \sum_{i=0}^{j-1} U_i^- + \sum_{i=0}^{j-1} U_i^+,\ j = 0, 1, 2, \ldots\right) \end{aligned} \tag{2.15}$$

in $\mathbb{R}^\infty$. *In particular,*

$$P\big(\big(1\big(E_j(n,\varepsilon),\ j \in \mathbb{Z}\big)\big) \in \cdot \,|\, E_0(n,\varepsilon)\big) \Rightarrow P\big(\big(V_j(\varepsilon),\ j \in \mathbb{Z}\big) \in \cdot\big) \tag{2.16}$$

in $\{0,1\}^\infty$. *Furthermore, for every fixed* $K_-, K_+ \geq 1$, *the conditional laws* $(\nu_n(K_-, K_+, \varepsilon))$ *in* (2.3) *satisfy*

$$\nu_n(K_-, K_+, \varepsilon)(\cdot) \Rightarrow P\left[\left(\sum_{j=-K_-}^{-1} V_j(\varepsilon), \sum_{j=1}^{K_+} V_j(\varepsilon)\right) \in \cdot\right] \quad \text{as } n \to \infty. \tag{2.17}$$

REMARK 2.2. The limiting process of overshoots appearing in the right hand side of (2.15) has a remarkable property. Define

$$Y_j = \begin{cases} \exp\left\{T^* + \sum_{i=j}^{-1} U_i^- - \sum_{i=j}^{-1} U_i^+\right\}, & j = \cdots, -2, -1, \\ \exp\left\{T^* - \sum_{i=0}^{j-1} U_i^- + \sum_{i=0}^{j-1} U_i^+\right\}, & j = 0, 1, 2, \ldots. \end{cases} \tag{2.18}$$

We claim that the process $\mathbf{Y} = (Y_j,\, j \in \mathbb{Z})$ satisfies the following time change formula:

$$E\big[H\big(tB^j\mathbf{Y}\big)\mathbf{1}(Y_{-j} > 1/t)\big] = t^{\tau(\varepsilon)/A}E\big[H(\mathbf{Y})\mathbf{1}(Y_j > t)\big] \tag{2.19}$$

for every $j \in \mathbb{Z}$, $t > 0$ and every measurable bounded functional H on $[0,\infty)^\infty$. Here B is the usual backward shift operator on $[0,\infty)^\infty$.

The time change property (2.19) has been known to be an important property of the so-called *tail process* in extreme value theory. Specifically, if $(W_j,\, j \in \mathbb{Z})$ is a nonnegative stationary process with multivariate regularly varying tails with exponent $\alpha > 0$, then the conditional law of $(W_j/W_0,\, j \in \mathbb{Z})$ given $W_0 > w$ has a weak limit, as $w \to \infty$. This limit is the law of the tail process, described by Basrak and Segers (2009), who also discovered the time change property, with the exponent $\tau(\varepsilon)/A$ replaced by α (allowing nonnecessarily nonnegative values, and in dimensions greater than 1). This property has become the subject of further investigations; we refer the reader to its recent presentation in a book form in Theorem 5.3.1 in Kulik and Soulier (2020).

The tail process in extreme value theory arises in the distributional context of tail behaviour of stationary processes with regularly varying tails. The limiting process $\mathbf{Y}$ in (2.18) arises in the context of large deviations of stationary moving average processes, without any presence of regular variation in the underlying model. This makes appearance of the time change property here unexpected. It would be interesting to investigate how widespread this phenomenon is in the realm of large deviations, but we are not pursuing this question any further in this paper.

In order to prove that the process $\mathbf{Y}$ in (2.18) satisfies (2.19), suppose first that $j \geq 0$. It is elementary that the Pareto-distributed random variable $W = e^{T*}$ has the property

$$t^{\tau(\varepsilon)/A}E\big[h(W)\mathbf{1}(W > t)\big] = E\big[h(tW)1(W > 1/t)\big], \tag{2.20}$$

valid for any $t > 0$ and any bounded measurable function h. Using this property we have

$$\begin{aligned}
&t^{\tau(\varepsilon)/A}E\big[H(\mathbf{Y})\mathbf{1}(Y_j > t)\big] \\
&\quad = t^{\tau(\varepsilon)/A}E\Bigg[H\Bigg(W\exp\Bigg\{\sum_{i=k_1}^{-1}(U_i^- - U_i^+)\Bigg\}, k_1 = \ldots, -2, -1, \\
&\qquad W\exp\Bigg\{\sum_{i=0}^{k_2-1}(U_i^+ - U_i^-)\Bigg\}, k_2 = 0, 1, \ldots\Bigg) \\
&\qquad \times \mathbf{1}\Bigg(W\exp\Bigg\{\sum_{i=0}^{j-1}(U_i^+ - U_i^-)\Bigg\} > t\Bigg)\Bigg] \\
&\quad = E\Bigg[H\Bigg(tW\exp\Bigg\{\sum_{i=k_1}^{-1}(U_i^- - U_i^+)\Bigg\}, k_1 = \ldots, -2, -1, \\
&\qquad tW\exp\Bigg\{\sum_{i=0}^{k_2-1}(U_i^+ - U_i^-)\Bigg\}, k_2 = 0, 1, \ldots\Bigg) \\
&\qquad \times \exp\Bigg\{\frac{\tau(\varepsilon)}{A}\sum_{i=0}^{j-1}(U_i^+ - U_i^-)\Bigg\}\mathbf{1}\Bigg(W\exp\Bigg\{\sum_{i=0}^{j-1}(U_i^- - U_i^+)\Bigg\} > 1/t\Bigg)\Bigg].
\end{aligned}$$

The factor

$$\exp\Bigg\{\frac{\tau(\varepsilon)}{A}\sum_{i=0}^{j-1}(U_i^+ - U_i^-)\Bigg\}$$

under the expectation in the right hand side introduces additional exponential tilting into the random variables $Z_d^{\pm}$ in (2.10). Specifically, it is a matter of straightforward calculus to check that, distributionally, Z_d^- becomes Z_{d-j}^- and Z_d^+ becomes Z_{d-j}^+ for every $d \in \mathbb{Z}$. That is, distributionally, $U_n^{\pm}$ becomes $U_{n-j}^{\pm}$ for every $n \in \mathbb{Z}$. Therefore,

$$\begin{aligned}
&t^{\tau(\varepsilon)/A} E\big[H(\mathbf{Y})\mathbf{1}(Y_j > t)\big] \\
&\quad = E\Bigg[H\Bigg(tW \exp\Bigg\{\sum_{i=k_1}^{-1}(U_{i-j}^- - U_{i-j}^+)\Bigg\}, k_1 = \dots, -2, -1, \\
&\qquad tW \exp\Bigg\{\sum_{i=0}^{k_2-1}(U_{i-j}^+ - U_{i-j}^-)\Bigg\}, k_2 = 0, 1, \dots\Bigg) \\
&\qquad \times \mathbf{1}\Bigg(W \exp\Bigg\{\sum_{i=0}^{j-1}(U_{i-j}^- - U_{i-j}^+)\Bigg\} > 1/t\Bigg)\Bigg] \\
&\quad = E\big[H\big(tB^j\mathbf{Y}\big)\mathbf{1}(Y_{-j} > 1/t)\big],
\end{aligned}$$

with the last step following from a rearrangement of indices. This proves (2.19) for $j \geq 0$. Finally, if $j < 0$, then we apply (2.19) to $\tilde{j} = -j$, $\tilde{t} = 1/t$ and $\tilde{H}(\cdot) = H(tB^j\cdot)$, proving (2.19) in this case as well.

REMARK 2.3. The conditional limiting behaviour described in Theorem 2.1 reflects the nature of large deviations for weighted sums of random variables with exponentially light tails: the terms in the sums "conspire to change their distributions just right" to make the rare event happen (and the change in distributions is reflected in the exponential tilting). As the tails become heavier, the nature of large deviations gradually changes from a "conspiration" to a "single extraordinary value" phenomenon. This will result in a change of how large deviation events cluster. We outline it in a simple example and leave a full discussion to a different occasion. Consider the finite moving average $X_n = Z_n + Z_{n-1}$, $n \in \mathbb{Z}$, where (Z_n) are i.i.d. 0 mean random variables whose right tail is regularly varying with exponent $\alpha > 1$. Then

$$S_n = Z_{-1} + Z_{n-1} + 2\sum_{i=0}^{n-2} Z_i,$$

and the event $E_0(n, \varepsilon)$ is, asymptotically, equivalent to the event that (exactly) one of the Z_i with $i = 0, 1, \dots, n-2$ is larger than $n\varepsilon/2$, with equal probabilities (it also possible that Z_1 or $Z_{n-1} > n\varepsilon$, but the probability of this goes to 0 as $n \to \infty$). For any $j \geq 1$, the probability that this exceptional i is in the range $j, \dots, n-2$ converges to 1, and the corresponding term $2Z_i$ is also a part of the sum $\sum_{j=i}^{j+n-1} Z_j$. Therefore, the conditional probability of $E_j(n, \varepsilon)$ given that $E_0(n, \varepsilon)$ occurs converges to 1 for every $j \geq 1$, as $n \to \infty$.

It is natural to interpret the statement of Theorem 1 as saying that a large deviation event $E_0(n, \varepsilon)$, upon occurring, leads to random clusters of large deviation events in the past and in the future. The limiting (as $n \to \infty$) total sizes of these clusters have the joint law of

$$(D_\varepsilon^-, D_\varepsilon^+) = \Bigg(\sum_{j=-\infty}^{-1} V_j(\varepsilon), \sum_{j=1}^{\infty} V_j(\varepsilon)\Bigg), \quad \varepsilon > 0. \tag{2.21}$$

Our second result of this section shows that for small $\varepsilon > 0$ these total cluster sizes are a.s. finite and describes their joint limiting behaviour as the overshoot $\varepsilon \to 0$.

THEOREM 2. *Under the assumptions of Theorem 1, the total cluster sizes D_ε^-, D_ε^+ are a.s. finite for $\varepsilon > 0$ small enough. Further, as $\varepsilon \to 0$,*

$$(\varepsilon^2 D_\varepsilon^-, \varepsilon^2 D_\varepsilon^+) \Rightarrow \biggl(A^2\sigma_Z^2 \int_0^\infty 1\bigl(T_0 \ge (\sqrt{2}B_t^- + t)\bigr)\,dt,\\ A^2\sigma_Z^2 \int_0^\infty 1\bigl(T_0 \ge (\sqrt{2}B_t^+ + t)\bigr)\,dt\biggr),$$

where A is the sum of the coefficients (2.4) and σ_Z^2 is the noise variance (1.4). Furthermore, T_0 is a standard exponential random variable independent of two independent standard Brownian motions $(B_t^\pm : t \ge 0)$.

We prove Theorem 1 first, and so $\varepsilon > 0$ (satisfying (2.8)) is for now fixed. The proof is via several technical lemmas, and we first sketch the flow of the argument. To simplify the notation we will write E_j instead of $E_j(n, \varepsilon)$ throughout. We start by deriving a nonlogarithmic asymptotic formula for the probability of E_0, which we use to show that, conditionally on E_0, all noise variables remain uniformly bounded in L_1 and, further, jointly weakly converge to the appropriately exponentially tilted laws. This allows us to prove that the sums of finitely truncated moving averages converge weakly, and this takes us very close to the finish.

We now embark on the technical details.

LEMMA 2.1. *Denote*

$$S_n = \sum_{i=0}^{n-1} X_i, \quad n \ge 1, \tag{2.22}$$

and let

$$\psi_n(t) = n^{-1} \log E\bigl(e^{tS_n}\bigr), \quad t \in \mathbb{R}, n \ge 1. \tag{2.23}$$

Then for all large enough n there exists a unique $\theta_n > 0$ such that

$$\psi_n'(\theta_n) = \varepsilon.$$

Furthermore,

$$P(E_0) \sim \frac{C}{\sqrt{n}} \exp\bigl(-n\bigl(\theta_n\varepsilon - \psi_n(\theta_n)\bigr)\bigr), \quad n \to \infty,$$

with

$$C = \frac{1}{\tau(\varepsilon)\sqrt{2\pi\varphi_Z''(\tau(\varepsilon))}},$$

and φ_Z and $\tau(\varepsilon)$ defined, respectively, in (2.6) and (2.9).

PROOF. We write

$$S_n = \sum_{j=0}^{n-1}(A_{n-1-j}^+ + A_j^-)Z_j + \sum_{j=n}^{\infty}(A_j^- - A_{j-n}^-)Z_j \\ + \sum_{j=1}^{\infty}(A_{j+n-1}^+ - A_{j-1}^+)Z_{-j}, \tag{2.24}$$

with A_n^+, A_n^- defined by (2.5) and check the conditions of Theorem 3 in the Appendix. As a first step we show that

$$\lim_{n\to\infty} \psi_n''(t) = A^2 \varphi_Z''(At) \tag{2.25}$$

locally uniformly in $t \in \mathbb{R}$. Indeed, by (2.24),

$$\begin{aligned}\psi_n(t) = \frac{1}{n}\Biggl[\sum_{j=0}^{n-1} \varphi_Z\bigl((A_{n-1-j}^+ + A_j^-)t\bigr) + \sum_{j=n}^{\infty} \varphi_Z\bigl((A_j^- - A_{j-n}^-)t\bigr) \\ + \sum_{j=1}^{\infty} \varphi_Z\bigl((A_{j+n-1}^+ - A_{j-1}^+)t\bigr)\Biggr].\end{aligned} \tag{2.26}$$

It is easy to see that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} (A_{n-1-j}^+ + A_j^-)^2 \varphi_Z''\bigl((A_{n-1-j}^+ + A_j^-)t\bigr) = A^2 \varphi_Z''(At). \tag{2.27}$$

Indeed, for every $0 < \eta < 1/2$,

$$\frac{1}{n}\sum_{j=[\eta n]}^{[(1-\eta)n]} (A_{n-1-j}^+ + A_j^=)^2 \varphi_Z''\bigl((A_{n-1-j}^+ + A_j^-)t\bigr) \to (1-2\eta)A^2\varphi_Z''(At),$$

since the terms in the sum converge uniformly to the limit. Furthermore, since φ_Z'' is locally bounded,

$$\frac{1}{n}\sum_{j=0}^{[\eta n]-1} (A_{n-1-j}^+ + A_j^-)^2 \varphi_Z''\bigl((A_{n-1-j}^+ + A_j^-)t\bigr) \le C\eta$$

for some finite constant C. Since the second remaining part of the sum in (2.27) can be bounded in the same way, by letting first $n \to \infty$ and then $\eta \to 0$ we obtain (2.27). Furthermore, as $n \to \infty$,

$$\begin{aligned}\sum_{j=1}^{\infty} (A_{j+n-1}^+ - A_{j-1}^+)^2 \varphi_Z''\bigl((A_{j+n-1}^+ - A_{j-1}^+)t\bigr) &= O\Biggl(\sum_{j=1}^{\infty} (A_{j+n-1}^+ - A_{j-1}^+)^2\Biggr) \\ &= O\Biggl(\sum_{j=1}^{\infty}\sum_{i=j}^{j+n-1} |a_i|\Biggr) \\ &= O\Biggl(\sum_{i=1}^{n}\sum_{j=i}^{\infty} |a_j|\Biggr) = o(n).\end{aligned}$$

Since the second sum in (2.26) can be estimated in the same way, and all these steps are locally uniform in $t \in \mathbb{R}$, (2.25) follows. Since this argument also shows that ψ_n'' is, uniformly in n, locally bounded, and the values of ψ_n, φ_Z and their respective first derivatives at 0 are 0, we also conclude that ψ_n' is, uniformly in n, locally bounded and for every $t \ge 0$,

$$\begin{aligned}\lim_{n\to\infty} \psi_n'(t) &= A\varphi_Z'(At), \\ \lim_{n\to\infty} \psi_n(t) &= \varphi_Z(At).\end{aligned} \tag{2.28}$$

The assumption (2.8) together with (2.28) implies that for large n there exists a unique $\theta_n > 0$ such that $\psi_n'(\theta_n) = \varepsilon$, and that

$$\lim_{n\to\infty} \theta_n = A^{-1}\tau(\varepsilon). \tag{2.29}$$

We choose n_1 so large that for $n \geq n_1$, θ_n is well defined, $A/2 \leq A^+_{n-1-j} + A^-_j \leq \sqrt{2}A$ for all $n/3 \leq j \leq 2n/3$ and $\theta_n \leq \sqrt{2}\tau(\varepsilon)/A$.

We claim next that for fixed $\delta, \lambda > 0$ there exists $\eta \in (0, 1)$ such that

$$\sup_{\delta \leq |t| \leq \lambda\theta_n} \left| \frac{1}{E(e^{\theta_n S_n})} E\big(e^{(\theta_n + it)S_n}\big) \right| = O(\eta^n), \quad n \to \infty, \tag{2.30}$$

with the convention that the supremum of the empty set is zero. To see this, note that $\phi : \mathbb{R}^2 \to \mathbb{C}$ defined by

$$\phi(\theta, t) = \frac{1}{E(e^{\theta Z})} E\big(e^{(\theta+it)Z}\big),$$

is continuous. For a fixed $\theta \in \mathbb{R}$, $\phi(\theta, \cdot)$ is the characteristic function of the distribution G_θ in (2.7). By (2.14), G_θ is not a lattice distribution and, hence, for any fixed $\lambda, \delta > 0$ and θ,

$$\sup_{A\delta/2 \leq |t| \leq 2\lambda\tau(\varepsilon)} \big|\phi(\theta, t)\big| < 1.$$

A standard compactness argument and (2.29) imply that

$$\eta := \Big(\sup_{n \geq n_1, n/3 \leq j \leq 2n/3, A\delta/2 \leq |t| \leq 2\lambda\tau(\varepsilon)} \big|\phi\big(\theta_n\big(A^+_{n-1-j} + A^-_j\big), t\big)\big| \Big)^{1/3} < 1,$$

while the choice of n_1 implies that for $n > j \geq n_1$ and $\delta \leq |t| \leq \lambda\theta_n$,

$$A\delta/2 \leq \big|\big(A^+_{n-1-j} + A^-_j\big)t\big| \leq 2\lambda\tau(\varepsilon).$$

Therefore, by (2.24) and the triangle inequality,

$$\begin{aligned}\frac{1}{E(e^{\theta_n S_n})} \big|E\big(e^{(\theta_n+it)S_n}\big)\big| &\leq \prod_{j=[n/3]+1}^{\lceil 2n/3 \rceil} \big|\phi\big(\theta_n\big(A^+_{n-1-j} + A^-_j\big), t\big(A^+_{n-1-j} + A^-_j\big)\big)\big| \\ &\leq \eta^n,\end{aligned}$$

establishing (2.30).

We have now verified all conditions of Theorem 3 for $T_n = S_n$, $a_n = n$, $m_n = \varepsilon$ and $\tau_n = \theta_n$, and (A.7) gives us the statement of the lemma. $\square$

We proceed with showing uniform boundedness of conditional moments of all noise variables.

LEMMA 2.2. *We have*

$$\sup_{n \geq 1, j \in \mathbb{Z}} E\big(|Z_j| | E_0\big) < \infty. \tag{2.31}$$

PROOF. Fix $j \in \mathbb{Z}$ and define

$$S_{n,j} = S_n - \beta_{n,j} Z_j, \quad n \geq 1,$$

where

$$\beta_{n,j} = \begin{cases} A^-_j - A^-_{j-n}, & 1 \leq n \leq j, \\ A^+_{n-1-j} + A^-_j, & n \geq j+1 \end{cases}$$

if $j \geq 0$ and

$$\beta_{n,j} = A^+_{n-1-j} - A^+_{-j-1}, \quad n \geq 1$$

if $j \le -1$, with S_n is as in (2.22). It follows from (2.24) that $S_{n,j}$ and Z_j are independent. We define

$$\tilde{\psi}_{n,j}(t) = n^{-1} \log E\big(e^{tS_{n,j}}\big), \quad t \in \mathbb{R}. \tag{2.32}$$

Since the numbers $(\beta_{n,j})$ are bounded uniformly in j and n, it follows that the functions in (2.23) and (2.32) satisfy

$$\psi_n'(\theta) = \tilde{\psi}_{n,j}'(\theta) + O(1/n),$$

with $O(1/n)$ uniform over j and θ in a compact interval. The same argument as in Lemma 2.1 shows that for large n there exists a unique $\tilde{\theta}_{n,j} > 0$ such that

$$\tilde{\psi}_{n,j}'(\tilde{\theta}_{n,j}) = \varepsilon.$$

Since φ_Z' is locally bounded away from zero, it follows from (2.28) that

$$\tilde{\theta}_{n,j} = \theta_n + O(1/n) \tag{2.33}$$

with $O(1/n)$ uniform over j. This also implies that

$$\psi_n(\tilde{\theta}_{n,j}) = \psi_n(\theta_n) + O(1/n). \tag{2.34}$$

For large n we can write

$$\begin{aligned} E\big(|Z_j| 1(E_0)\big) &= \int_{\mathbb{R}} |z| P(Z_j \in dz) \int_{\mathbb{R}} 1(s + \beta_{n,j} z \ge n\varepsilon) P(S_{n,j} \in ds) \\ &= \exp\{-n\big(\tilde{\theta}_{n,j}\varepsilon - \psi_n(\tilde{\theta}_{n,j})\big)\} \int_{\mathbb{R}} |z| e^{-\tilde{\theta}_{n,j}\beta_{n,j} z} P\big(Z_{n,j}^* \in dz\big) \\ &\quad \times \int_{[n\varepsilon - \beta_{n,j} z, \infty)} \exp\big(-\tilde{\theta}_{n,j}(s - n\varepsilon)\big) P\big(S_{n,j}^* \in ds\big), \end{aligned}$$

where $S_{n,j}^*$ and $Z_{n,j}^*$ are independent random variables with $Z_{n,j}^*$ having distribution $G_{\tilde{\theta}_{n,j}\beta_{n,j}}$ and

$$P\big(S_{n,j}^* \in ds\big) = \frac{1}{E(\exp(\tilde{\theta}_{n,j} S_{n,j}))} e^{\tilde{\theta}_{n,j} s} P(S_{n,j} \in ds).$$

It follows from (2.33) and (2.34) that, uniformly over j,

$$\begin{aligned} \exp\{-n\big(\tilde{\theta}_{n,j}\varepsilon - \psi_n(\tilde{\theta}_{n,j})\big)\} &= O\big(\exp\big(-n\big(\theta_n\varepsilon - \psi_n(\theta_n)\big)\big)\big) \\ &= O\big(\sqrt{n} P(E_0)\big), \end{aligned}$$

with the second line implied by Lemma 2.1. Therefore, to complete the proof it suffices to show that, uniformly in j,

$$\begin{aligned} &\int_{\mathbb{R}} |z| e^{-\tilde{\theta}_{n,j}\beta_{n,j} z} P\big(Z_{n,j}^* \in dz\big) \int_{[n\varepsilon - \beta_{n,j} z, \infty)} \exp\big(-\tilde{\theta}_{n,j}(s - n\varepsilon)\big) P\big(S_{n,j}^* \in ds\big) \\ &\quad = O\big(n^{-1/2}\big). \end{aligned} \tag{2.35}$$

This will follow from the following claim: there is $C' > 0$ such that for all n large,

$$P\big(y \le S_{n,j}^* \le y + 1\big) \le C' n^{-1/2}, \quad y \in \mathbb{R}, \tag{2.36}$$

uniformly in j. Indeed, suppose that this is the case. Then for large n and every $z \in \mathbb{R}$,

$$\begin{aligned}
&\int_{[n\varepsilon-\beta_{n,j}z,\infty)} \exp\bigl(-\tilde{\theta}_{n,j}(s-n\varepsilon)\bigr) P\bigl(S^*_{n,j} \in ds\bigr) \\
&\quad \le \sum_{j=1}^{\infty} e^{-\tilde{\theta}_{n,j}(j-1-\beta_{n,j}z)} P\bigl(S^*_{n,j} - n\varepsilon - \beta_{n,j}z \in [j-1,j)\bigr) \\
&\quad \le C' n^{-1/2} e^{\tilde{\theta}_{n,j}\beta_{n,j}z}\bigl(1-e^{-\tilde{\theta}_{n,j}}\bigr)^{-1},
\end{aligned}$$

which shows (2.35).

It remains to prove (2.36). We start by observing that for $M > 0$

$$P\bigl(S^*_{n,j} > nM\bigr) \le \exp\bigl\{n\bigl[\tilde{\psi}_{n,j}(\tilde{\theta}_{n,j}+1) - \tilde{\psi}_{n,j}(\tilde{\theta}_{n,j}) - M\tilde{\theta}_{n,j}\bigr]\bigr\}.$$

Since the values of both $\tilde{\psi}_{n,j}(\tilde{\theta}_{n,j}+1)$ and $\tilde{\psi}_{n,j}(\tilde{\theta}_{n,j})$ remain within a compact set independent of n and j, while $\theta_{n,j}$ converges, uniformly in j, to $A^{-1}\tau(\varepsilon) > 0$, we see that by choosing M large enough we can ensure that there is $c > 0$ such that for all n large enough,

$$P\bigl(S^*_{n,j} > nM\bigr) \le e^{-cn} \quad \text{for all } j.$$

An identical argument shows that, if $M > 0$ is large enough, then there is $c > 0$ such that for all n large enough,

$$P\bigl(S^*_{n,j} < -nM\bigr) \le e^{-cn} \quad \text{for all } j.$$

That means that it suffices to prove that (2.36) holds for all $|y| \le nM$, uniformly in j.

Notice that by part (b) of Theorem 3, for any $h > 0$ there is $C'_h > 0$ such that

$$P\bigl(y \le S^*_n \le y+h\bigr) \le C'_h n^{-1/2}, \quad y \in \mathbb{R}, \tag{2.37}$$

where S^*_n is a random variable with the law

$$P\bigl(S^*_n \in ds\bigr) = \frac{1}{E(\exp(\theta_n S_n))} e^{\theta_n s} P(S_n \in ds).$$

Write

$$\begin{aligned}
P\bigl(y \le S^*_n \le y+h\bigr) &= \frac{E(\exp(\tilde{\theta}_{n,j}S_{n,j}))}{E(\exp(\theta_n S_n))} \frac{1}{E(\exp(\tilde{\theta}_{n,j}S_{n,j}))} \\
&\quad \times \int_{[y,y+h]} \exp\{(\theta_n - \tilde{\theta}_{n,j})s\} e^{\tilde{\theta}_{n,j}s} P(S_n \in ds).
\end{aligned}$$

By (2.33), the factor $\exp\{(\theta_n - \tilde{\theta}_{n,j})s\}$ above is uniformly bounded away from zero over $s \in [y, y+h]$, $|y| \le nM$ and j. Furthermore,

$$\frac{E(\exp(\tilde{\theta}_{n,j}S_{n,j}))}{E(\exp(\theta_n S_n))} = \exp\bigl\{-n\bigl[\psi_n(\theta_n) - \psi_n(\tilde{\theta}_{n,j})\bigr] - \phi_Z(\tilde{\theta}_{n,j}\beta_{n,j})\bigr\},$$

and it follows from (2.34) and uniform boundedness of the argument of ϕ_Z that the ratio above is bounded away from zero over n and j. We conclude that for some $c > 0$, for all n large enough and $|y| \le nM$,

$$\begin{aligned}
P\bigl(y \le S^*_n \le y+h\bigr) &\ge c\frac{1}{E(\exp(\tilde{\theta}_{n,j}S_{n,j}))} \int_{[y,y+h]} e^{\tilde{\theta}_{n,j}s} P(S_n \in ds) \\
&\ge cP(0 \le \beta_{n,j}Z \le h-1) P\bigl(y \le S^*_{n,j} \le y+1\bigr).
\end{aligned}$$

Since $\beta_{n,j}$ is uniformly bounded, we can choose h large enough such that $P(0 \le \beta_{n,j}Z \le h-1)$ is uniformly bounded away from zero, and (2.36) follows from (2.37). □

The next, final, lemma is a major ingredient in the proof of Theorem 1.

LEMMA 2.3. *For a fixed $k \ge 1$, the conditional law of $(S_n - n\varepsilon, Z_{-k}, \dots \dots, Z_k, Z_{n-k}, \dots, Z_{n+k})$ given E_0 converges weakly, as $n \to \infty$, to the law of*

$$(T^*, Z^-_{-k}, \dots, Z^-_k, Z^+_{-k}, \dots, Z^+_k).$$

PROOF. Consider the following truncated version of S_n:

$$\bar{S}_n = \sum_{j=k+1}^{n-k-1} (A^+_{n-1-j} + A^-_j) Z_j + \sum_{j=n+k}^{\infty} (A^-_j - A^-_{j-n}) Z_j + \sum_{j=k+1}^{\infty} (A^+_{j+n-1} - A^+_{j-1}) Z_{-j},$$

$n \ge 2(k+1)$. We claim that there exists $c_n > 0$ such that for any $x \in \mathbb{R}$ and any sequence $x_n \to x$,

$$P(\bar{S}_n \ge n\varepsilon + x_n) \sim c_n e^{-x\tau(\varepsilon)/A}. \tag{2.38}$$

To show this we proceed as in the proof of Lemma 2.1. Let

$$\bar{\psi}_n(t) = n^{-1} E(e^{t\bar{S}_n}), \quad t \in \mathbb{R}.$$

Repeating the argument in Lemma 2.1 shows that

$$\lim_{n\to\infty} \bar{\psi}''_n(t) = A^2 \varphi''_Z(At) \tag{2.39}$$

locally uniformly in $t \in \mathbb{R}$, and that for large n there exists $\bar{\theta}_n > 0$ such that

$$\begin{aligned} &\bar{\psi}'_n(\bar{\theta}_n) = \varepsilon, \\ &\lim_{n\to\infty} \bar{\theta}_n = A^{-1}\tau(\varepsilon) \end{aligned} \tag{2.40}$$

and

$$P(\bar{S}_n \ge n\varepsilon) \sim \frac{C}{\sqrt{n}} \exp(-n(\bar{\theta}_n\varepsilon - \bar{\psi}_n(\bar{\theta}_n))), \quad n \to \infty, \tag{2.41}$$

with C as in Lemma 2.1. The same argument shows that, if $x_n \to x$, then for large n there exists $\bar{\theta}_{n,x} > 0$ such that

$$\begin{aligned} &\bar{\psi}'_n(\bar{\theta}_{n,x}) = \varepsilon + n^{-1}x_n, \\ &\lim_{n\to\infty} \bar{\theta}_{n,x} = A^{-1}\tau(\varepsilon) \end{aligned} \tag{2.42}$$

and

$$P(\bar{S}_n \ge n\varepsilon + x_n) \sim \frac{C}{\sqrt{n}} \exp(-n(\bar{\theta}_{n,x}(\varepsilon + n^{-1}x_n) - \bar{\psi}_n(\bar{\theta}_{n,x}))), \quad n \to \infty. \tag{2.43}$$

The mean value theorem applied to (2.40) and (2.42), together with (2.39) implies that

$$A^2\varphi''_Z(\tau(\varepsilon))(\bar{\theta}_{n,x} - \bar{\theta}_n) = n^{-1}x_n + o(n^{-1}) = n^{-1}x + o(n^{-1}).$$

We use this fact together with another application of the mean value theorem. Keeping in mind (2.40) and the locally uniform boundedness of the second derivative implied by (2.39), we see that

$$\bar{\psi}_n(\bar{\theta}_{n,x}) - \bar{\psi}_n(\bar{\theta}_n) = \frac{1}{n} \frac{\varepsilon x}{A^2\varphi''_Z(\tau(\varepsilon))} + o(n^{-1}).$$

Putting together the above two displays, we see that

$$(\bar{\theta}_{n,x} - \bar{\theta}_n)\varepsilon - \bar{\psi}_n(\bar{\theta}_{n,x}) + \bar{\psi}_n(\bar{\theta}_n) = o(n^{-1}),$$

which in conjunction with (2.43) establishes (2.38) with c_n given by the right-hand side of (2.41).

Finally, for $t \geq 0$ and a compact rectangle $R \subset \mathbb{R}^{4k+2}$,

$$\begin{aligned}
&P([S_n - n\varepsilon \geq t, (Z_{-k}, \dots, Z_k, Z_{n-k}, \dots, Z_{n+k}) \in R] \cap E_0) \\
&\quad = P(S_n - n\varepsilon \geq t, (Z_{-k}, \dots, Z_k, Z_{n-k}, \dots, Z_{n+k}) \in R) \\
&\quad = \int_{(x_{-k},\dots,x_k,y_{-k},\dots,y_k)\in R} P\Biggl(\bar{S}_n \geq n\varepsilon + t - \sum_{j=0}^{k}(A^+_{n-1-j} + A^-_j)x_j \\
&\qquad - \sum_{j=1}^{k}(A^+_{j+n-1} - A^+_{j-1})x_{-j} - \sum_{j=1}^{k}(A^+_{j-1} + A^-_{n-j})y_{-j} - \sum_{j=0}^{k}(A^-_{n+j} - A^-_j)y_j\Biggr) \\
&\qquad \times F_Z(dx_{-k})\dots F_Z(dx_k)F_Z(dy_{-k})\dots F_Z(dy_k) \\
&\quad \sim c_n e^{-\tau(\varepsilon)t/A} \int_{(x_{-k},\dots,x_k,y_{-k},\dots,y_k)\in R} \exp\Biggl\{\frac{\tau(\varepsilon)}{A}\Biggl(\sum_{j=0}^{k}(A^+ + A^-_j)x_j \\
&\qquad + \sum_{j=1}^{k}(A^+ - A^+_{j-1})x_{-j} + \sum_{j=1}^{k}(A^+_{j-1} + A^-)y_{-j} + \sum_{j=0}^{k}(A^- - A^-_j)y_j\Biggr)\Biggr\} \\
&\qquad \times F_Z(dx_{-k})\dots F_Z(dx_k)F_Z(dy_{-k})\dots F_Z(dy_k)
\end{aligned} \tag{2.44}$$

as $n \to \infty$. In order to justify the asymptotic equivalence above, note that for each fixed $x_{-k}, \dots, x_k, y_{-k}, \dots, y_k$, c_n^{-1} times the integrand of (2.44) converges, by (2.38), to

$$\begin{aligned}
&\exp\Biggl\{\frac{\tau(\varepsilon)}{A}\Biggl(-t + \sum_{j=0}^{k}(A^+ + A^-_j)x_j + \sum_{j=1}^{k}(A^+ - A^+_{j-1})x_{-j} \\
&\qquad + \sum_{j=1}^{k}(A^+_{j-1} + A^-)y_{-j} + \sum_{j=0}^{k}(A^- - A^-_j)y_j\Biggr)\Biggr\}.
\end{aligned}$$

The absolute value of each of the variables $x_{-k}, \dots, x_k, y_{-k}, \dots, y_k$ in the rectangle R has a finite upper bound. Replacing each of these variables by the corresponding upper bound of its absolute value and using (2.38) once again, provides a bound to use in the dominated convergence theorem.

Continuing to keep k an arbitrary fixed positive integer, we claim that, as $n \to \infty$,

$$\begin{aligned}
P(E_0) \sim c_n \int_{\mathbb{R}^{4k+2}} &\exp\Biggl\{\frac{\tau(\varepsilon)}{A}\Biggl(\sum_{j=0}^{k}(A^+ + A^-_j)x_j + \sum_{j=1}^{k}(A^+ - A^+_{j-1})x_{-j} \\
&\qquad + \sum_{j=1}^{k}(A^+_{j-1} + A^-)y_{-j} + \sum_{j=0}^{k}(A^- - A^-_j)y_j\Biggr)\Biggr\} \\
&\qquad \times F_Z(dx_{-k})\dots F_Z(dx_k)F_Z(dy_{-k})\dots F_Z(dy_k).
\end{aligned} \tag{2.45}$$

Once this has been established, the claim of the lemma will follow from (2.44) and (2.45), completing the argument. To prove (2.45), we notice that by (2.38) and Fatou's lemma,

$$
\begin{aligned}
&\liminf_{n\to\infty} c_n^{-1} P(E_0) \\
&\quad \geq \int_{\mathbb{R}^{4k+2}} \exp\biggl\{\frac{\tau(\varepsilon)}{A}\biggl(\sum_{j=0}^{k}(A^+ + A_j^-)x_j + \sum_{j=1}^{k}(A^+ - A_{j-1}^+)x_{-j} \\
&\qquad + \sum_{j=1}^{k}(A_{j-1}^+ + A^-)y_{-j} + \sum_{j=0}^{k}(A^- - A_j^-)y_j\biggr)\biggr\} \\
&\qquad \times F_Z(dx_{-k})\dots F_Z(dx_k)F_Z(dy_{-k})\dots F_Z(dy_k).
\end{aligned}
\tag{2.46}
$$

By Lemma 2.2 the sequence of the conditional laws of $(Z_{-k},\dots,Z_k,Z_{n-k},\dots\dots,Z_{n+k})$ given E_0 is tight in $\mathbb{R}^{4k+2}$. Let ν be a subsequential limit of this sequence. It follows from the above inequality and (2.44) with $t=0$ that

$$
\nu(R) \leq P\bigl((Z_{-k}^-,\dots,Z_k^-,Z_{-k}^+,\dots,Z_k^+)\in R\bigr)
$$

for any compact rectangle R in $\mathbb{R}^{4k+2}$, which can only happen if ν is, in fact, the law of the random vector $(Z_{-k}^-,\dots,Z_k^-,Z_{-k}^+,\dots,Z_k^+)$. Therefore, (2.46) must hold as an equality. $\square$

We are now ready to prove the first of our main theorems.

PROOF OF THEOREM 1. We start with showing that for every fixed $k\geq 1$, conditionally on E_0 as $n\to\infty$,

$$
\begin{aligned}
&(S_n - n\varepsilon, X_{-k},\dots,X_k,X_{n-k},\dots,X_{n+k}) \\
&\qquad \Rightarrow (T^*, U_{-k}^-,\dots,U_k^-,U_{-k}^+,\dots,U_k^+),
\end{aligned}
\tag{2.47}
$$

with (U_k^-) and (U_k^+) defined in (2.11). For all $i\geq 1$ let

$$
X_m^{(i)} = \sum_{j=-i}^{i} a_j Z_{m-j}, \quad m\in\mathbb{Z}.
$$

Lemma 2.3 implies that for a fixed i,

$$
\bigl(S_n - n\varepsilon, X_{-k}^{(i)},\dots,X_k^{(i)},X_{n-k}^{(i)},\dots,X_{n+k}^{(i)}\bigr)
$$

converges weakly as $n\to\infty$, conditionally on E_0, to

$$
\bigl(T^*, U_{-k}^{-(i)},\dots,U_k^{-(i)},U_{-k}^{+(i)},\dots,U_k^{+(i)}\bigr),
$$

where

$$
U_m^{\pm(i)} = \sum_{j=-i}^{i} a_j Z_{m-j}^{\pm}, \quad m\in\mathbb{Z}.
$$

Note that by Lemma 2.2, for every $\delta>0$,

$$
\sup_{n\geq 1}\sup_{m\in\mathbb{Z}} P\bigl(\bigl|X_m^{(i)} - X_m\bigr| > \delta|E_0\bigr) \leq \frac{1}{\delta}\Bigl[\sup_{n\geq 1,m\in\mathbb{Z}} E\bigl(|Z_m||E_0\bigr)\Bigr]\sum_{|j|>i}|a_j| \to 0
$$

as $i\to\infty$. Since the two series in (2.11) converge in probability, the claim (2.47) follows from Theorem 3.2 in Billingsley (1999).

Notice that

$$
\begin{aligned}
1(E_j) &= 1\left(\sum_{i=j}^{j+n-1} X_i \ge n\varepsilon\right) \\
&= \begin{cases} 1\left(S_n - n\varepsilon - \sum_{i=0}^{j-1} X_i + \sum_{i=n}^{n+j-1} X_i \ge 0\right) & \text{if } j \ge 0, \\ 1\left(S_n - n\varepsilon + \sum_{i=j}^{-1} X_i - \sum_{i=n+j}^{n-1} X_i \ge 0\right) & \text{if } j < 0. \end{cases}
\end{aligned}
$$

We conclude by (2.47) and the continuous mapping theorem that for $K \ge 1$,

$$
\begin{aligned}
&\big(1(E_j), j = -K, \dots, K\big) \\
&\quad \Rightarrow \left(1\left(T^* + \sum_{i=j}^{-1} U_i^- - \sum_{i=j}^{-1} U_i^+ \ge 0\right), j = -K, \dots, -1,\right. \\
&\qquad \left. 1\left(T^* - \sum_{i=0}^{j-1} U_i^- + \sum_{i=0}^{j-1} U_i^+ \ge 0\right), j = 1, \dots, K\right) \\
&\qquad = \big(V_j(\varepsilon), j = -K, \dots, K\big)
\end{aligned}
$$

as $n \to \infty$, where the law of the vector in the left hand side is computed conditionally on E_0. Indeed, the continuity of the exponential random variable T^* means that the boundary of the K-dimensional set above has limiting probability zero. This proves (2.16). □

Finally, we prove our second main result.

PROOF OF THEOREM 2. We take the stochastic process

$$
W_\varepsilon^+(t) = \frac{\tau(\varepsilon)}{A} \sum_{i=0}^{[t\varepsilon^{-2}]} (U_i^- - U_i^+), \quad t \ge 0
$$

and its version in the direction of the negative time and prove for them joint functional weak convergence. The claim of the theorem will then follow by an application of the continuous mapping theorem. We start with some variance calculations. For a large m,

$$
\begin{aligned}
\mathrm{Var}\left(\sum_{i=0}^{m} U_i^-\right) &= \mathrm{Var}\left(\sum_{i=0}^{m}\sum_{k=-\infty}^{-1} a_{i-k} Z_k^-\right) + \mathrm{Var}\left(\sum_{i=0}^{m}\sum_{k=0}^{\infty} a_{i-k} Z_k^-\right) \\
&= \sum_{k=1}^{\infty}\left(\sum_{i=k}^{m+k} a_i\right)^2 \mathrm{Var}(Z_{-k}^-) + \sum_{k=0}^{m} (A_{m-k}^+ + A_k^-)^2 \mathrm{Var}(Z_k^-) \\
&\quad + \sum_{k=m+1}^{\infty}\left(\sum_{i=-k}^{m-k} a_i\right)^2 \mathrm{Var}(Z_k^-).
\end{aligned}
$$

It is elementary that

$$
\mathrm{Var}(Z_k^-) \to \sigma_Z^2 \quad \text{as } \varepsilon \to 0
$$

uniformly in $k \in \mathbb{Z}$. Therefore,

$$
\sum_{k=0}^{m} (A_{m-k}^+ + A_k^-)^2 \mathrm{Var}(Z_k^-) \sim \sigma_Z^2 \sum_{k=0}^{m} (A_{m-k}^+ + A_k^-)^2 \sim m\sigma_Z^2 A^2
$$

as $\varepsilon \to 0$, $m \to \infty$. Furthermore,

$$\sum_{k=1}^{\infty}\left(\sum_{i=k}^{m+k} a_i\right)^2 \operatorname{Var}(Z_{-k}^-) \sim \sigma_Z^2 \sum_{k=1}^{\infty}\left(\sum_{i=k}^{m+k} a_i\right)^2 = o(m)$$

as $\varepsilon \to 0$, $m \to \infty$, with the last statement an easy consequence of the absolute summability of (a_i). Similarly,

$$\sum_{k=m+1}^{\infty}\left(\sum_{i=-k}^{m-k} a_i\right)^2 \operatorname{Var}(Z_k^-) = o(m)$$

as $\varepsilon \to 0$, $m \to \infty$.

The same argument shows that we also have

$$\sum_{k=0}^{m}(A_{m-k}^+ + A_k^-)^2 \operatorname{Var}(Z_k^+) \sim \sigma_Z^2 \sum_{k=0}^{m}(A_{m-k}^+ + A_k^-)^2 \sim m\sigma_Z^2 A^2,$$

$$\sum_{k=1}^{\infty}\left(\sum_{i=k}^{m+k} a_i\right)^2 \operatorname{Var}(Z_{-k}^+) \sim \sigma_Z^2 \sum_{k=1}^{\infty}\left(\sum_{i=k}^{m+k} a_i\right)^2 = o(m),$$

$$\sum_{k=m+1}^{\infty}\left(\sum_{i=-k}^{m-k} a_i\right)^2 \operatorname{Var}(Z_k^+) = o(m)$$

as $\varepsilon \to 0$, $m \to \infty$. We write

$$\text{(2.48)} \quad W_\varepsilon^+(t) = W_\varepsilon(t)$$

$$\text{(2.49)} \qquad = \frac{\tau(\varepsilon)}{A}\sum_{i=0}^{[t\varepsilon^{-2}]}(U_i^- - U_i^+)$$

$$= \left[\frac{\tau(\varepsilon)}{A}\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=-\infty}^{-1} a_{i-k}Z_k^- - \frac{\tau(\varepsilon)}{A}\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=-\infty}^{-1} a_{i-k}Z_k^+\right]$$

$$+ \left[\frac{\tau(\varepsilon)}{A}\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=0}^{[t\varepsilon^{-2}]} a_{i-k}Z_k^- - \frac{\tau(\varepsilon)}{A}\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=0}^{[t\varepsilon^{-2}]} a_{i-k}Z_k^+\right]$$

$$+ \left[\frac{\tau(\varepsilon)}{A}\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=[t\varepsilon^{-2}]+1}^{\infty} a_{i-k}Z_k^- - \frac{\tau(\varepsilon)}{A}\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=[t\varepsilon^{-2}]+1}^{\infty} a_{i-k}Z_k^+\right]$$

$$=: W_\varepsilon^{(1)}(t) + W_\varepsilon^{(2)}(t) + W_\varepsilon^{(3)}(t), \qquad \varepsilon > 0, t \geq 0.$$

For typographical convenience we will omit the superscript in W_ε^+ for now and bring it back at a certain point later on. The assumption $EZ = 0$ and (2.9) imply that, as $\varepsilon \to 0$,

$$\text{(2.50)} \qquad \varepsilon/A \sim \tau(\varepsilon)\varphi_Z''(0) = \sigma_Z^2\tau(\varepsilon).$$

We have, therefore, verified that $\operatorname{Var}(W_\varepsilon^{(j)}(t)) \to 0$ as $\varepsilon \to 0$ for every t and $j = 1, 3$, so for every such j,

$$\text{(2.51)} \qquad W_\varepsilon^{(j)}(t) - E(W_\varepsilon^{(j)}(t)) \to 0 \quad \text{in probability as } \varepsilon \to 0.$$

Furthermore, for every t, as $\varepsilon \to 0$,

$$\operatorname{Var}(W_\varepsilon^{(2)}(t)) \to \frac{2t}{A^2\sigma_Z^2}.$$

A similar calculation shows that for $0 \le s \le t$,

$$\lim_{\varepsilon\to 0} \operatorname{Cov}\big(W_\varepsilon^{(2)}(s), W_\varepsilon^{(2)}(t)\big) = \frac{2s}{A^2\sigma_Z^2}.$$

Observe next that the third absolute moment of both (Z_k^-) and (Z_k^+) is bounded uniformly in ε and k. Therefore, the Lindeberg condition is satisfied by the triangular array defined by any finite linear combination of the type $\theta_1 W_\varepsilon^{(2)}(t_1) + \cdots + \theta_d W_\varepsilon^{(2)}(t_d)$. Applying the Lindeberg central limit theorem (see, e.g., Theorem 27.2 in Billingsley (1995)) and the Cramér–Wold device we conclude that the finite-dimensional distributions of $W_\varepsilon^{(2)}(t) - E(W_\varepsilon^{(2)}(t))$ converge to those of $(A\sigma_Z)^{-1}\sqrt{2}B_t$, where B_t is a standard Brownian motion. It follows from (2.51) that the finite-dimensional distributions of $W_\varepsilon(t) - E(W_\varepsilon(t))$ converge to the same limit.

Next, let $0 \le s < t$. If $\varepsilon^2 > (t-s)$, then for any $s \le r \le t$ either

$$W_\varepsilon(t) - E\big(W_\varepsilon(t)\big) = W_\varepsilon(r) - E\big(W_\varepsilon(r)\big) \quad \text{a.s.}$$

or

$$W_\varepsilon(s) - E\big(W_\varepsilon(s)\big) = W_\varepsilon(r) - E\big(W_\varepsilon(r)\big) \quad \text{a.s.},$$

so that

$$\begin{aligned} &E\big[\big(W_\varepsilon(t) - W_\varepsilon(r) - E\big(W_\varepsilon(t) - W_\varepsilon(r)\big)\big)^2 \\ &\quad \times \big(W_\varepsilon(r) - W_\varepsilon(s) - E\big(W_\varepsilon(r) - W_\varepsilon(s)\big)\big)^2\big] = 0. \end{aligned}$$

Suppose now that $\varepsilon^2 \le (t-s)$. We have

$$\begin{aligned} &E\big[\big(W_\varepsilon(t) - W_\varepsilon(s) - E\big(W_\varepsilon(t) - W_\varepsilon(s)\big)\big)^4\big] \\ &\quad \le 12E\big[\big(W_\varepsilon^{(1)}(t) - W_\varepsilon^{(1)}(s) - E\big(W_\varepsilon^{(1)}(t) - W_\varepsilon^{(1)}(s)\big)\big)^4\big] \\ &\qquad + 12E\big[\big(W_\varepsilon^{(2)}(t) - W_\varepsilon^{(2)}(s) - E\big(W_\varepsilon^{(2)}(t) - W_\varepsilon^{(2)}(s)\big)\big)^4\big] \\ &\qquad + 12E\big[\big(W_\varepsilon^{(3)}(t) - W_\varepsilon^{(2)}(s) - E\big(W_\varepsilon^{(3)}(t) - W_\varepsilon^{(2)}(s)\big)\big)^4\big]. \end{aligned}$$

For a positive constant C independent of ε, s, t, that may change from appearance to appearance,

$$\begin{aligned} &E\Bigg[\Bigg(\frac{\tau(\varepsilon)}{A}\Bigg(\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=0}^{[t\varepsilon^{-2}]} a_{i-k}\big(Z_k^- - E\big(Z_k^-\big)\big) \\ &\qquad - \sum_{i=0}^{[s\varepsilon^{-2}]}\sum_{k=0}^{[s\varepsilon^{-2}]} a_{i-k}\big(Z_k^- - E\big(Z_k^-\big)\big)\Bigg)\Bigg)^4\Bigg] \\ &\quad \le C\varepsilon^4 E\Bigg(\sum_{k=0}^{[s\varepsilon^{-2}]}\big(Z_k^- - E\big(Z_k^-\big)\big)\sum_{i=[s\varepsilon^{-2}]-k}^{[t\varepsilon^{-2}]-k} a_i\Bigg)^4 \\ &\qquad + C\varepsilon^4 E\Bigg(\sum_{k=[s\varepsilon^{-2}]+1}^{[t\varepsilon^{-2}]}\big(Z_k^- - E\big(Z_k^-\big)\big)\sum_{i=-k}^{[t\varepsilon^{-2}]-k} a_i\Bigg)^4. \end{aligned} \tag{2.52}$$

Since the fourth moments of (Z_k^-) are bounded uniformly in ε and k, and the coefficients (a_i) are absolutely summable, the first term in the right hand side can be bounded by

$$
\begin{aligned}
&C\varepsilon^4 \sum_{k=0}^{[s\varepsilon^{-2}]} E\big(Z_k^- - E(Z_k^-)\big)^4 \Bigg(\sum_{i=[s\varepsilon^{-2}]-k}^{[t\varepsilon^{-2}]-k} |a_i|\Bigg)^4 \\
&\quad + \Bigg[C\varepsilon^2 \sum_{k=0}^{[s\varepsilon^{-2}]} E\big(Z_k^- - E(Z_k^-)\big)^2 \Bigg(\sum_{i=[s\varepsilon^{-2}]-k}^{[t\varepsilon^{-2}]-k} |a_i|\Bigg)^2\Bigg]^2 \\
&\le C\varepsilon^4 \sum_{k=0}^{\infty}\sum_{i=-\infty}^{\infty} |a_{i-k}| + \Bigg[C\varepsilon^2 \sum_{k=0}^{\infty} \sum_{i=[s\varepsilon^{-2}]-k}^{[t\varepsilon^{-2}]-k} |a_i|\Bigg]^2.
\end{aligned}
$$

Since

$$
\begin{aligned}
\sum_{k=0}^{\infty} \sum_{i=[s\varepsilon^{-2}]-k}^{[t\varepsilon^{-2}]-k} |a_i| &\le \big([t\varepsilon^{-2}] - [s\varepsilon^{-2}]\big) \sum_{i=-\infty}^{\infty} |a_i| \\
&\le \big(t\varepsilon^{-2} - (s\varepsilon^{-2} - 1)\big) \sum_{i=-\infty}^{\infty} |a_i| \le 2(t-s)\varepsilon^{-2} \sum_{i=-\infty}^{\infty} |a_i|,
\end{aligned}
$$

we conclude that the first term in the right hand side of (2.52) is bounded by $C(t-s)^2$. In a similar way one can show that the second term in the right hand side of (2.52) is also bounded by $C(t-s)^2$.

The same argument shows that

$$
\begin{aligned}
E\Bigg[\Bigg(\frac{\tau(\varepsilon)}{A}\Bigg(&\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=0}^{[t\varepsilon^{-2}]} a_{i-k}\big(Z_k^+ - E(Z_k^+)\big) \\
&- \sum_{i=0}^{[s\varepsilon^{-2}]}\sum_{k=0}^{[s\varepsilon^{-2}]} a_{i-k}\big(Z_k^+ - E(Z_k^+)\big)\Bigg)\Bigg)^4\Bigg] \\
&\le C(t-s)^2,
\end{aligned}
$$

so that

$$
E\big[\big(W_\varepsilon^{(2)}(t) - W_\varepsilon^{(2)}(s) - E\big(W_\varepsilon^{(2)}(t) - W_\varepsilon^{(2)}(s)\big)\big)^4\big] \le C(t-s)^2.
$$

In the same way we can check that

$$
E\big[\big(W_\varepsilon^{(j)}(t) - W_\varepsilon^{(1)}(s) - E\big(W_\varepsilon^{(j)}(t) - W_\varepsilon^{(1)}(s)\big)\big)^4\big] \le C(t-s)^2, \quad j = 1, 3,
$$

so we conclude that

$$
E\big[\big(W_\varepsilon(t) - W_\varepsilon(s) - E\big(W_\varepsilon(t) - W_\varepsilon(s)\big)\big)^4\big] \le C(t-s)^2 \tag{2.53}
$$

if $\varepsilon^2 \le (t-s)$. By the Cauchy–Schwarz inequality, for any $0 \le s \le r \le t$ we have

$$
\begin{aligned}
&E\big[\big(W_\varepsilon(t) - W_\varepsilon(r) - E\big(W_\varepsilon(t) - W_\varepsilon(r)\big)\big)^2 \\
&\quad \times \big(W_\varepsilon(r) - W_\varepsilon(s) - E\big(W_\varepsilon(r) - W_\varepsilon(s)\big)\big)^2\big] \\
&\le C(t-s)^2
\end{aligned}
$$

if $\varepsilon^2 \le (t-s)$. Appealing to Theorem 13.5 of Billingsley (1999) we conclude that for any fixed T, the family

$$\{(W_\varepsilon(t) - E(W_\varepsilon(t)) : 0 \le t \le T) : \varepsilon > 0\}$$

is tight in $D[0,T]$ endowed with the Skorohod J_1 topology. Therefore, as $\varepsilon \to 0$,

$$(W_\varepsilon(t) - E(W_\varepsilon(t)) : 0 \le t \le T) \Rightarrow ((A\sigma_Z)^{-1}\sqrt{2}B_t : 0 \le t \le T), \tag{2.54}$$

in $D[0,T]$. Furthermore,

$$\begin{aligned} &E(W_\varepsilon(t)) \\ &\quad = \frac{\tau(\varepsilon)}{A}\Bigg[\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=-\infty}^{-1} a_{i-k}Z_k^- + \sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=0}^{[t\varepsilon^{-2}]} a_{i-k}Z_k^- + \sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=[t\varepsilon^{-2}]+1}^{\infty} a_{i-k}Z_k^- \\ &\qquad - \sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=-\infty}^{-1} a_{i-k}Z_k^+ - \sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=0}^{[t\varepsilon^{-2}]} a_{i-k}Z_k^+ - \sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=[t\varepsilon^{-2}]+1}^{\infty} a_{i-k}Z_k^+\Bigg]. \end{aligned}$$

Clearly, $|EZ_k^\pm| = O(\tau(\varepsilon))$ uniformly in ε and $k \in \mathbb{Z}$. Therefore,

$$\left|\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=-\infty}^{-1} a_{i-k}EZ_k^-\right| \le O(\tau(\varepsilon))\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=i+1}^{\infty}|a_k| = o(\tau(\varepsilon)\varepsilon^{-2})$$

uniformly in t in a compact set. Similarly,

$$\left|\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=[t\varepsilon^{-2}]+1}^{\infty} a_{i-k}Z_k^-\right| = o(\tau(\varepsilon)\varepsilon^{-2}),$$

and by the same argument,

$$\left|\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=-\infty}^{-1} a_{i-k}EZ_k^+\right| = o(\tau(\varepsilon)\varepsilon^{-2}),$$

$$\left|\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=[t\varepsilon^{-2}]+1}^{\infty} a_{i-k}Z_k^+\right| = o(\tau(\varepsilon)\varepsilon^{-2}),$$

all uniformly in t in a compact set.

Finally, $EZ_k^- \sim \tau(\varepsilon)\sigma_Z^2(A^+ + A_k^-)/A$ as $\varepsilon \to 0$ uniformly in $k \ge 0$, so

$$\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=0}^{[t\varepsilon^{-2}]} a_{i-k}EZ_k^- \sim A\tau(\varepsilon)\sigma_Z^2 t\varepsilon^{-2}.$$

Similarly, $EZ_k^+ \sim \tau(\varepsilon)\sigma_Z^2(A^- - A_k^-)/A$ as $\varepsilon \to 0$ uniformly in $k \ge 0$, so

$$\sum_{i=0}^{[t\varepsilon^{-2}]}\sum_{k=0}^{[t\varepsilon^{-2}]} a_{i-k}EZ_k^+ = o(\tau(\varepsilon)\varepsilon^{-2}),$$

all uniformly in t in a compact set. We conclude by (2.50) that for all $\varepsilon > 0$ small enough,

$$E(W_\varepsilon(t)) \ge \frac{t}{2A^2\sigma_Z^2}, \quad t \ge 1 \tag{2.55}$$

and

$$E\big(W_\varepsilon(t)\big) \to \frac{t}{A^2\sigma_Z^2}, \quad \varepsilon \to 0,$$

uniformly in t in a compact set. Since the addition in $D[0,T]$ is continuous at continuous functions, this along with (2.54) shows that

$$\big(W_\varepsilon^+(t) : 0 \le t \le T\big) \Rightarrow \big((A\sigma_Z)^{-1}\sqrt{2}B_t^+ + (A\sigma_Z)^{-2}t : 0 \le t \le T\big), \tag{2.56}$$

in $D[0,T]$ as $\varepsilon \to 0$. Notice that we have brought back the superscript in W_ε^+ omitted since (2.48), and we have also added a superscript to the standard Brownian motion B^+.

Clearly we can also define

$$W_\varepsilon^-(t) = \frac{\tau(\varepsilon)}{A} \sum_{i=-[t\varepsilon^{-2}]}^{-1} \big(U_i^+ - U_i^-\big), \quad t \ge 0, \tag{2.57}$$

and use the same argument (or, even simpler, just appeal to time inversion) to show that for any $T > 0$,

$$\big(W_\varepsilon^-(t) : 0 \le t \le T\big) \Rightarrow \big((A\sigma_Z)^{-1}\sqrt{2}B_t^- + (A\sigma_Z)^{-2}t : 0 \le t \le T\big), \tag{2.58}$$

in $D[0.T]$ as $\varepsilon \to 0$, where B^- is a standard Brownian motion.

We claim that, in fact, we have joint convergence

$$\begin{aligned} &\big(\big(W_\varepsilon^+(t), : 0 \le t \le T\big), \big(W_\varepsilon^-(t), : 0 \le t \le T\big)\big) \\ &\quad \Rightarrow \big(\big((A\sigma_Z)^{-1}\sqrt{2}B_t^+ + (A\sigma_Z)^{-2}t : 0 \le t \le T\big), \\ &\qquad \big((A\sigma_Z)^{-1}\sqrt{2}B_t^- + (A\sigma_Z)^{-2}t : 0 \le t \le T\big)\big), \end{aligned} \tag{2.59}$$

in $D[0.T] \times D[0.T]$ as $\varepsilon \to 0$, where the standard Brownian motions in the right hand side are independent. To see this, recall that the only term in (2.48) that contributes to the randomness in the weak limit of $(W_\varepsilon^+(t), : 0 \le t \le T)$ is the term $(W_\varepsilon^{(2)}(t), : 0 \le t \le T)$, which is a function of $(Z_k^\pm)$ with $k \ge 0$. An identical argument shows that the only term in the same expansion of $(W_\varepsilon^-(t), : 0 \le t \le T)$ that contributes to the randomness in the limit is a function of $(Z_k^\pm)$ with $k < 0$. Since the random variables $(Z_k^\pm)$ are independent, we obtain the claimed joint weak convergence in (2.59), with independent components in the limit.

For any real λ the function $\varphi : D[0,T] \to \mathbb{R}$ defined by

$$\varphi(f) = \int_0^T 1\big(\lambda \ge f(t)\big)\,dt$$

is continuous at any continuous f that takes value λ only on a set of measure 0. Therefore, for any such λ, by the continuous mapping theorem,

$$\begin{aligned} \int_0^T 1\big(\lambda \ge W_\varepsilon^\pm(t)\big)\,dt &\Rightarrow \int_0^T 1\big(\lambda \ge (A\sigma_Z)^{-1}\sqrt{2}B_t^\pm + (A\sigma_Z)^{-2}t\big)\,dt \\ &\overset{d}{=} \int_0^T 1\big(\lambda \ge \sqrt{2}B_{(A\sigma_Z)^{-2}t}^\pm + (A\sigma_Z)^{-2}t\big)\,dt \\ &= A^2\sigma_Z^2 \int_0^{(A\sigma_Z)^2 T} 1\big(\lambda \ge \sqrt{2}B_t^\pm + t\big)\,dt. \end{aligned}$$

Noticing that we can write

$$V_j^+(\varepsilon) = 1\left(T_0 \ge \frac{\tau(\varepsilon)}{A} \sum_{i=0}^{j-1} \big(U_i^- - U_i^+\big)\right), \quad j \ge 1,$$

where T_0 is a standard exponential random variable independent of the collection $(Z_j^u : j \in \mathbb{Z}, u = + \text{ or } -)$, we conclude that for any $T > 0$,

$$\varepsilon^2 \sum_{j=0}^{[T\varepsilon^{-2}]} V_j^+(\varepsilon) = \int_0^T \mathbf{1}\big(T_0 \ge W_\varepsilon^+(t)\big)\, dt - V_{[T\varepsilon^{-2}]}^+\big(T - \varepsilon^2[T\varepsilon^{-2}]\big)$$

$$\Rightarrow A^2\sigma_Z^2 \int_0^{(A\sigma_Z)^2 T} \mathbf{1}\big(T_0 \ge \sqrt{2}B_t^+ + t\big)\, dt.$$

It is clear that the latter integral converges a.s., as $T \to \infty$, to the integral prescribed in the theorem. Therefore, we can appeal to Theorem 3.2 in Billingsley (1999), which requires us to show that for any $\delta > 0$,

$$\lim_{T\to\infty} \limsup_{\varepsilon\to 0} P\left(\varepsilon^2 \sum_{j=[T\varepsilon^{-2}]+1}^{\infty} V_j^+(\varepsilon) > \delta\right) = 0. \tag{2.60}$$

However, by Markov's inequality

$$P\left(\varepsilon^2 \sum_{j=[T\varepsilon^{-2}]+1}^{\infty} V_j^+(\varepsilon) > \delta\right) \le \varepsilon^2\delta^{-1} \sum_{j=[T\varepsilon^{-2}]+1}^{\infty} P\big(T_0 \ge W_\varepsilon^+((j-1)\varepsilon^2)\big)$$

$$\le \varepsilon^2\delta^{-1} \sum_{j=[T\varepsilon^{-2}]+1}^{\infty} P\left(W_\varepsilon^+((j-1)\varepsilon^2) \le \frac{(j-1)\varepsilon^2}{4A^2\sigma_Z^2}\right)$$

$$+ \varepsilon^2\delta^{-1} \sum_{j=[T\varepsilon^{-2}]+1}^{\infty} \exp\left\{-\frac{(j-1)\varepsilon^2}{4A^2\sigma_Z^2}\right\}.$$

By (2.55) and (2.53) we have for some positive constant C,

$$P\left(W_\varepsilon^+((j-1)\varepsilon^2) \le \frac{(j-1)\varepsilon^2}{4A^2\sigma_Z^2}\right) \le C\varepsilon^{-4}(j-1)^{-2}$$

for all $j > [T\varepsilon^{-2}]$, $T \ge 1$ and $\varepsilon > 0$ small enough. This estimate suffices to establish (2.60).

Note that this argument also shows that that for small $\varepsilon > 0$, $ED_\varepsilon^+ < \infty$, so $D_\varepsilon^+ < \infty$ a.s.

Since a similar argument can be applied to $\varepsilon^2 \sum_{j=0}^{[T\varepsilon^{-2}]} V_j^-(\varepsilon)$, the proof is complete. □

3. Discussion. As is usually the case with large deviations, the limiting distributions obtained in (2.16) and (2.17) of Theorem 1 depend on the underlying model through the distribution of the noise variables F_Z and the coefficients (a_i). This dependence largely disappears in Theorem 2 where the limiting distribution depends only on the noise variance σ_Z^2 and the sum of the coefficients A. This can be understood by viewing the case of a small overshoot ε as approaching the regime of moderate deviations. Indeed, in the case of moderate deviations one expects that the central limit behaviour becomes visible and leads to a collapse of the model ingredients necessary to describe the limit to a bare minimum consisting of second order information.

This naturally leads to the question of a difference of how large deviations cluster between the short memory moving average processes and long memory moving average processes. It is common to say that the coefficients of the moving average process (1.1) with long memory are square summable but not absolutely summable. Assuming certain regularity of the coefficients (a_i) (e.g., their regular variation), one can show that for any fixed $j \ge 1$ and $\varepsilon > 0$,

$$\lim_{n\to\infty} P\big(E_j(n,\varepsilon) \,|\, E_0(n,\varepsilon)\big) = 1,$$

so one expects infinitely many events $(E_j(n, \varepsilon))$ to happen once $E_0(n, \varepsilon)$ does. This necessitates different limiting procedures when studying large deviations clustering of such long memory processes. It is important to note that for these long memory moving average processes, in the notation of (2.22), $n = o(\mathrm{Var}(S_n))$, so one can view the events $(E_j(n, \varepsilon))$ as moderate deviation events and not large deviation events. Indeed, it turns out that a natural limiting procedure leads to a collapse in the amount of information about the model needed to describe the limit, which is similar to the situation with Theorem 2 in the present, short memory case. This is described in details in Chakrabarty and Samorodnitsky (2023).

APPENDIX: SOME USEFUL FACTS

The following nonlogarithmic version of a large deviation statement and a related estimate are from Chaganty and Sethuraman (1993).

THEOREM 3. *Let $\{T_n\}$ be a sequence of random variables with*

$$E\left(e^{zT_n}\right) < \infty \quad \text{for any } z \in \mathbb{R}, n \geq 1.$$

For a sequence $\{a_n\}$ of positive numbers with

$$\lim_{n\to\infty} a_n = \infty \tag{A.1}$$

we denote

$$\psi_n(z) = a_n^{-1} \log E\left(e^{zT_n}\right), \quad z \in \mathbb{R}, n \geq 1.$$

Let $\{m_n\}$ be a bounded sequence of real numbers. Assume that there exists a bounded positive sequence $\{\tau_n\}$ satisfying

$$\psi_n'(\tau_n) = m_n, \quad n \geq 1, \tag{A.2}$$

$$a_n^{-1/2} = o(\tau_n), \quad n \to \infty, \tag{A.3}$$

and such that for all fixed $\delta, \lambda > 0$,

$$\sup_{\delta \leq |t| \leq \lambda\tau_n} \left| \frac{1}{E(e^{\tau_n T_n})} E\left(e^{(\tau_n + it)T_n}\right) \right| = o\left(a_n^{-1/2}\right), \quad n \to \infty, \tag{A.4}$$

(*with the supremum of the empty set defined as zero*). *Furthermore, assume that*

$$\sup_{n\geq 1, z\in[-a,a]} \left|\psi_n(z)\right| < \infty \quad \text{for any } a > 0 \tag{A.5}$$

and that

$$\inf_{n\geq 1} \psi_n''(\tau_n) > 0. \tag{A.6}$$

(a) *Under the above assumption,*

$$P\left(a_n^{-1} T_n \geq m_n\right) \sim \frac{1}{\tau_n \sqrt{2\pi a_n \psi_n''(\tau_n)}} \exp\left\{-a_n\left(m_n \tau_n - \psi_n(\tau_n)\right)\right\}, \quad n \to \infty. \tag{A.7}$$

(b) *Let*

$$b_n = \tau_n \sqrt{a_n \psi_n''(\tau_n)},$$

and let T_n^* be a random variable with the law

$$P\left(T_n^* \in du\right) = \frac{1}{E(e^{\tau_n T_n})} e^{u\tau_n} P(T_n \in du).$$

Then

$$\sup_{n\geq 1, y\in\mathbb{R}} b_n P\left(y \leq \tau_n T_n^* \leq y + 1\right) < \infty.$$

PROOF. The first part of the theorem is Theorem 3.3 in Chaganty and Sethuraman (1993). Furthermore, Lemmas 3.1 and 3.2 ibid. show that the hypotheses (2.7) and (2.8) of Theorem 2.3 therein hold, and the second part of Theorem 3 follows from (2.9) of that paper. □

Acknowledgments. Two anonymous referees provided the authors with incredibly useful and insightful comments. We are very grateful.

Funding. The second author was partially supported by NSF grant DMS-2015242 and AFOSR grant FA9550-22-1-0091 at Cornell University.

REFERENCES

BASRAK, B. and SEGERS, J. (2009). Regularly varying multivariate time series. *Stochastic Process. Appl.* **119** 1055–1080. MR2508565 https://doi.org/10.1016/j.spa.2008.05.004

BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR1324786

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. Wiley, New York. MR1700749 https://doi.org/10.1002/9780470316962

CHAGANTY, N. R. and SETHURAMAN, J. (1993). Strong large deviation and local limit theorems. *Ann. Probab.* **21** 1671–1690. MR1235434

CHAKRABARTY, A. and SAMORODNITSKY, G. (2023). Clustering of large deviations in moving average processes: The long memory regime. *Stochastic Process. Appl.* **163** 387–423. MR4612303 https://doi.org/10.1016/j.spa.2023.06.009

DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. Springer, New York. MR1619036 https://doi.org/10.1007/978-1-4612-5320-4

DEUSCHEL, J.-D. and STROOCK, D. W. (1989). *Large Deviations*. *Pure and Applied Mathematics* **137**. Academic Press, Boston, MA. MR0997938

DJELLOUT, H. and GUILLIN, A. (2001). Large and moderate deviations for moving average processes. *Ann. Fac. Sci. Toulouse Math.* (6) **10** 23–31. MR1928987

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (2003). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin. 4th corrected printing.

GOYAL, A. and WELCH, I. (2003). Predicting the equity premium with dividend ratios. *Manage. Sci.* **49** 639–654.

JIANG, T. F., WANG, X. C. and RAO, M. B. (1992). Moderate deviations for some weakly dependent random processes. *Statist. Probab. Lett.* **15** 71–76. MR1190249 https://doi.org/10.1016/0167-7152(92)90287-F

KULIK, R. and SOULIER, P. (2020). *Heavy-Tailed Time Series*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR4174389 https://doi.org/10.1007/978-1-0716-0737-4

MIKOSCH, T. and NAGAEV, A. V. (1998). Large deviations of heavy-tailed sums with applications in insurance. *Extremes* **1** 81–110. MR1652936 https://doi.org/10.1023/A:1009913901219

RYBSKI, D., BUNDE, A., HAVLIN, S. and VON STORCH, H. (2006). Long-term persistence in climate and detection problem. *Geophys. Res. Lett.* **33**. https://doi.org/10.1029/2005GL02559

SAMORODNITSKY, G. (2016). *Stochastic Processes and Long Range Dependence*. *Springer Series in Operations Research and Financial Engineering*. Springer, Cham. MR3561100 https://doi.org/10.1007/978-3-319-45575-4