

## Edgeworth expansions for volatility models

Moritz Jirak\*

### Abstract

Motivated from option and derivative pricing, this note develops Edgeworth expansions both in the Kolmogorov and Wasserstein metric for many different types of discrete time volatility models and their possible transformations. This includes, among others, Hölder-type functions of (augmented) Garch processes of any order, iterated random functions or Volterra-processes.

**Keywords:** edgeworth expansions; volatility models; weak dependence; Garch.

**MSC2020 subject classifications:** 60F05; 60F25; 60G10.

Submitted to EJP on November 3, 2021, final version accepted on September 11, 2023.

## 1 Introduction

Consider a strictly stationary sequence  $(X_k)_{k \in \mathbb{Z}}$  of real-valued random variables with  $\mathbb{E}X_k = 0$  and  $\mathbb{E}X_k^2 < \infty$ . If the sequence exhibits weak dependence in a certain sense, then the distribution of

$$n^{-1/2}S_n, \text{ where } S_n = X_1 + X_2 + \cdots + X_n,$$

is asymptotically normal, see for instance [48] and the references therein. This fact has made the central limit theorem one of the most important tools in probability theory and statistics. On the other hand, it was already noticed by Chebyshev [11] and Edgeworth [20] that normal approximations can be improved in terms of (Edgeworth) expansions  $\Psi_n$ , implying the approximation

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(S_n \leq x\sqrt{n}) - \Psi_n(x)| = o(n^{-\frac{1}{2}}) \quad (\text{or even better}) \quad (1.1)$$

in the Kolmogorov metric, where we used the usual  $o$ ,  $O$  Landau symbols, and

$$\Psi_n(x) = \Phi(x/s_n) + (\kappa_n/s_n)^3(1 - x^2/s_n^2)\phi(x/s_n), \quad (1.2)$$

where  $\Phi$  and  $\phi$  are the distribution function and density of the standard normal distribution, respectively, and

$$s_n^2 = n^{-1}\mathbb{E}S_n^2, \quad \kappa_n^3 = n^{-3/2}\mathbb{E}S_n^3. \quad (1.3)$$

---

\*University of Vienna, Austria. E-mail: moritz.jirak@univie.ac.at

Motivated by applications in actuarial science, Cramér gave rigorous proofs in [12], and ever since, Edgeworth expansions have been an indispensable tool in actuarial science and finance, see the discussion below and for instance [15] for overviews and further historic comments. They are also studied in the context of dynamical system theory and Markovian setups, see e.g. [25, 36, 42] and the references therein for some more recent results, and also [32, 45, 53] for a general, weakly dependent framework and [29, 51, 44, 60] for some continuous time results.<sup>1</sup> On the other hand, in a very influential work, Efron [21] broadened the view on resampling techniques (e.g. bootstrapping) and demonstrated their significant superior performance compared to normal approximations, see [22, 34, 46] for an overview. Not surprisingly, among the key tools for analysing, and, in particular, showing superiority of resampling methods, are again Edgeworth expansions.

One of our main motivations here stems more from actuarial, finance and risk management considerations. Very prominent models in these areas in a discrete time setting are (augmented) Garch processes, e.g. [9, 17, 19, 27, 37]. It is well-known that already the Black-Scholes formula for option pricing has serious shortcomings, see e.g. [31]. To address these problems, a common and quite successful approach is to employ more complex models and use (Edgeworth) expansions or large deviation techniques to salvage comparatively simple and easy to evaluate formulas, both in the continuous time and discrete time setting. A large body of literature with quite different techniques has evolved around these ideas, for more details, we refer for instance to [1, 3, 18, 23, 26, 28, 30, 38, 43, 59, 62], and the references therein. To illustrate this further, consider the (standard) model

$$\log P_{nt} = \sum_{k=1}^{nt} V_{k-1} \frac{\epsilon_k}{\sqrt{n}} - \frac{1}{2n} \sum_{k=1}^{nt} h(V_{k-1}^2), \quad t \in [0, 1], \quad (1.4)$$

where  $\log P_{nt}$  is the log-price of some derivative under a martingale measure for appropriate function  $h$ . Here,  $\epsilon_k$  are the innovations and  $V_k^2$  is some volatility process with  $V_k^2 \in \sigma(\epsilon_k, \epsilon_{k-1}, \dots)$ . Then, one seeks an approximation of the type

$$\mathbb{E}f(\log P_{nt}) \approx \int f(x) \Lambda_{nt}(dx),$$

where the function  $f(x)$  describes the pay-off of some option and  $\Lambda_{nt}$  is a ‘convenient’, signed measure. Since we may express  $\log P_{nt}$  in terms of a normalised (centred) sum of dependent random variables, the connection to Edgeworth expansions is obvious. A rather prominent example in this context are European Put-Options, where  $f(x) = \max\{K - e^x, 0\}$  for some strike price  $K > 0$ . Now  $|f(x) - f(y)| \leq K|x - y|$ , and hence  $f(x)$  is Lipschitz-continuous. The latter is true for many options, and thus another natural metric to measure the quality of Edgeworth expansions is in terms of the Wasserstein metric  $W_1$ , using Kantorovich-Rubinstein duality. In the latter, a Gamma-type approximation is more convenient in our setting, see (2.5) for more details and definitions. A nice feature of the  $W_1$ -metric is that it provides a uniform bound for different Lipschitz functions without the necessity of deriving explicit bounds and constants. Another interesting application of the  $W_1$ -metric are approximation bounds for the conditional value at risk, see for instance [54].

Our main contribution here is to establish the validity of Edgeworth expansions both in the Kolmogorov and Wasserstein metric for various classes of popular volatility type models. This includes in particular – for the first time, to the best of the author’s

<sup>1</sup>These lists of references are by no means complete and only represent a mere fraction of the vast literature in this area.

knowledge – functions of (augmented) Garch(p, q)-processes of any order. Previously, only the case  $p = q = 1$  appears to have been treated in the literature. In addition, it seems that there are almost no results concerning Edgeworth expansions for weakly dependent processes in terms of the Wasserstein distance in general. From a more technical side, our key novelty is to handle higher Fourier frequencies with the help of small-ball conditions.

This note is structured as follows. In Section 2, we present the setup and our main global results. We then show how to use these to derive the validity of Edgeworth expansions both in the Kolmogorov and Wasserstein metric for Hölder type functionals of various volatility type models, see Section 3.1 (augmented Garch processes), Section 3.2 (iterated random functions), Section 3.3 (linear processes) and Section 3.4 (Volterra processes) for details.

## 2 Setup and main global results

For a random variable  $X$ , we write  $\mathbb{E}X$  for expectation,  $\|X\|_p$  for  $(\mathbb{E}|X|^p)^{1/p}$ ,  $p \geq 1$ , and sometimes  $\mathbb{E}_{\mathcal{H}}X = \mathbb{E}[X|\mathcal{H}]$  for conditional expectation, and in analogy  $\mathbb{P}_{\mathcal{H}}(\cdot)$  for conditional probabilities.  $\lesssim, \gtrsim, (\approx)$  denote (two-sided) inequalities involving a multiplicative constant. For  $a, b \in \mathbb{R}$ , we put  $a \vee b = \max\{a, b\}$ ,  $a \wedge b = \min\{a, b\}$ . For two random variables  $X, Y$ , we write  $X \stackrel{d}{=} Y$  for equality in distribution. For an i.i.d. sequence  $(\epsilon_k)_{k \in \mathbb{Z}}$ , let  $\mathcal{E}_k = \sigma(\epsilon_j, j \leq k)$ , and  $\mathcal{E}_k^+ = \sigma(\epsilon_j, j \geq k)$ .

Consider a sequence of real-valued, measurable random variables  $X_1, \dots, X_n$ . It is well known (cf. [55]), that this sequence can be assumed to satisfy  $X_k \in \mathcal{E}_k$ , that is, we have

$$X_k = g_k(\epsilon_k, \epsilon_{k-1}, \dots) \tag{2.1}$$

for some measurable functions  $g_k$ ,<sup>2</sup> where  $(\epsilon_k)_{k \in \mathbb{Z}}$  is a sequence of independent and identically distributed random variables. For notational convenience, we sometimes assume  $g_k = g$ , that is, the function  $g$  does not depend on  $k$ . Such processes are usually referred to as (time-homogenous) Bernoulli-shift processes.

Representation (2.1) allows to give simple, yet very efficient and general dependence conditions. Following [57], let  $(\epsilon'_k)_{k \in \mathbb{Z}}$  be an independent copy of  $(\epsilon_k)_{k \in \mathbb{Z}}$  on the same probability space, and define the ‘filter’  $\theta_k^{(l,*)}$  as

$$\theta_k^{(l,*)} = (\epsilon_k, \epsilon_{k-1}, \dots, \epsilon'_{k-l}, \epsilon'_{k-l-1}, \epsilon'_{k-l-2}, \dots). \tag{2.2}$$

We write

$$\theta_k^* = \theta_k^{(k,*)} = (\epsilon_k, \epsilon_{k-1}, \dots, \epsilon'_0, \epsilon'_{-1}, \epsilon'_{-2}, \dots),$$

$X_k^{(l,*)} = g_k(\theta_k^{(l,*)})$ , and  $X_k^* = X_k^{(k,*)}$ . As dependence measure, one may then define

$$\vartheta_l^*(p) = \sup_{k \in \mathbb{Z}} \|X_k - X_k^{(l,*)}\|_p. \tag{2.3}$$

If  $g = g_k$  does not depend on  $k$  (time-homogenous case),  $\vartheta_l^*(p)$  simplifies to

$$\vartheta_l^*(p) = \|X_l - X_l^*\|_p.$$

In the literature, a considerable number of other notions of mixing and (weak) dependence coefficients have been established (cf. [13, 14]). A particular popular one is

<sup>2</sup>In fact,  $g_k$  can be selected as a map from  $\mathbb{R}^k$  to  $\mathbb{R}$ .

$\alpha$ -mixing, which, however, can be difficult to verify and may fail to hold even for models as simple as AR(1)-processes, see for instance [2, 14]. On the other hand, notions such as (2.3) are not new and go back at least to [39] and [7]. In case of exponential decay, they have also been employed in [33], see also [32].

Our basic condition regarding weak dependence is now the following.

**Assumption 2.1.** For  $p > 3$ ,  $(X_k)_{k \in \mathbb{Z}}$  is stationary and satisfies

- (A1)  $\mathbb{E}|X_k|^p < \infty$ ,  $\mathbb{E}X_k = 0$ ,
- (A2)  $\sum_{k=1}^{\infty} k^2 \vartheta_k^*(p) < \infty$ ,
- (A3)  $\mathfrak{s}^2 > 0$ , where  $\mathfrak{s}^2 = \sum_{k \in \mathbb{Z}} \mathbb{E}X_0 X_k$ .

Our requirement of stationarity is more a convenience condition, and can be replaced with quenched or locally stationary setups. Also note that  $\mathfrak{s}^2 < \infty$  is implied by (A2), see for instance [40].

Subject to the above conditions, Edgeworth expansions were recently characterised in [40] in terms of the *Berry-Esseen characteristic*, which essentially encodes the higher order Fourier frequencies of magnitude at least  $\sqrt{n}$ , see Section 4 for more details. Unfortunately though, the results in [40] tell nothing about the (non-trivial) nature of the Berry-Esseen characteristic or how to control it. However, as is well known, validity of Edgeworth expansions is not for free and requires at the very least some non-lattice condition to control Fourier frequencies of order  $\sqrt{n}$  and larger. In general, this is a highly non-trivial task, see e.g. [33, 44, 62] for results based on the conditional Cramér-method. A considerable difficulty that has to be overcome with this approach is that conditioning has to be performed both on the ‘past’ and ‘future’, making it difficult to implement and validate, see for instance [35, p. 901] for a comment. In this note, we bypass some of the problems attached to a conditional Cramér-type argument by utilising a method based on small-ball probabilities. To this end, we require the following regularity assumptions regarding the underlying distribution.

**Assumption 2.2.** Consider  $\mathcal{E}'_l = \sigma(e'_k, k \leq l)$ , which is independent of  $\mathcal{E}_l$ . For any  $\delta > 0$  and  $l \in \mathbb{Z}$ , there exists a family of random variables  $(X_k^+) \in \mathcal{G}_l \stackrel{\text{def}}{=} \sigma(\mathcal{E}_l^+ \cup \mathcal{E}'_l)$ , such that

- (B1)  $\mathbb{P}\left(\sum_{k \geq l} \mathbb{E}_{\mathcal{G}_l} |X_k - X_k^+| \leq \delta\right) > 0$ ,
- (B2)  $\mathbb{E}|\mathbb{E}_{\mathcal{E}_{l-1}} e^{i\xi X_l}| < 1$  for any  $\xi \neq 0$ .

Observe that condition (B2) is a non-lattice condition, and will be easy to verify in case of our applications. The key to our results is (B1) though, which is a small-ball condition. While it is not true in general, we show below that it does hold for a huge class of volatility models and their Hölder-continuous transformations (and even more), requiring only little additional regularity conditions. Note that validity of (B1) does not imply that  $\sum_{k \geq l} \mathbb{E}_{\mathcal{E}_l^+} |X_k - X_k^+|$  or even  $X_k$  is non-lattice. Our first result is the following.

**Theorem 2.3.** Assume that Assumptions 2.1 and 2.2 hold. Then

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(S_n \leq x\sqrt{n}) - \Psi_n(x)| = o(n^{-1/2}).$$

In particular, there exists  $b_n \rightarrow \infty$  and  $\delta > 0$  such that for any (fixed)  $a > 0$

$$\sup_{\xi \in [a, b_n]} \left| \mathbb{E} e^{i\xi S_n} \right| \leq C n^{-1/2-\delta}, \tag{2.4}$$

where  $C > 0$  does not depend on  $n$ .

Recall that Assumption 2.2 (only) rules out non-lattice distributions, hence an error term of magnitude  $o(n^{-1/2})$  is the maximum we can obtain here in general. This can be seen by considering the case where  $(X_k)_{k \in \mathbb{Z}}$  are i.i.d., in which case **(B1)** is trivially true and **(B2)** is equivalent with  $X \stackrel{d}{=} X_k$  being non-lattice, see for instance Lemma 4 in [24].

Next, we turn to the Wasserstein metric  $W_1$ . For two probability measures  $\mathbb{P}_1, \mathbb{P}_2$ , let  $\mathcal{L}(\mathbb{P}_1, \mathbb{P}_2)$  be the set of all probability measures on  $\mathbb{R}^2$  with marginals  $\mathbb{P}_1, \mathbb{P}_2$ . The Wasserstein metric (of order one) is defined as the minimal coupling  $\mathbb{L}^1$ -distance, that is,

$$W_1(\mathbb{P}_1, \mathbb{P}_2) = \inf \left\{ \int_{\mathbb{R}^2} |x - y| \mathbb{P}(dx, dy) : \mathbb{P} \in \mathcal{L}(\mathbb{P}_1, \mathbb{P}_2) \right\}. \tag{2.5}$$

Let  $\mathbb{V}_n$  be the (signed) measure induced by  $\Psi_n$ . Then a priori, the distance  $W_1(\mathbb{P}_1, \mathbb{V}_n)$  is not defined in general. In [8], generalized transport distances are introduced that also allow for signed measures. In order to maintain the original definition in terms of couplings, we follow [40] and replace  $\Psi_n$  with a probability measure that is induced by a sequence of i.i.d. random variables. Let  $Z$  be a zero mean Gaussian random variable  $\mathcal{N}(0, \sigma^2)$  with variance  $\sigma^2 = s_n^2$ , and  $G$  follow a Gamma distribution  $\Gamma(\alpha, \beta)$  with shape parameter  $\alpha = s_n^2 \beta$  and rate  $\beta = 2s_n^2(\sqrt{n}\kappa_n)^{-3}$  (if  $\kappa_n \neq 0$ ), independent of  $Z$  (recall (1.3) for  $s_n, \kappa_n$ ). Then, for  $1 \leq k \leq n$ , let

$$M_k \stackrel{d}{=} \begin{cases} (Z + G - \mathbb{E}G)/\sqrt{2} & \text{if } \kappa_n^3 > 0, \\ (Z - G + \mathbb{E}G)/\sqrt{2} & \text{if } \kappa_n^3 < 0, \\ Z & \text{if } \kappa_n^3 = 0 \end{cases} \tag{2.6}$$

be i.i.d., and denote by  $\mathbb{P}_{L_n}$  the probability measure induced by  $L_n = n^{-1/2} \sum_{k=1}^n M_k$ . Observe that  $\mathbb{E}L_n^2 = s_n^2$  and  $\mathbb{E}L_n^3 = \kappa_n^3$ . Also note that  $\sqrt{2}L_n \stackrel{d}{=} Z_0 \pm (G_0 - \mathbb{E}G_0)$ , with  $Z_0, G_0$  independent and  $Z_0 \stackrel{d}{=} \mathcal{N}(0, s_n^2)$ ,  $G_0 \stackrel{d}{=} \Gamma(n\alpha, \beta/\sqrt{n})$  with  $\alpha, \beta$  as above. We now have the following result.

**Theorem 2.4.** *Grant Assumption 2.1, and suppose that (2.4) holds. Then*

$$W_1(\mathbb{P}_{S_n/\sqrt{n}}, \mathbb{P}_{L_n}) = o(n^{-1/2}).$$

Due to Theorem 2.3, an immediate consequence is the following.

**Corollary 2.5.** *Grant Assumptions 2.1 and 2.2. Then*

$$W_1(\mathbb{P}_{S_n/\sqrt{n}}, \mathbb{P}_{L_n}) = o(n^{-1/2}).$$

### 3 Volatility models

Over the past decades, the following basic model has emerged as a key building block in econometrics, finance and actuarial science for an underlying process  $(Y_k)_{k \in \mathbb{Z}}$ :

$$Y_k = \epsilon_k V_{k-1}, \quad k \in \mathbb{Z}, \tag{3.1}$$

where  $(\epsilon_k)_{k \in \mathbb{Z}}$  are i.i.d. and  $V_k \in \mathcal{E}_k$  is some volatility process. The actual models for asset prices (and related) are then obtained by appropriate transformations, compensations or by passing on to the limit to obtain stochastic differential equations. A sheer endless amount of models and processes of this type have been established and discussed. Since our focus here lies on discrete time, we mention for instance [9, 17, 19, 27, 37] which, however, presents only an almost infinitesimal fraction of the literature.

Our basic setup here is the following. We consider processes  $Y_k$  of type (3.1), where we assume that  $V_k \in \mathcal{E}_k$  is stationary. In the sequel, it will be convenient to use  $\varepsilon \stackrel{d}{=} \epsilon_k$ . Let

$$X_k = f(Y_k) + h_n(V_{k-1}) \quad (f, h_n \text{ specified below, see (3.7)), \tag{3.2}$$

and  $S_k = \sum_{j=1}^k X_j$ . Consider the price of an asset  $P_k = e^{n^{-1/2}S_k}$ . If we select  $h_n \equiv 0$  and  $f$  such that

$$f(x) = x - \frac{x^2}{2\sqrt{n}} - \frac{2x^3}{3n} + r(x), \quad r(Y_k) = O_{\mathbb{P}}(n^{-3/2}), \tag{3.3}$$

then, given sufficiently many (exponential) moments,<sup>3</sup> we obtain

$$\mathbb{E}[P_{k+1} | \mathcal{E}_k] = P_k + O_{\mathbb{P}}(n^{-2}). \tag{3.4}$$

Hence  $P_k$  is almost a martingale, and the actual error in (3.4) can be made arbitrarily small by further specifying  $r$ . On the other hand, formally letting

$$X_k = Y_k - h_n(V_{k-1}), \quad h_n(V_{k-1}) = n^{1/2} \log \mathbb{E}[e^{n^{-1/2}Y_k} | V_{k-1}], \tag{3.5}$$

it follows that  $P_k$  is a martingale. Note in particular that if  $\varepsilon$  has a standard Gaussian distribution  $\mathcal{N}(0, 1)$ , then we get the well-known form  $X_k = Y_k - V_{k-1}^2/(2n^{1/2})$ . More generally, a formal Taylor expansion around zero with  $\mathbb{E}\varepsilon = 0$  leads to

$$h_n(V_{k-1}) = \frac{V_{k-1}^2}{2\sqrt{n}} \mathbb{E}\varepsilon^2 + O_{\mathbb{P}}(n^{-1}). \tag{3.6}$$

We wish to apply Theorem 2.3 to  $X_k = f(Y_k) + h_n(V_{k-1})$ , where we assume  $\mathbb{E}X_k = 0$ . Having in mind (3.3), (3.6), but also statistical applications (power transformations), we consider functions  $f, h_n$  satisfying the generalised Hölder condition

$$|F(x) - F(y)| \leq L|x - y|^\beta (1 + |x|^\alpha + |y|^\alpha), \quad \alpha \geq 0, \beta, L > 0. \tag{3.7}$$

For future reference, we denote this class with  $\mathcal{H}(L, \alpha, \beta)$ . Moreover, we assume that

$$\lim_{n \rightarrow \infty} \mathbb{E}|h_n(V_k)| = 0, \tag{3.8}$$

which in light of (3.6) is a mild condition.

Our basic condition to verify Assumption 2.2 is the following.

**Assumption 3.1.** Assume that  $Y_k$  is stationary, is of the form (3.1),  $f, h_n \in \mathcal{H}(L, \alpha, \beta)$  (uniformly in  $n$ ),  $\varsigma^2 = \sum_{k \in \mathbb{Z}} (\mathbb{E}f(Y_k)f(Y_0) - \mathbb{E}^2f(Y_0)) > 0$ , and in addition:

- (V1) For any  $\delta > 0$ , we have  $\mathbb{P}(|\varepsilon| \leq \delta) > 0$ .
- (V2) There exists a compact set  $\mathcal{V}$  with  $\mathbb{P}(V_0 \in \mathcal{V}) > 0$ , such that for any  $v \in \mathcal{V}$  and  $\xi \neq 0$ , we have  $|\mathbb{E}e^{i\xi f(\varepsilon v)}| < 1$ .

**Remark 3.2.** If we strengthen (V1) a little bit, we can simplify condition (V2): Suppose that  $\mathbb{P}(|\varepsilon| \leq \delta, \varepsilon \neq 0) > 0$  for all  $\delta > 0$ . Then it is sufficient to demand that  $f$  is non-constant in a surrounding of zero.

We now discuss three particular, yet quite general class of models. We note, however, that the method of proof is more general and can also be applied to other classes of models. At this stage, it should also be mentioned that (Edgeworth) expansions for martingales are also present in the literature (cf. [49, 50, 60, 61]). However, it seems that the results presented here require different arguments and cannot be derived from the literature.

---

<sup>3</sup>Using smooth truncation functions, moment conditions can be drastically reduced here while maintaining the result.

### 3.1 Augmented Garch

Arch, Garch and augmented Garch models have had a huge impact both in theory and practice, see [9, 19] and [27]. In one of its more general forms, it can be stated as

$$\Lambda(V_k^2) = \sum_{i=1}^p g_i(\epsilon_{k-i}) + \sum_{i=1}^q c_i(\epsilon_{k-i})\Lambda(V_{k-i}^2), \tag{3.9}$$

where  $g_i, c_i \geq 0$  are positive functions, and  $g_{i_0} \geq \omega > 0$  for some  $1 \leq i_0 \leq p$ . Motivated from the Box-Cox transformation, the function  $\Lambda(\cdot)$  typically comprises the cases  $\Lambda(x) = \log x$  or  $\Lambda(x) = x^\lambda, \lambda > 0$ . We will consider the latter. For  $q \geq 1$ , an important quantity is

$$\gamma_c = \sum_{i=1}^{\tau} \|c_i(\epsilon_0)\|_q, \quad \text{with } \tau = \max\{p, q\},$$

where we replace possible undefined  $c_i$  (and  $g_i$ ) with zero. If  $\gamma_c < 1$ , then  $(V_k)_{k \in \mathbb{Z}}$  is stationary. In particular, one can show the representation

$$V_k^{2\lambda} = \sum_{m=1}^{\infty} \sum_{1 \leq l_1, \dots, l_m \leq \tau} g_{l_m}(\epsilon_{k-l_1-\dots-l_m}) \prod_{i=1}^{m-1} c_{l_i}(\epsilon_{k-l_1-\dots-l_i}),$$

see [5] for comments and references on this matter. In particular,  $V_k$  is a time-homogenous Bernoulli-shift process, that is,  $g_k = g$  does not depend on  $k$  in representation (2.1).

In the sequel, we will only treat the case  $\lambda \geq 1/2$ , the case  $\lambda < 1/2$  follows in essentially the same manner, requiring more moment conditions though. Our assumptions are now the following.

**Assumption 3.3.** *Let  $q > \max\{(\alpha \vee \beta)/\lambda, 3(\alpha + \beta)\}$  ( $\alpha, \beta$  as in (3.7)),  $\lambda \geq 1/2$ . For all  $1 \leq i \leq \tau$ , there exists  $\delta > 0$  such that  $\sup_{|x| \leq \delta} c_i(x) \leq \|c_i(\epsilon_0)\|_q$  and  $\sup_{|x| \leq \delta} g_i(x) \leq C_g$ . Moreover, we have  $\gamma_c < 1$  and  $\|g_i(\epsilon)\|_q, \mathbb{E}|\epsilon|^{2\sqrt{q}} < \infty$ .*

We then have the following result.

**Theorem 3.4.** *Grant Assumptions 3.1 and 3.3. Then both Assumptions 2.1 and 2.2 hold. In particular, Theorem 2.3 and Corollary 2.5 apply.*

### 3.2 Iterated random functions

An iterated random function system on the state space  $\mathbb{R}$  is defined as

$$V_k = F_{\epsilon_k}(V_{k-1}), \quad k \in \mathbb{N}, \tag{3.10}$$

where  $\epsilon_k \in \mathbb{R}$  are i.i.d. with  $\epsilon \stackrel{d}{=} \epsilon_k$ . Here,  $F_\epsilon(\cdot) = F(\cdot, \epsilon)$  is the  $\epsilon$ -section of a jointly measurable function  $F : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ . Many dynamical systems, Markov processes and non-linear time series are within this framework, see for instance [16]. For  $y \in \mathbb{R}$ , let  $V_k(y) = F_{\epsilon_k} \circ F_{\epsilon_{k-1}} \circ \dots \circ F_{\epsilon_0}(y)$ , and, given  $y, y' \in \mathcal{Y}$  and  $\gamma > 0$ , we say that the system is  $\gamma$ -moment contracting if

$$\sup_{y, y'} \mathbb{E}|V_k(y) - V_k(y')|^\gamma \leq C\rho^k, \quad \rho \in (0, 1). \tag{3.11}$$

We note that some variations exist in the literature. A key quantity for verifying the moment contraction (3.11) is

$$L_\epsilon = \sup_{y \neq y'} \frac{|F_\epsilon(y) - F_\epsilon(y')|}{|y - y'|}.$$

Essentially (subject to some mild regularity conditions), (3.11) holds if  $\mathbb{E}L_\epsilon^\gamma < \infty$  and  $\mathbb{E} \log L_\epsilon < 0$ , see [16, 58]. Note that (3.11) implies  $\vartheta_k^*(\gamma) \leq C\rho^k$  for  $\gamma \geq 1$ .

**Assumption 3.5.** Let  $q > \max\{3(\alpha + \beta), 1\}$  ( $\alpha, \beta$  as in (3.7)). There exists  $\delta > 0$ , such that  $\sup_{|\varepsilon| \leq \delta} L_\varepsilon < 1$  almost surely. Moreover, we have  $\|L_\varepsilon\|_q < 1$  and  $\mathbb{E}|F_\varepsilon(0)|^q, \mathbb{E}|\varepsilon|^{2Vq} < \infty$ .

We then have the following result.

**Theorem 3.6.** Grant Assumptions 3.1 and 3.5. Then both Assumptions 2.1 and 2.2 hold. In particular, Theorem 2.3 and Corollary 2.5 apply.

### 3.3 Functions of linear processes

Linear processes and their transformations are key in time series analysis, see e.g. [10]. For  $\alpha_k \in \mathbb{R}$  and real-valued functions  $c_i(\cdot)$ , we formally define the linear process

$$G_k = \sum_{i=0}^{\infty} a_i c_i(\epsilon_{k-i}), \quad k \in \mathbb{Z}.$$

Let  $g \in H(L, \gamma, \delta)$ , and consider

$$V_k = g(G_k), \quad k \in \mathbb{Z}.$$

The following is our main assumption.

**Assumption 3.7.** Let  $q > \max\{3(\alpha + \beta)(\gamma + \delta), 1\}$  ( $\alpha, \beta$  as in (3.7)) and suppose that  $\sup_i \|c_i(\varepsilon)\|_q < \infty, \sum_{i \geq 1} i^2 |a_i|^{\beta\delta} < \infty$  and  $\mathbb{E}|\varepsilon|^{2(\alpha\vee 1)} < \infty$ .

We then have the following result.

**Theorem 3.8.** Grant Assumptions 3.1 and 3.7. Then both Assumptions 2.1 and 2.2 hold. In particular, Theorem 2.3 and Corollary 2.5 apply.

### 3.4 Volterra processes

In the study of nonlinear processes, Volterra processes are of fundamental importance, see for instance [4, 52] or [56]. We consider

$$V_k = \sum_{i=1}^{\infty} \sum_{0 \leq j_1 < \dots < j_i} a_k(j_1, \dots, j_i) \epsilon_{k-j_1} \dots \epsilon_{k-j_i},$$

where  $\|\epsilon_k\|_p < \infty$  for  $p \geq 2$ , and  $a_k$  are called the  $k$ -th Volterra kernel. Let

$$A_{k,i} = \sum_{k \in \{j_1, \dots, j_i\}, 0 \leq j_1 < \dots < j_i} |a_k(j_1, \dots, j_i)|.$$

Then by the triangle inequality, there exists a constant  $C$  such that

$$\|V_k - V_k^*\|_p \leq C \sum_{i=1}^{\infty} \|\epsilon_0\|_p^i \sum_{l \geq k} A_{l,i}.$$

We thus require the following assumption.

**Assumption 3.9.** Let  $\mathbb{E}|\varepsilon|^{2Vq} < \infty, q > \max\{3(\alpha + \beta)\}$  ( $\alpha, \beta$  as in (3.7)), such that

$$\sum_{k \geq 1} k^2 \left( \sum_{i=1}^{\infty} \|\epsilon_0\|_q^i \sum_{l \geq k} A_{l,i} \right)^\beta < \infty.$$

We then have the following result.

**Theorem 3.10.** Grant Assumptions 3.1 and 3.9. Then both Assumptions 2.1 and 2.2 hold. In particular, Theorem 2.3 and Corollary 2.5 apply.



#### 4 Proofs of Theorem 2.3 and Theorem 2.4

Throughout the proof, for notational convenience, we assume for simplicity that we are in the time-homogenous Bernoulli-shift case. This means that  $g_k = g$  in (2.1), and, in particular, the quantities in Assumption 2.2 are invariant in  $l \in \mathbb{Z}$ . The general case only requires notational adjustments, the actual proof remains the same.

For  $0 \leq a \leq b$ , define the *Berry-Esseen tail*

$$\mathfrak{T}_a^b(x) = \int_{a \leq |\xi| \leq b} e^{-i\xi x} \mathbb{E}[e^{i\xi S_n / \sqrt{n}}] \left(1 - \frac{|\xi|}{b}\right) \frac{1}{\xi} d\xi, \quad (4.1)$$

which arises naturally in Berry's smoothing inequality. Using  $\mathfrak{T}_a^b$ , for  $a > 0$ , we consider the *Berry-Esseen characteristic*

$$\mathfrak{C}_a = \inf_{b \geq a} \left( \sup_{x \in \mathbb{R}} |\mathfrak{T}_a^b(x)| + 1/b \right). \quad (4.2)$$

For  $m \in \mathbb{N}$ , define the following  $\sigma$ -algebra

$$\mathcal{F}_m = \sigma(\epsilon_{-m+1}, \dots, \epsilon_0, \epsilon_{m+1}, \dots, \epsilon_{2m}, \epsilon_{3m+1}, \dots). \quad (4.3)$$

Moreover, for  $1 \leq j \leq n$ , let  $(\epsilon_k^{(j)})_{k \in \mathbb{Z}}$  be independent copies of  $(\epsilon_k)_{k \in \mathbb{Z}}$ . For each  $2(j-1)m+1 \leq k \leq 2jm$ , define

$$X_{k,m} = f(\epsilon_k, \epsilon_{k-1}, \dots, \epsilon_{k-m+1}, \epsilon_{k-m}^{(j)}, \epsilon_{k-m-1}^{(j)}, \dots), \quad (4.4)$$

and note that  $X_k \stackrel{d}{=} X_{k,m}$ . Finally, let us introduce the quantities

$$A_{i:j} = \sum_{k=i}^j X_{k,m}, \quad B_j = A_{2(j-1)m+1:2jm} = \sum_{k=2(j-1)m+1}^{2jm} X_{k,m}, \quad S_{n,m} = \sum_{k=1}^n X_{k,m}.$$

We recall parts of Theorem 2.7 in [40], which we restate as the following lemma for the sake of reference.

**Lemma 4.1.** *Grant Assumption 2.1. Then there exists  $\delta > 0$ , such that*

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(S_n \leq x\sqrt{n}) - \Psi_n(x)| \lesssim n^{-\frac{1}{2}-\delta} + \mathfrak{C}_{T_n},$$

where  $T_n \geq c\sqrt{n}$  for some  $c > 0$ .

In addition, we require the following technical result in the sequel.

**Lemma 4.2.** *Grant Assumption 2.1. Then there exists  $C > 0$ , such that*

$$\|S_n - S_{n,m}\|_1 \leq C\sqrt{nm}^{-2}.$$

*Proof of Lemma 4.2.* This is an easy consequence of Equation (50) in [41] (observe that  $\sum_{k \geq 1} k^\alpha \|X_k - X_k^{(l,*)}\|_p \lesssim \sum_{k \geq 1} k^\alpha \|X_k - X_k^{(l,*)}\|_p$ ,  $\alpha \geq 0$ ,  $p \geq 1$ , see also Theorem 1 in [47]). Note that the construction of  $X_{k,m}$  is slightly different in [41], but the argument remains equally valid.  $\square$

*Proof of Theorem 2.3.* Let  $\bar{n} = \lfloor n/m \rfloor$  for  $m \in \mathbb{N}$ . The proof works with any choice  $m \asymp n^m$ ,  $m \in (1/2, 1)$ . In order to establish the claim, it suffices to show that for any  $a > 0$ , the Berry-Esseen Characteristic  $\mathfrak{C}_{a\sqrt{\bar{n}}}$  (cf. (4.2)) satisfies

$$\mathfrak{C}_{a\sqrt{\bar{n}}} \lesssim n^{-1/2-\delta}, \quad \delta > 0. \quad (4.5)$$

To this end, we will study  $\mathbb{E}e^{i\xi S_n}$  more closely, subject to Assumptions 2.1 and 2.2. Due to  $|e^{ix} - 1| \leq |x|$ ,  $|e^{ix}| = 1$ , and Lemma 4.2, we have

$$\begin{aligned} |\mathbb{E}e^{i\xi S_n} - \mathbb{E}e^{i\xi S_{n,m}}| &\leq |\xi| \|S_n - S_{n,m}\|_1 \\ &\lesssim |\xi\sqrt{nm}^{-2}| \lesssim |\xi|n^{-1/2-\delta}, \quad \delta > 0. \end{aligned} \tag{4.6}$$

Observe that  $(B_j)_{1 \leq j \leq \bar{n}}$  is conditionally independent with respect to  $\mathcal{F}_m$ , and is a one-dependent sequence in general. Let  $\mathcal{I} = \{1, 3, \dots, 2\lfloor \bar{n}/2 \rfloor - 1\}$ . Then  $(B_j)_{j \in \mathcal{I}}$  is an independent sequence. Hence using  $|e^{ix}| = 1$ , we have

$$\begin{aligned} \|\mathbb{E}_{\mathcal{F}_m} e^{i\xi S_{n,m}}\|_1 &\leq \left\| \prod_{j=1}^{\bar{n}} \mathbb{E}_{\mathcal{F}_m} e^{i\xi B_j} \right\|_1 \\ &\leq \left\| \prod_{j \in \mathcal{I}} \mathbb{E}_{\mathcal{F}_m} e^{i\xi B_j} \right\|_1 = \prod_{j \in \mathcal{I}} \|\mathbb{E}_{\mathcal{F}_m} e^{i\xi B_j}\|_1. \end{aligned} \tag{4.7}$$

Next, put  $\mathcal{H} = \sigma(\mathcal{E}_0 \cup \mathcal{E}'_{m+1} \cup \mathcal{E}^+_{m+1})$ , and observe that for  $1 \leq j \leq \bar{n}$ , we have the identity

$$\|\mathbb{E}_{\mathcal{F}_m} e^{i\xi B_j}\|_1 = \|\mathbb{E}_{\mathcal{H}} e^{i\xi B_1}\|_1.$$

Let  $A^+ \in \mathcal{G}_{m+1}$ , where we recall that  $\mathcal{G}_{m+1} = \sigma(\mathcal{E}'_{m+1} \cup \mathcal{E}^+_{m+1})$ . Since  $|e^{ix}| = 1$ , it follows that

$$\begin{aligned} |\mathbb{E}_{\mathcal{H}} e^{i\xi B_1}| &\leq |\mathbb{E}_{\mathcal{H}} e^{i\xi(B_1 - A^+)}| \\ &\leq |\mathbb{E}_{\mathcal{H}} e^{i\xi A_{1:m}}| + |\mathbb{E}_{\mathcal{H}} e^{i\xi A_{1:m}} (e^{i\xi(A_{m+1:2m} - A^+)} - 1)| \\ &\leq |\mathbb{E}_{\mathcal{H}} e^{i\xi A_{1:m}}| + \mathbb{E}_{\mathcal{H}} |e^{i\xi(A_{m+1:2m} - A^+)} - 1|. \end{aligned}$$

Observe that  $A_{m+1:2m} - A^+$  is independent of  $\mathcal{E}_0$ . Using  $|e^{ix} - 1| \leq |x|$ , we thus further obtain

$$\mathbb{E}_{\mathcal{H}} |e^{i\xi(A_{m+1:2m} - A^+)} - 1| \leq |\xi| \mathbb{E}_{\mathcal{G}_{m+1}} |A_{m+1:2m} - A^+|.$$

Then due to **(B1)**, for any  $\delta > 0$ , there exists a set  $\mathcal{A}_\delta \in \mathcal{G}_{m+1}$  with  $\mathbb{P}(\mathcal{A}_\delta) \geq c_\delta > 0$ , such that for appropriate choice of  $A^+$

$$\mathbb{E}_{\mathcal{G}_{m+1}} |A_{m+1:2m} - A^+| \mathbb{1}_{\mathcal{A}_\delta} \leq \delta. \tag{4.8}$$

Since  $\mathcal{A}_\delta \in \mathcal{H}$ , we obtain from the above and  $|e^{ix}| = 1$

$$\begin{aligned} |\mathbb{E}_{\mathcal{H}} e^{i\xi B_1}| &\leq |\mathbb{E}_{\mathcal{H}} e^{i\xi B_1} \mathbb{1}_{\mathcal{A}_\delta}| + \mathbb{E}_{\mathcal{H}} \mathbb{1}_{\mathcal{A}_\delta^c} \\ &\leq \left( |\mathbb{E}_{\mathcal{H}} e^{i\xi A_{1:m}}| + \delta |\xi| \right) \mathbb{1}_{\mathcal{A}_\delta} + \mathbb{1}_{\mathcal{A}_\delta^c}. \end{aligned}$$

Next, observe

$$\mathbb{E}_{\mathcal{H}} e^{i\xi A_{1:m}} = \mathbb{E}_{\mathcal{E}_0} e^{i\xi A_{1:m}}.$$

Since  $\mathcal{A}_\delta$  is independent from  $\mathcal{E}_0$ , we get

$$\begin{aligned} \mathbb{E} |\mathbb{E}_{\mathcal{H}} e^{i\xi B_1}| &\leq \mathbb{E} \left( |\mathbb{E}_{\mathcal{E}_0} e^{i\xi A_{1:m}}| + \delta |\xi| \right) \mathbb{1}_{\mathcal{A}_\delta} + \mathbb{E} \mathbb{1}_{\mathcal{A}_\delta^c} \\ &= \left( \mathbb{E} |\mathbb{E}_{\mathcal{E}_0} e^{i\xi A_{1:m}}| + \delta |\xi| \right) \mathbb{P}(\mathcal{A}_\delta) + \mathbb{P}(\mathcal{A}_\delta^c). \end{aligned} \tag{4.9}$$

We next deal with  $\mathbb{E}|\mathbb{E}_{\mathcal{E}_0} e^{i\xi A_{1:m}}|$ . To this end, let  $\mathcal{E}_m^{(1)} = \sigma(\epsilon_k^{(1)}, k \leq m)$ . Since  $X_{k,m} \stackrel{d}{=} X_k$  and  $|e^{ix}| = 1$ , we have

$$\begin{aligned} \mathbb{E}|\mathbb{E}_{\mathcal{E}_0} e^{i\xi A_{1:m}}| &\leq \mathbb{E}|\mathbb{E}_{\sigma(\mathcal{E}_{m-1}, \mathcal{E}_{m-1}^{(1)})} e^{i\xi A_{1:m}}| \\ &\leq \mathbb{E}|\mathbb{E}_{\mathcal{E}_{-1}} e^{i\xi X_0}|. \end{aligned}$$

One readily shows that the map  $g : \mathbb{R} \rightarrow [0, 1]$ , given by

$$g(\xi) = \mathbb{E}|\mathbb{E}_{\mathcal{E}_{-1}} e^{i\xi X_0}|,$$

is continuous. Let  $I = [a, b]$ ,  $0 < a < b$ . Then by compactness, there exists  $\xi^* \in I$  such that

$$\sup_{\xi \in I} g(\xi) = g(\xi^*) < 1,$$

where we used **(B2)** for the last inequality. Hence for any  $0 < a < b$ , there exists  $\eta_{ab} < 1$  such that

$$\sup_{\xi \in [a,b]} \mathbb{E}|\mathbb{E}_{\mathcal{E}_{-1}} e^{i\xi X_0}| \leq \eta_{ab} < 1. \tag{4.10}$$

Setting  $\delta = (1 - \eta_{ab})/2b$ , we obtain from (4.9) and (4.10) that for any  $\xi \in [a, b]$ , there exists  $\rho_{ab} < 1$  such that (recall  $\mathbb{P}(\mathcal{A}_\delta) \geq c_\delta > 0$ )

$$\begin{aligned} \mathbb{E}|\mathbb{E}_{\mathcal{H}} e^{i\xi B_1}| &\leq \left( \mathbb{E}|\mathbb{E}_{\mathcal{E}_0} e^{i\xi A_{1:m}}| + \delta|\xi| \right) \mathbb{P}(\mathcal{A}_\delta) + \mathbb{P}(\mathcal{A}_\delta^c) \\ &\leq \frac{1 + \eta_{ab}}{2} \mathbb{P}(\mathcal{A}_\delta) + \mathbb{P}(\mathcal{A}_\delta^c) \leq \rho_{ab}. \end{aligned}$$

Consequently, since  $|\mathcal{I}| \geq \bar{n}/3$  for  $n$  large enough, we get

$$\sup_{\xi \in [a,b]} \prod_{j \in \mathcal{I}} \|\mathbb{E}_{\mathcal{F}_m} e^{i\xi B_j}\|_1 \leq \rho_{ab}^{\frac{\bar{n}}{3}}.$$

Combining (4.7) and the above, we conclude that there exists  $b_n \rightarrow \infty$  such that for any  $a > 0$ , we have

$$\sup_{\xi \in [a,b_n]} \prod_{j \in \mathcal{I}} \|\mathbb{E}_{\mathcal{F}_m} e^{i\xi B_j}\|_1 \lesssim n^{-1/2-\delta'}, \quad \delta' > 0.$$

Using (4.6) and selecting  $b_n \rightarrow \infty$  sufficiently slow, it follows that

$$\sup_{\xi \in [a,b_n]} \left| \mathbb{E} e^{i\xi S_n} \right| \lesssim n^{-1/2-\delta'}, \tag{4.11}$$

and hence

$$\mathfrak{C}_{a,\sqrt{n}} \leq \left( \sup_{x \in \mathbb{R}} |\mathfrak{T}_{a,\sqrt{n}}^{b_n\sqrt{n}}(x)| + 1/(b_n\sqrt{n}) \right) = o(n^{-1/2}),$$

which, by virtue of Lemma 4.1, completes the proof.  $\square$

For the proof of Theorem 2.4, we require some additional notation. For  $e > 0$  and  $f \in \mathbb{N}$  even, let  $G_{e,f}$  be a real valued random variable with density function

$$g_{e,f}(x) = c_f e \left| \frac{\sin(ex)}{ex} \right|^f, \quad x \in \mathbb{R}, \tag{4.12}$$

for some constant  $c_f > 0$  only depending on  $f$ . It is well-known (cf. [6], Section 10) that for even  $f$  the Fourier transform  $\hat{g}_{e,f}$  satisfies

$$\hat{g}_{e,f}(t) = \begin{cases} 2\pi c_f u^{*f}[-e, e](t) & \text{if } |t| \leq ef, \\ 0 & \text{otherwise,} \end{cases} \tag{4.13}$$

where  $u^{*f}[-e, e]$  denotes the  $f$ -fold convolution of the density of the uniform distribution on  $[-e, e]$ , that is  $u[-e, e](t) = \frac{1}{2e} \mathbb{1}_{[-e, e]}(t)$ . For  $f \geq 6$ , let  $(H_k)_{k \in \mathbb{Z}}$  be i.i.d. with  $H_k \stackrel{d}{=} G_{e,f}$  and independent of  $S_n$ . For  $\eta > 0$ , define

$$X_k^\diamond = X_k + \eta H_k - \eta H_{k-1}, \quad S_n^\diamond = \sum_{k=1}^n X_k^\diamond = S_n + \eta H_n - \eta H_0, \tag{4.14}$$

and  $L_n^\diamond$  in analogy.

*Proof of Theorem 2.4.* Using standard properties of the Wasserstein distance and the triangle inequality, we arrive at

$$W_1(\mathbb{P}_{S_n/\sqrt{n}}, \mathbb{P}_{L_n}) \leq W_1(\mathbb{P}_{S_n^\diamond/\sqrt{n}}, \mathbb{P}_{L_n^\diamond}) + 4\eta \mathbb{E}|G_{e,f}|/\sqrt{n}. \tag{4.15}$$

For  $f \geq 6$  and small enough  $e > 0$ , we get, using (2.4), for any  $c, \eta > 0$

$$\sup_{|\xi| \geq c} |\mathbb{E}e^{i\xi S_n^\diamond}| \leq \sup_{c \leq |\xi| \leq ef/\eta} |\mathbb{E}e^{i\xi S_n}| \leq C_{c,\delta} n^{-1/2-\delta}, \quad \delta > 0. \tag{4.16}$$

Following the proof of Theorem 3.6. in [40], we obtain

$$W_1(\mathbb{P}_{S_n^\diamond/\sqrt{n}}, \mathbb{P}_{L_n^\diamond}) \lesssim n^{-p/2+1} + \int_{|x| \leq \tau_n} |\mathbb{P}(S_n^\diamond \leq x\sqrt{n}) - \mathbb{P}(L_n^\diamond \leq x)| dx, \tag{4.17}$$

where  $\tau_n \lesssim \sqrt{\log n}$ . By (4.16), we have for  $\mathfrak{T}_a^b(x)$  (defined with respect to  $S_n^\diamond$ )

$$|\mathfrak{T}_{a\sqrt{n}}^{b\sqrt{n}}(x)| \leq \int_{a \leq |\xi|/\sqrt{n} \leq b} \left| e^{-i\xi x} \mathbb{E}e^{i\xi S_n^\diamond/\sqrt{n}} \left(1 - \frac{|\xi|}{b\sqrt{n}}\right) \frac{1}{\xi} \right| d\xi \leq C_{a,\delta} \frac{\log(nb)}{n^{-1/2-\delta}},$$

which does not depend on  $x$ . Hence for  $\mathfrak{C}_{a\sqrt{n}}$  (defined with respect to  $S_n^\diamond$ )

$$\mathfrak{C}_{a\sqrt{n}} = \inf_{b \geq a} \left( \sup_{x \in \mathbb{R}} |\mathfrak{T}_{a\sqrt{n}}^{b\sqrt{n}}(x)| + 1/(b\sqrt{n}) \right) \leq C'_{a,\delta} \frac{\log(n)}{n^{-1/2-\delta}}.$$

An application of Lemma 4.1 then yields

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(S_n^\diamond \leq x\sqrt{n}) - \mathbb{P}(L_n^\diamond \leq x)| \leq C''_{a,\delta} n^{-1/2-\delta'}, \quad \delta' > 0.$$

Plugging this into (4.17), we obtain

$$W_1(\mathbb{P}_{S_n^\diamond/\sqrt{n}}, \mathbb{P}_{L_n^\diamond}) \leq C'''_{\delta,\delta} n^{-1/2-\delta'} (1 + \log n). \tag{4.18}$$

Selecting  $\eta = \eta_n \rightarrow 0$  sufficiently slow ((2.4) must be valid, see (4.16)), the claim follows by combining (4.15) and (4.18).  $\square$

## 5 Proofs of volatility models

We first state the following elementary lemma.

**Lemma 5.1.** *Suppose that the function  $f$  satisfies (3.7), and assume  $\|Y_0\|_{p(\alpha+\beta)} + \sum_{k \in \mathbb{N}} k^2 \|Y_k - Y_k^*\|_{p(\alpha+\beta)}^\beta < \infty$  for  $p \geq 3$ . Then  $X_k = f(Y_k) - \mathbb{E}f(Y_k)$  satisfies **(A2)**.*

*Proof of Lemma 5.1.* Using Hölders inequality with  $r = \alpha/\beta + 1$ ,  $s = (\alpha + \beta)/\alpha$ , we get

$$\begin{aligned} \|X_k - X_k^*\|_p &\leq \| |Y_k - Y_k^*|^\beta (L + |Y_k|^\alpha + |Y_k^*|^\alpha) \|_p \leq \|Y_k - Y_k^*\|_{rp\beta}^\beta (L + 2\|Y_k\|_{sp\alpha}^\alpha) \\ &= \|Y_k - Y_k^*\|_{p(\alpha+\beta)}^\beta (L + 2\|Y_k\|_{p(\alpha+\beta)}^\alpha). \end{aligned} \quad \square$$

**5.1 Proof of Theorem 3.4**

*Proof of Theorem 3.4.* In order to apply Theorem 2.3, we need to validate both Assumptions 2.1 and 2.2, based on Assumptions 3.1 and 3.3. We will do so below. Since, as mentioned above,  $V_k$  is a time-homogenous Bernoulli-shift process, the quantities in Assumption 2.2 do not depend on  $l \in \mathbb{Z}$ , simplifying the notation. We first consider the case  $h_n \equiv 0$ .

**(B1):** We first validate **(B1)**, which requires most attention. To this end, we first introduce some necessary quantities. Let

$$(V_k^+)^{2\lambda} = \sum_{m=1}^{\infty} \sum_{1 \leq l_1, \dots, l_m \leq \tau} g_{l_m}(\varepsilon_{k-l_1-\dots-l_m}) \mathbb{1}(k > l_1 + \dots + l_m) \times \prod_{i=1}^{m-1} c_{l_i}(\varepsilon_{k-l_1-\dots-l_i}) \mathbb{1}(k > l_1 + \dots + l_i),$$

and  $X_k^+ = f(\varepsilon_k V_k^+) \in \mathcal{E}_1^+$  for  $k \geq 1$  (note that  $V_k \in \mathcal{E}_{k-1}$  here by construction). Observe the bound

$$|(V_k^+)^{2\lambda} - V_k^{2\lambda}| \leq \sum_{m > \lfloor k/\tau \rfloor} \sum_{1 \leq l_1, \dots, l_m \leq \tau} g_{l_m}(\varepsilon_{k-l_1-\dots-l_m}) \prod_{i=1}^{m-1} c_{l_i}(\varepsilon_{k-l_1-\dots-l_i}), \tag{5.1}$$

which we will repeatedly use. For  $\delta > 0$ , denote with

$$\mathcal{A}_{1j\delta} = \left\{ \sum_{k=1}^j \mathbb{E}_{\mathcal{E}_1^+} |X_k - X_k^+| \leq \delta \right\},$$

and let  $\mathcal{E}_{ij} = \sigma(\varepsilon_k, i \leq k \leq j)$  (note  $\mathcal{E}_{1\infty} = \mathcal{E}_1^+$  and  $\mathcal{A}_{1j\delta} \in \mathcal{E}_{1j}$ ). Then

$$\begin{aligned} \mathbb{P} \left( \sum_{k=1}^{\infty} \mathbb{E}_{\mathcal{E}_1^+} |X_k - X_k^+| \leq 2\delta \right) &\geq \mathbb{P} \left( \sum_{k>j}^{\infty} \mathbb{E}_{\mathcal{E}_1^+} |X_k - X_k^+| \leq \delta \cap \mathcal{A}_{1j\delta} \right) \\ &= \mathbb{E} \mathbb{P}_{\mathcal{E}_{1j}} \left( \sum_{k>j}^{\infty} \mathbb{E}_{\mathcal{E}_1^+} |X_k - X_k^+| \leq \delta \right) \mathbb{1}_{\mathcal{A}_{1j\delta}} \\ &\geq \mathbb{E} \left( 1 - \delta^{-1} \sum_{k>j}^{\infty} \mathbb{E}_{\mathcal{E}_{1j}} |X_k - X_k^+| \right) \mathbb{1}_{\mathcal{A}_{1j\delta}}, \end{aligned} \tag{5.2}$$

where we used Markov’s inequality in the last step. This simple lower bound is the key for establishing **(B1)**.

Let  $\mathcal{B}_{1j\eta} = \{|\varepsilon_k| \leq \eta, 1 \leq k \leq j\}$ . Select  $0 < \eta \leq 1$  such that  $\sup_{|\varepsilon| \leq \eta} c_i(\varepsilon) \leq \|c_i(\varepsilon_0)\|_q$ ,  $\sup_{|x| \leq \eta} g_i(x) \leq C_g$  for  $1 \leq i \leq \tau$ , which is always possible by Assumption 3.3. Assume without loss of generality  $\|g_i(\varepsilon)\|_q \leq C_g$ . Then by Assumption 3.3, we have on the event  $\mathcal{B}_{1j\eta}$

$$\left( \mathbb{E}_{\mathcal{E}_{1j}} \left| \sum_{1 \leq l_1, \dots, l_m \leq \tau} g_{l_m}(\varepsilon_{k-l_1-\dots-l_m}) \prod_{i=1}^{m-1} c_{l_i}(\varepsilon_{k-l_1-\dots-l_i}) \right|^q \right)^{1/q} \leq C_g \tau \gamma_c^{m-1},$$

where we recall  $\|c_1(\varepsilon)\|_q + \dots + \|c_\tau(\varepsilon)\|_q \leq \gamma_c < 1$ . From the above, we conclude (on the event  $\mathcal{B}_{1j\eta}$ ),

$$\left( \mathbb{E}_{\mathcal{E}_{1j}} |V_k^{2\lambda} - (V_k^+)^{2\lambda}|^q \right)^{1/q} \leq C_g \tau \frac{\gamma_c^{\lfloor k/\tau \rfloor}}{1 - \gamma_c}, \tag{5.3}$$

and similarly, one obtains (still on the event  $\mathcal{B}_{1j\eta}$ )

$$(\mathbb{E}_{\mathcal{E}_{1j}}|V_k^{2\lambda}|^q)^{1/q} \leq C_g \mathfrak{r} \frac{1}{1 - \gamma_c}. \tag{5.4}$$

In the following derivations below, all norms  $\|\cdot\|_r$  are taken with respect to  $\mathbb{P}_{\mathcal{E}_{1j}}$ . Since  $V_k \geq V_k^+$  and  $\lambda \geq 1/2$ , we have from  $f \in \mathcal{H}(L, \alpha, \beta)$ ,  $|x^{1/(2\lambda)} - y^{1/(2\lambda)}| \leq |x - y|^{1/(2\lambda)}$  and Cauchy-Schwarz

$$\begin{aligned} \|X_k - X_k^+\|_1 &\leq L \|\varepsilon_k|^\beta |V_k - V_k^+|^\beta (1 + |2\varepsilon_k V_k|^\alpha)\|_1 \\ &\leq L \|\varepsilon_k|^\beta |V_k^{2\lambda} - (V_k^+)^{2\lambda}|^{\beta/(2\lambda)} (1 + |2\varepsilon_k V_k|^\alpha)\|_1 \\ &\leq L \|\varepsilon_k|^\beta |V_k^{2\lambda} - (V_k^+)^{2\lambda}|^{\beta/(2\lambda)}\|_2 \|1 + |2\varepsilon_k V_k|^\alpha\|_2. \end{aligned} \tag{5.5}$$

By independence, (5.3) and Jensen’s inequality (if  $\beta/\lambda < 1$ , we again apply Jensen’s inequality in addition)

$$\|\varepsilon_k|^{2\beta} |V_k^{2\lambda} - (V_k^+)^{2\lambda}|^{\beta/\lambda}\|_1^{\lambda/\beta} \leq \|\varepsilon_k^2\|_1^\lambda C_g \mathfrak{r} \frac{\gamma_c^{\lfloor k/\mathfrak{r} \rfloor}}{1 - \gamma_c}. \tag{5.6}$$

Similarly, by independence and (5.4)

$$\|1 + |2\varepsilon_k V_k|^\alpha\|_2 \leq 1 + 2 \|\varepsilon_k\|_{2\alpha}^\alpha \left(C_g \mathfrak{r} \frac{1}{1 - \gamma_c}\right)^{\frac{\alpha}{2\lambda}}. \tag{5.7}$$

All in all, on the event  $\mathcal{B}_{1j\eta}$ , combining (5.5), (5.6), and (5.7), we arrive at

$$\mathbb{E}_{\mathcal{E}_{1j}}|X_k - X_k^+|_1 \leq C_+ \rho^k (\mathbb{E}_{\mathcal{E}_{1j}}|\varepsilon_k|^2)^{\beta/2},$$

where  $C_+$  does not depend on  $\eta$ , and  $\rho < 1$  only depends on  $\gamma_c$ ,  $\lambda$  and  $\beta$ . Moreover, we have  $\mathbb{E}_{\mathcal{E}_{1j}}\varepsilon_k^2 = \varepsilon_k^2 \leq \eta^2$  for  $1 \leq k \leq j$  on the event  $\mathcal{B}_{1j\eta}$ , and consequently

$$\sum_{k=1}^j \mathbb{E}_{\mathcal{E}_{1j}}|X_k - X_k^+| \leq C_+ \eta^\beta \frac{1}{1 - \rho}. \tag{5.8}$$

Hence, selecting  $\eta$  such that  $C_+ \eta^\beta \frac{1}{1 - \rho} < \delta$ , we conclude  $\mathcal{B}_{1j\eta} \subseteq \mathcal{A}_{1j\delta}$ , and **(V1)** yields

$$\mathbb{P}(\mathcal{A}_{1j\delta}) \geq \mathbb{P}(\mathcal{B}_{1j\eta}) > 0. \tag{5.9}$$

Similarly, we obtain on  $\mathcal{B}_{1j\eta}$ , for  $k > j$ , the estimate

$$\sum_{k>j} \mathbb{E}_{\mathcal{E}_{1j}}|X_k - X_k^+| \leq C_+ \frac{\rho^j}{1 - \rho} (\mathbb{E}|\varepsilon_0|^2)^{\beta/2}.$$

Selecting  $j_0$  sufficiently large, we get for  $j \geq j_0$

$$\begin{aligned} &\mathbb{E}\left(1 - \delta^{-1} \sum_{k>j} \mathbb{E}_{\mathcal{E}_{1j}}|X_k - X_k^+|\right) \mathbb{1}_{\mathcal{A}_{1j\delta}} \\ &\geq \mathbb{E}\left(1 - \delta^{-1} C_+ \frac{\rho^j}{1 - \rho} (\mathbb{E}|\varepsilon_0|^2)^{\beta/2}\right) \mathbb{1}_{\mathcal{B}_{1j\eta}} \\ &\geq \mathbb{E}\left(1 - \frac{1}{2}\right) \mathbb{1}_{\mathcal{B}_{1j\eta}} = \frac{\mathbb{P}(\mathcal{B}_{1j\eta})}{2} > 0. \end{aligned}$$

Hence **(B1)** holds. Let us now consider the case  $h_n \neq 0$ , which turns out to be just a minor extension. Arguing as above in (5.5), (5.6), and (5.7), we obtain  $\sum_{k=1}^\infty \|h_n(V_k) -$

$h_n(V_k^+) \|_1 \leq C$ , where  $C$  does not depend on  $n$ . Hence by the triangle and Jensen's inequality

$$\left\| \sum_{k=1}^{\infty} \mathbb{E}_{\mathcal{E}_1^+} |h_n(V_k) - h_n(V_k^+)| \right\|_1 \leq C. \tag{5.10}$$

On the other hand, (3.8), the triangle and Jensen's inequality imply that there exists  $l_n \rightarrow \infty$ , such that

$$\left\| \sum_{k=1}^{l_n} \mathbb{E}_{\mathcal{E}_1^+} |h_n(V_k)| \right\|_1 \leq \sum_{k=1}^{l_n} \mathbb{E} |h_n(V_k)| \rightarrow 0 \tag{5.11}$$

as  $n$  increases. Setting  $h_n^+(V_k) = h_n(V_k^+)$  for  $k \geq l_n$  and  $h_n^+(V_k) \equiv 0$  otherwise, we conclude from the above

$$\lim_{n \rightarrow \infty} \left\| \sum_{k=1}^{\infty} \mathbb{E}_{\mathcal{E}_1^+} |h_n(V_k) - h_n^+(V_k)| \right\|_1 = 0. \tag{5.12}$$

Piecing everything together, the validity of **(B1)** follows.

**(B2)**: Note that  $v \mapsto \mathbb{E} e^{i\xi f(v\epsilon)}$  is continuous in  $v$ . Then by compactness of  $\mathcal{V}$  and **(V2)**, we have  $\sup_{v \in \mathcal{V}} |\mathbb{E} e^{i\xi f(v\epsilon)}| \mathbb{1}_{\mathcal{V}} < 1$ , and due to  $|e^{iz}| = 1$ ,  $h_n(V_k) \in \mathcal{E}_{k-1}$  and **(V2)**, we thus conclude

$$\begin{aligned} \mathbb{E} |\mathbb{E}_{\mathcal{E}_{-1}} e^{i\xi X_0}| &\leq \mathbb{E} \sup_{v \in \mathcal{V}} |\mathbb{E} e^{i\xi f(v\epsilon)}| \mathbb{1}_{\mathcal{V}} + \mathbb{P}(V_0 \in \mathcal{V}^c) \\ &< \mathbb{P}(V_0 \in \mathcal{V}) + \mathbb{P}(V_0 \in \mathcal{V}^c) = 1. \end{aligned}$$

Hence **(B2)** holds.

**(A2)**: Arguing similarly as above for establishing **(B1)**, one derives

$$\|\varepsilon_k V_k - \varepsilon_k V_k^*\|_q, \|V_k - V_k^*\|_q \leq C \rho^k, \quad \rho < 1. \tag{5.13}$$

The claim now follows from Lemma 5.1 and the triangle inequality.

**(A1)**: We may repeat arguments employed in **(A2)** (resp. **(B1)**). □

### 5.2 Proof of Theorem 3.6

*Proof of Theorem 3.6.* Let  $V_k^+ = F_{\varepsilon_k} \circ F_{\varepsilon_{k-1}} \circ \dots \circ F_{\varepsilon_1}(0)$ . Then  $V_k^+ \in \mathcal{E}_l^+$ . Although we are no longer in the time-homogenous Bernoulli-shift setup, it is obvious that we can repeat the proof of Theorem 3.4, almost verbatim. In fact, due to the more explicit iterative structure, some computations are even simpler. □

### 5.3 Proof of Theorem 3.8

*Proof of Theorem 3.8.* As for augmented Garch sequences, we are again in the time-homogenous Bernoulli-shift case. Let

$$G_k^+ = \sum_{i=0}^{k-1} a_i c_i (\epsilon_{k-i}), \quad V_k^+ = g(G_k^+).$$

Then  $V_k^+ \in \mathcal{E}_1^+$ . It is again obvious that we can repeat the proof of Theorem 3.4, almost verbatim. As in the case of Theorem 3.6, the actual proof is even simpler. □

### 5.4 Proof of Theorem 3.10

*Proof of Theorem 3.10.* As for augmented Garch sequences and functions of linear processes, we are again in the time-homogenous Bernoulli-shift case. Let

$$V_k^+ = \sum_{i=1}^{\infty} \sum_{0 \leq j_1 < \dots < j_i \leq k} a_k(j_1, \dots, j_i) \epsilon_{k-j_1} \cdots \epsilon_{k-j_i}.$$

Since clearly  $V_1^+ \in \mathcal{E}_1^+$ , we may now repeat the proof of Theorem 3.4. As in previous cases, the actual proof is even simpler.  $\square$

## References

- [1] E. Alòs, *A generalization of the Hull and White formula with applications to option pricing approximation*, *Finance Stoch.* **10** (2006), no. 3, 353–365. MR2244350
- [2] D. W. K. Andrews, *Non-strong mixing autoregressive processes*, *Journal of Applied Probability* **21** (1984), no. 4, pp. 930–934 (English). MR0766830
- [3] C. Bayer, P. Friz, A. Gulisashvili, B. Horvath, and B. Stemper, *Short-time near-the-money skew in rough fractional volatility models*, *Quantitative Finance* **19** (2017). MR3939657
- [4] J.S. Bendat, *Nonlinear system analysis and identification from random data*, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1990. MR1045198
- [5] I. Berkes, S. Hörmann, and L. Horváth, *The functional central limit theorem for a family of GARCH observations with applications*, *Statist. Probab. Lett.* **78** (2008), no. 16, 2725–2730. MR2465114
- [6] R. N. Bhattacharya and R.R. Rao, *Normal approximation and asymptotic expansions*, *Classics in Applied Mathematics*, vol. 64, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010. MR3396213
- [7] P. Billingsley, *Convergence of probability measures*, John Wiley & Sons Inc., New York, 1968. MR0233396
- [8] S. G. Bobkov, *Central limit theorem and Diophantine approximations*, *J. Theoret. Probab.* **31** (2018), no. 4, 2390–2411. MR3866618
- [9] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, *J. Econometrics* **31** (1986), no. 3, 307–327. MR853051
- [10] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*, second ed., *Springer Series in Statistics*, Springer-Verlag, New York, 1991. MR1093459
- [11] P. Chebyshev, *Sur deux théorèmes relatifs aux probabilités*, *Acta Math.* **14** (1890), no. 1, 305–315. MR1554802
- [12] H. Cramér, *On the composition of elementary errors.*, **11**.
- [13] J. Dedecker, P. Doukhan, G. Lang, R. León, S. Louhichi, and Cl. Prieur, *Weak dependence: with examples and applications*, *Lecture Notes in Statistics*, vol. 190, Springer, New York, 2007. MR2338725
- [14] J. Dedecker and C. Prieur, *New dependence coefficients. Examples and applications to statistics*, *Probab. Theory Related Fields* **132** (2005), no. 2, 203–236. MR2199291
- [15] M. Denker and E. C. Waymire (eds.), *Rabi N. Bhattacharya—selected papers*, *Contemporary Mathematicians*, Birkhäuser/Springer, [Cham], 2016. MR3470100
- [16] P. Diaconis and D. Freedman, *Iterated random functions*, *SIAM Rev.* **41** (1999), no. 1, 45–76. MR1669737
- [17] J.-C. Duan, *The garch option pricing model*, *Mathematical Finance* **5** (1995), no. 1, 13–32. MR1322698
- [18] J.-C. Duan, G. Geneviève, and J.-G. Simonato, *An analytical approximation for the garch option pricing model*, *Journal of Computational Finance* **2** (2000).
- [19] J. C. Duan, *Augmented GARCH(p, q) process and its diffusion limit*, *J. Econometrics* **79** (1997), no. 1, 97–127. MR1457699



- [20] F. Y. Edgeworth, *The asymmetrical probability curve.*, Proceedings of the Royal Society of London **56** (1894), no. 336-339, 271–272.
- [21] B. Efron, *Bootstrap methods: another look at the jackknife*, Ann. Statist. **7** (1979), no. 1, 1–26. MR515681
- [22] B. Efron, *The jackknife, the bootstrap and other resampling plans*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 38, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1982. MR659849
- [23] O. E. Euch, M. Fukasawa, J. Gatheral, and M. Rosenbaum, *Short-term at-the-money asymptotics under stochastic volatility models*, SIAM J. Financial Math. **10** (2019), no. 2, 491–511. MR3945237
- [24] W. Feller, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971. MR0270403
- [25] K. Fernando and C. Liverani, *Edgeworth expansions for weakly dependent random variables*, Annales de l'Institut Henri Poincaré (B) Probability and Statistics (to appear). MR4255182
- [26] J.-P. Fouque, G. Papanicolaou, R. Sircar, and K. Solna, *Singular perturbations in option pricing*, SIAM J. Appl. Math. **63** (2003), no. 5, 1648–1665. MR2001213
- [27] C. Francq and J.-M. Zakoïan, *GARCH models*, John Wiley & Sons, Ltd., Chichester, 2010, Structure, statistical inference and financial applications. MR3186556
- [28] P. K. Friz, P. Gassiat, and P. Pigato, *Precise asymptotics: Robust stochastic volatility models*, The Annals of Applied Probability **31** (2021), no. 2, 896 – 940. MR4254499
- [29] M. Fukasawa, *Edgeworth expansion for ergodic diffusions*, Probab. Theory Related Fields **142** (2008), no. 1-2, 1–20. MR2413265
- [30] M. Fukasawa, *Asymptotic analysis for stochastic volatility: Edgeworth expansion*, Electron. J. Probab. **16** (2011), no. 25, 764–791. MR2793240
- [31] J. Gatheral, *The volatility surface: A practitioner's guide*, John Wiley & Sons, Ltd, 2012.
- [32] F. Götze and C. Hipp, *Asymptotic expansions for sums of weakly dependent random vectors*, Z. Wahrsch. Verw. Gebiete **64** (1983), no. 2, 211–239. MR714144
- [33] F. Götze and C. Hipp, *Asymptotic distribution of statistics in time series*, Ann. Statist. **22** (1994), no. 4, 2062–2088. MR1329183
- [34] P. Hall, *The bootstrap and Edgeworth expansion*, Springer Series in Statistics, Springer-Verlag, New York, 1992. MR1145237
- [35] P. Hall and J. L. Horowitz, *Bootstrap critical values for tests based on generalized-method-of-moments estimators*, Econometrica **64** (1996), no. 4, 891–916. MR1399222
- [36] L. Hervé and F. Pène, *The Nagaev-Guivarc'h method via the Keller-Liverani theorem*, Bull. Soc. Math. France **138** (2010), no. 3, 415–489. MR2729019
- [37] S. L. Heston and S. Nandi, *A closed-form garch option valuation model*, Rev. Financ. Stud. **13** (2000), no. 3, 585–625.
- [38] Z. Huang, T. Wang, and P. R. Hansen, *Option Pricing with the Realized GARCH Model: An Analytical Approximation Approach*, Journal of Futures Markets **37** (2017), no. 4, 328–358.
- [39] I. A. Ibragimov, *On the accuracy of approximation by the normal distribution of distribution functions of sums of independent random variables*, Teor. Veroyatnost. i Primenen **11** (1966), 632–655. MR0212853
- [40] M. Jirak, W. B. Wu, and O. Zhao, *Sharp connections between Berry-Esseen characteristics and Edgeworth expansions for stationary processes*, Trans. Amer. Math. Soc. **374** (2021), no. 6, 4129–4183. MR4251225
- [41] M. Jirak, *A Berry-Esseen bound with (almost) sharp dependence conditions*, Bernoulli **29** (2023), no. 2, 1219–1245. MR4550221
- [42] I. Kontoyiannis and S. P. Meyn, *Spectral theory and limit theorems for geometrically ergodic Markov processes*, Ann. Appl. Probab. **13** (2003), no. 1, 304–362. MR1952001
- [43] N. Kunitomo and A. Takahashi, *The asymptotic expansion approach to the valuation of interest rate contingent claims*, Math. Finance **11** (2001), no. 1, 117–151. MR1807851

- [44] S. Kusuoka and N. Yoshida, *Malliavin calculus, geometric mixing, and expansion of diffusion functionals*, Probab. Theory Related Fields **116** (2000), no. 4, 457–484. MR1757596
- [45] S. N. Lahiri, *Refinements in asymptotic expansions for sums of weakly dependent random vectors*, Ann. Probab. **21** (1993), no. 2, 791–799. MR1217565
- [46] S. N. Lahiri, *Resampling methods for dependent data*, Springer Series in Statistics, Springer-Verlag, New York, 2003. MR2001447
- [47] W. Liu, H. Xiao, and W. B. Wu, *Probability and moment inequalities under dependence*, Statist. Sinica **23** (2013), no. 3, 1257–1272. MR3114713
- [48] M. Maxwell and M. Woodroffe, *Central limit theorems for additive functionals of Markov chains*, Ann. Probab. **28** (2000), no. 2, 713–724. MR1782272
- [49] P. A. Mykland, *Asymptotic expansions for martingales*, Ann. Probab. **21** (1993), no. 2, 800–818. MR1217566
- [50] P. A. Mykland, *Martingale expansions and second order inference*, Ann. Statist. **23** (1995), no. 3, 707–731. MR1345195
- [51] D. Nualart and N. Yoshida, *Asymptotic expansion of Skorohod integrals*, Electron. J. Probab. **24** (2019), Paper No. 119, 64. MR4029422
- [52] M. B. Priestley, *Nonlinear and nonstationary time series analysis*, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, 1988. MR991969
- [53] Y. Rinott and V. Rotar, *On Edgeworth expansions for dependency-neighborhoods chain structures and Stein’s method*, Probab. Theory Related Fields **126** (2003), no. 4, 528–570. MR2001197
- [54] E. Rio, *About the conditional value at risk of partial sums*, C. R. Math. Acad. Sci. Paris **355** (2017), no. 11, 1190–1195. MR3724884
- [55] M. Rosenblatt, *Markov processes: structure and asymptotic behavior*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen; 184, Springer-Verlag, Berlin, New York (eng). MR0329037
- [56] W. J. Rugh, *Nonlinear system theory*, Johns Hopkins Series in Information Sciences and Systems, Johns Hopkins University Press, Baltimore, Md., 1981, The Volterra-Wiener approach. MR713994
- [57] W. B. Wu, *Nonlinear system theory: Another look at dependence*, Proceedings of the National Academy of Sciences USA. **102** (2005), 14150–14154. MR2172215
- [58] W. B. Wu and X. Shao, *Limit theorems for iterated random functions*, J. Appl. Probab. **41** (2004), no. 2, 425–436. MR2052582
- [59] N. Yoshida, *Asymptotic expansion for statistics related to small diffusions*, J. Japan Statist. Soc. **22** (1992), no. 2, 139–159. MR1212246
- [60] N. Yoshida, *Malliavin calculus and asymptotic expansion for martingales*, Probab. Theory Related Fields **109** (1997), no. 3, 301–342. MR1481124
- [61] N. Yoshida, *Malliavin calculus and martingale expansion*, vol. 125, 2001, Rencontre Franco-Japonaise de Probabilités (Paris, 2000), pp. 431–456. MR1869987
- [62] N. Yoshida, *Partial mixing and Edgeworth expansion*, Probab. Theory Related Fields **129** (2004), no. 4, 559–624. MR2078982

**Acknowledgments.** I would like to thank the reviewers for a careful reading and the many thoughtful comments, pointing in particular to related and attached literature.