LARGE AND MODERATE DEVIATIONS FOR STOCHASTIC VOLTERRA SYSTEMS

ANTOINE JACQUIER AND ALEXANDRE PANNIER

Abstract. We provide a unified treatment of pathwise Large and Moderate deviations principles for a general class of multidimensional stochastic Volterra equations with singular kernels, not necessarily of convolution form. Our methodology is based on the weak convergence approach by Budhijara, Dupuis and Ellis [14, 36]. We show in particular how this framework encompasses most rough volatility models used in mathematical finance, and generalises many recent results in the literature.

1. Introduction

This paper sheds new light on the asymptotic behaviour of the class of stochastic Volterra equations (SVEs)

$$X_t = X_0 + \int_0^t K(t,s)b(s,X_s)\mathrm{d}s + \int_0^t K(t,s)\sigma(s,X_s)\mathrm{d}W_s, \quad t \in [0,T], \tag{1.1}$$

where $X_0 \in \mathbb{R}^d$, $d \geq 1$, W is a multidimensional Brownian motion, K is a kernel that may be singular, and the coefficients are such that a unique pathwise solution exists. This class of models has been investigated in many fields, including nonlinear filtering [29] using fractional Brownian motion kernels, pharmacokinetic models [63] (Langevin equation driven by fractional Brownian motion), fluid turbulence [21], and turbulence modelling in atmospheric winds or energy prices [3, 28] using Brownian Semistationary processes.

Mathematical finance has however been the most dynamic area by far in terms of applications of SVEs, and an in-depth study of (1.1) with convolution kernels was recently carried out by Abi Jaber, Larsson and Pulido [1]. Following previous analyses supporting non-Markovian systems [2, 26, 25, 23, 24, 47], the investigation of high-frequency data in [51] revealed the roughness, in the sense of low Hölder regularity, of the observed time series of the instantaneous volatility of stock price processes. This suggested that fractional Brownian motion (fBm) with small Hurst parameter ($H \approx 0.1$) is an accurate driver for its dynamics. Since this seminal observation, more advanced results [39] have proposed that that the drift and the diffusion coefficients should be state dependent, giving rise to the widespread development of (1.1) in quantitative finance.

For option pricing purposes, the asymptotic results in [2, 7, 48] showed that the short-maturity behaviour of option prices is captured much more accurately by these rough volatility models rather than by Markovian diffusions. Reconciling the stylised facts of the markets from both the statistical and the option pricing viewpoints is the tour de force that make these models so important today. However, the loss in tractability compared to classical Itô diffusions is not negligible. The solution to (1.1) is in general not a semimartingale nor a Markov process, preventing the use of Itô calculus or Feynman-Kac type formulas. Path-dependent versions of the latter are available in some cases, in particular for affine rough volatility models [1, 31, 41, 52], but general results are scarce. Rough path theory is a natural route but is not available for $H \leq 1/4$, although a regularity structure approach

Date: April 22, 2020.

2010 *Mathematics Subject Classification.* 60F10, 60G22, 91G20.

Key words and phrases. stochastic Volterra equations, large deviations, moderate deviations, weak convergence, rough volatility.

was recently developed [5]. In this context, one could turn to numerical methods to understand the dynamics or to price options but, despite new advances based on Monte-Carlo methods [6, 8, 64], rough Donsker theorem [57] or Fourier methods [41], the roughness and memory of the process seriously complicate the task.

Asymptotic methods have been used, both to provide clearer understanding of models in extreme parameter configurations and to act as proxies to numerical schemes. Large Deviations Principles (LDP), in particular, have been widely explored in mathematical finance, and we refer the interested reader to [69] for an overview. Let $\{X^\varepsilon\}_{\varepsilon>0}$ be a sequence in some Polish space $\mathcal{X}$, converging in probability to a deterministic limit $\overline{X}$ as ε goes to zero. This sequence is said to satisfy an LDP with speed ε^{-1} and rate function $I : \mathcal{X} \to [0, +\infty]$ if for all Borel subsets $B \subset \mathcal{X}$, the inequalities

$$-\inf_{x\in B^\circ} I(x) \le \liminf_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{P}\big(X^\varepsilon \in B\big) \le \limsup_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{P}\big(X^\varepsilon \in B\big) \le -\inf_{x\in \overline{B}} I(x)$$

hold, and the level sets $\{x \in \mathcal{X} : I(x) \le N\}$ of I are compact for all $N > 0$. This rate function encompasses in a (relatively) concise formula first-order information about the asymptotic behaviour of complex dynamical systems. If X satisfies (1.1) and $\mathcal{X} = \mathbb{R}^d$, one can consider finite-dimensional LDP for $\{X_t\}_{t\ge 0}$ (also called small-time LDP if the limit takes place as t goes to zero), or pathwise LDP for some rescaling of X with $\mathcal{X} = \mathcal{C}([0, T] : \mathbb{R}^d)$. The former is easily recovered from the latter by a projection argument. Moderate deviations however are concerned with deviations of a lower order than large deviations, and thus apply to 'less rare events'. We indeed say that $\{X^\varepsilon\}_{\varepsilon>0}$ satisfies a moderate deviations principle (MDP) if $\{\eta^\varepsilon\}_{\varepsilon>0}$ satisfies an LDP with speed h_ε, where

$$\eta^\varepsilon := \frac{X^\varepsilon - \overline{X}}{\varepsilon h_\varepsilon}, \quad \text{for all } \varepsilon > 0,$$

with $\lim_{\varepsilon\downarrow 0} h_\varepsilon = +\infty$ and $\lim_{\varepsilon\downarrow 0} \varepsilon h_\varepsilon = 0$. Since the speed of convergence of h_ε is not fixed, an MDP essentially bridges the gap between the central limit regime where $h_\varepsilon = 1$ and the LDP regime where $h_\varepsilon = \varepsilon^{-1}$. An edifying example of the relevance of moderate deviations appeared in [46] where, interested in option pricing asymptotics, the authors judiciously rescale the strikes with respect to time to expiry. Indeed, as time to expiry becomes smaller, the range of pertinent strikes naturally shrinks, and this 'moderately-out-of-the-money' regime becomes more realistic.

Large deviations for SVEs were originally studied in [68, 71] with regular kernels. In the context of rough volatility, Forde and Zhang [43] introduced the first finite-dimensional LDP where the log-volatility is modelled by a fractional Brownian motion, and refined versions followed in [5, 7, 45], while pathwise LDP for similar models were studied in [20, 55]. Departing from regular conditions on the behaviour of the coefficients led to specific requirements, and finite-dimensional large deviations for the fractional Heston model were carried out in [42, 54], while more elaborate pathwise LDPs were derived for the rough Stein-Stein model with random starting point [56], for the rough Bergomi model [58], and small-time LDPs for the multifactor rough Bergomi appeared in [61].

The Gärtner-Ellis theorem [33, Theorem 2.3.6] is the main ingredient of a finite-dimensional LDP and depends on explicit computations of certain limits of the Laplace transform. This is only available though, when the process is either Gaussian [43] or affine [42]. Pathwise LDP on the other hand, have mainly been derived using the Freidlin-Wentzell approach [44]: starting from known large deviations for the driving (Gaussian) process [34, Theorem 3.4.5], they follow from a combination of approximations and continuous mapping, keeping track of the rate function. While this methodology is clear, it requires a case-by-case tailored path for each model, and in general leads to a cumbersome rate function. Furthermore, pathwise moderate deviations are so far out of reach in this approach, partially explaining the small number of related results compared to LDP.

A radically different method, introduced by Dupuis and Ellis in the monograph [36] and developed further by Budhiraja and Dupuis [14], relies on the equivalence between the LDP and the Laplace principle. The family $\{X^\varepsilon\}_{\varepsilon>0}$ is said to satisfy the Laplace principle with speed ε^{-1} and rate function $I : \mathcal{X} \to [0, +\infty]$ if for all continuous bounded maps $F : \mathcal{X} \to \mathbb{R}$,

$$\lim_{\varepsilon \downarrow 0} -\varepsilon \log \mathbb{E}\left[\exp\left\{-\frac{F(X^\varepsilon)}{\varepsilon}\right\}\right] = \inf_{x\in\mathcal{X}} \big\{I(x) + F(x)\big\}. \tag{1.2}$$

This alternative, called the weak convergence approach, consists in proving a Laplace principle where the left-hand side pre-limit of (1.2) can be represented as a variational principle for expectations of functionals of Brownian motion [11, Theorem 3.1]:

Lemma 1.1 (Boué-Dupuis)**.** *Let W be an $\mathbb{R}^m$-Brownian motion and F be a bounded Borel-measurable function mapping $\mathcal{C}([0,T] : \mathbb{R}^m)$ into $\mathbb{R}$. Then*

$$-\log \mathbb{E}\left[\mathrm{e}^{-F(W)}\right] = \inf_{v\in\mathcal{A}} \mathbb{E}\left[\frac{1}{2}\int_0^T |v_s|^2\,\mathrm{d}s + F\left(W + \int_0^\cdot v_s \mathrm{d}s\right)\right], \tag{1.3}$$

where

$$\mathcal{A} := \left\{v : [0,T] \to \mathbb{R}^m \text{ progressively measurable, } \mathbb{E}\left[\int_0^T |v_t|^2 \mathrm{d}t\right] < +\infty\right\}. \tag{1.4}$$

The representation (1.3) contains in a single formula the usual tools used in the proof of an LDP. The first term on the right-hand side comes from the relative entropy between the Wiener measure and the measure shifted by $\int_0^\cdot v_s \mathrm{d}s$ via Girsanov's theorem, under which $W + \int_0^\cdot v_s \mathrm{d}s$ is a Brownian motion. It can be interpreted as the cost of deviating from the original path and clearly indicates where the form of the rate function comes from. In essence, this representation replaces the non-linear analysis of the Freidlin-Wentzell approach with the linear theory of weak convergence. Instead of exponential estimates, only qualitative properties of the shifted process need to be established, such as strong existence and uniqueness and tightness.

The extensive literature on the topic, summarised in [14] and the references therein, demonstrates the strength of this generic approach which can be applied to a variety of models without appealing to their particular features. It has been used to derive LDPs, in the continuous-time case, for diffusions [22], multiscale systems [16, 37, 72], SDEs driven by infinite-dimensional Brownian motions [17], by Poisson random measures [18] or both [12], including stochastic PDEs. Contrary to the Freidlin-Wentzell approach, this method has also proved efficient to obtain MDPs. SVEs with Lipschitz continuous kernels [62], SDEs with jumps [15] and slow-fast systems [65] are a few relevant examples. The latter were then tailored to the setting of stochastic volatility models in [59], developing the first application of the weak convergence approach in mathematical finance, and extending the MDP results in [46] to a pathwise setting. A further appealing feature of moderate deviations is the simple form, often quadratic, of the rate function, as opposed to that provided by large deviations, thereby opening the gates to the use of importance sampling and variance reduction techniques [70, 38, 66].

Building on this powerful approach, we provide a unified treatment of (finite-dimensional and pathwise) large and moderate deviations in the general framework (1.1) by showing the weak convergence of a perturbed system. We relax the uniqueness requirement for the limiting Volterra equation, as in [27, 35] for the diffusion case, allowing us to consider coefficients that are not Lipschitz continuous and do not necessarily have sublinear growth.

The paper is organised in the following way: Section 2 introduces the framework and useful definitions. In section 3, we present abstract criteria for the validity of an LDP, extending the results by Budhiraja and Dupuis [13, Theorem 4.1]. Our main results, Theorem 3.8 for LDP and Theorem 3.16 for MDP, are then stated in the case of convolution kernels and extended to non-convolution kernels in Theorem 3.25. In Section 4, we show how these results apply to rough volatility models, and give precise formulae for the rough Stein-Stein, the (multifactor) rough Bergomi and the rough Heston models. We finally gather technical proofs in the appendix.

2. General framework

2.1. **Notations.** We consider a fixed time horizon $T > 0$, and denote $\mathbb{T} := [0, T]$, and $\mathbb{R}_+ := [0, +\infty]$. For $d_1 \geq 1$, $d_2 \geq 1$, $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^{d_1}$ and the Frobenius norm in $\mathbb{R}^{d_1 \times d_2}$. For $p \geq 1$, L^p stands short for $L^p(\mathbb{T})$, and $\|\cdot\|_2$ is the usual L^2 norm. Furthermore, for $d \geq 1$, $\mathcal{W}^d := \mathcal{C}(\mathbb{T} : \mathbb{R}^d)$ represents the space of continuous functions from $\mathbb{T}$ to $\mathbb{R}^d$, equipped with the supremum norm $\|\varphi\|_{\mathbb{T}} := \sup_{t\in\mathbb{T}} |\varphi_t|$ for any $\varphi \in \mathcal{W}^d$. Finally, for any $d_1 \geq 1$, $d_2 \geq 1$, $M > 0$, $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, we write $\|f\|_M := \sup\{|f(x)| : |x| \leq M\}$. Unless stated otherwise, constants will be denoted by C (with possible subscript) and may be different from one proof to another. Every statement involving ε stands for all $\varepsilon > 0$ small enough. A family of random variables will be called tight if the corresponding measures are tight [36, Appendix A]. We also use the classical convention that the infimum over an empty set is equal to infinity. Finally, we recall the following definitions for clarity and notations:

Definition 2.1. Let g be a function from $\mathbb{R}^d$ to $\mathbb{R}^n$.

- It has linear growth if there exists $C_L > 0$ such that $|g(x)| \leq C_L(1 + |x|)$, for all $x \in \mathbb{R}^d$;
- if it is uniformly continuous, it admits a continuous and increasing modulus of continuity $\rho_g : \mathbb{R}_+ \to \mathbb{R}_+$, with $\rho_g(0) = 0$ and $|g(x) - g(y)| \leq \rho_g(|x - y|)$, for all $x, y \in \mathbb{R}^d$;
- it is locally δ-Hölder continuous with $\delta \in (0, 1)$ if, for all $M > 0$, there exists $C_M > 0$ such that $|g(x) - g(y)| \leq C_M |x - y|^\delta$, for all $|x| \vee |y| \leq M$.

2.2. **Framework.** We consider small-noise convolution stochastic Volterra equations (SVE)

$$X_t^\varepsilon = X_0^\varepsilon + \int_0^t K(t-s) b_\varepsilon(s, X_s^\varepsilon)\mathrm{d}s + \vartheta_\varepsilon \int_0^t K(t-s)\sigma_\varepsilon(s, X_s^\varepsilon)\mathrm{d}W_s, \tag{2.1}$$

taking values in $\mathbb{R}^d$ with $d \geq 1$, where $\varepsilon > 0$, and $\vartheta_\varepsilon > 0$ tends to zero as ε goes to zero. For each $\varepsilon > 0$, $X_0^\varepsilon \in \mathbb{R}^d$, $b_\varepsilon : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma_\varepsilon : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are Borel-measurable functions, and W is an m-dimensional Brownian motion on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathbb{T}}, \mathbb{P})$ satisfying the natural conditions. The kernel function $K : \mathbb{T} \to \mathbb{R}^d \times \mathbb{R}^d$, of convolution type, is allowed to be singular, thus encompassing fractional processes, in particular the recent literature on rough volatility [2, 6, 39, 51]. Components of the system are in general correlated, the correlation matrix being implicitly encoded in the diffusion coefficient σ_ε. General existence and uniqueness results for such stochastic Volterra equations are so far out of reach, and our conditions below are sufficient and general enough for most applications. In order to state them precisely, we first introduce several definitions and concepts:

Definition 2.2. For any $\varepsilon > 0$, a solution to (2.1) is an $\mathbb{R}^d$-valued progressively measurable stochastic process X^ε satisfying (2.1) almost surely and such that

$$\mathbb{P}\left(\int_0^t \Big\{ |K(t-s)b_\varepsilon(s, X_s^\varepsilon)| + |K(t-s)\sigma_\varepsilon(s, X_s^\varepsilon)|^2 \Big\}\mathrm{d}s < \infty, \text{ for all } t \in \mathbb{T}\right) = 1.$$

We call it exact if it is pathwise unique.

We shall always assume that the (singular) convolution kernel satisfies the following condition, which is essentially a multivariate version of the one given in [1, Condition (2.5)]:

Assumption 2.3. The kernel $K : \mathbb{T} \to \mathbb{R}^{d\times d}$ is an upper triangular matrix satisfying the following conditions: $K \in L^2_{\mathrm{loc}}(\mathbb{T}; \mathbb{R}^{d\times d})$ and there exists $\gamma \in (0, 2]$ such that, for h small enough,

$$\int_0^h |K(t)|^2 \,\mathrm{d}t + \int_0^T |K(t+h) - K(t)|^2 \,\mathrm{d}t = \mathcal{O}(h^\gamma).$$

We refer to [1, Example 2.3] for a broad range of kernels that satisfy this assumption. Of particular interest in mathematical finance is the Riemann-Liouville kernel $K(t) = t^{H-\frac{1}{2}}$, for $H \in (0, \frac{1}{2})$ implying $\gamma = 2H$. Moreover if $\widetilde{K}$ is locally Lipschitz and K satisfies Assumption 2.3 then so does the product $K\widetilde{K}$; this includes the Gamma and Power-law kernels which are related to the class of Brownian Semistationary processes [4].

Remark 2.4. *This setup covers in particular the following two useful forms for the kernel:*

- $K = \mathrm{Diag}(k_1, \cdots, k_d)$ *is a diagonal matrix, where each* $k_i : \mathbb{T} \to \mathbb{R}$ *satisfies Assumption 2.3.*
- *The drift and diffusion coefficients of any sub-system of (2.1) can be convoluted with different kernels. As an example, the one-dimensional SVE*

$$Y_t^\varepsilon = Y_0^\varepsilon + \int_0^t K_1(t-s) b_\varepsilon^Y(s, Y_s^\varepsilon)\,\mathrm{d}s + \int_0^t K_2(t-s)\sigma_\varepsilon^Y(s, Y_s^\varepsilon)\,\mathrm{d}W_s^{(1)}$$

is the first component of (2.1) with $d = 2$ *and*

$$K = \begin{bmatrix} K_1 & K_2 \\ 0 & 0 \end{bmatrix}, \quad b_\varepsilon = \begin{bmatrix} b_\varepsilon^Y \\ 0 \end{bmatrix}, \quad \sigma_\varepsilon = \begin{bmatrix} 0 & 0 \\ \sigma_\varepsilon^Y & 0 \end{bmatrix}.$$

Volterra systems appearing in the literature, and in particular in the mathematical finance one, have a specific structure in the sense that only one component satisfies an SVE with (singular) kernel, and can be dealt with independently of the other component. This particular structure allows us to relax some conditions on the coefficients, and we shall leverage on it whenever needed. We make this more specific through the following two definitions:

Definition 2.5. Let $\Upsilon \subset [\![1, d]\!]$ and $\Gamma : \mathbb{R}^{|\Upsilon|} \to \mathbb{R}_+$. We define $\mathcal{S}_\Upsilon^\Gamma$ as the set of functions $f : \mathbb{R}^d \to \mathbb{R}$ for which there exists a strictly positive constant C_Υ such that, for all $x \in \mathbb{R}^d$,

$$|f(x)| \le C_\Upsilon \Big(1 + |x|_{\Upsilon^c} + \Gamma\big(x^{(\Upsilon)}\big)\Big), \tag{2.2}$$

where $|x|_{\Upsilon^c} := \sum_{i\in\Upsilon^c} \left|x^{(i)}\right|$ and $x^{(\Upsilon)} := \{x^{(i)}, i \in \Upsilon\}$.

Definition 2.6. The process X^ε admits an autonomous $\mathcal{S}_\Upsilon^\Gamma$-subsystem $\{X^{\varepsilon,(l)}\}_{l\in\Upsilon}$ if for $\xi \in \{b_\varepsilon, \sigma_\varepsilon\}$,

- if $l \in \Upsilon$, $\xi^{(l)}(t,\cdot)$ has linear growth and does not depend on $X^{\varepsilon,(j)}$ for $j \in \Upsilon^c$;
- if $l \in \Upsilon^c$, $\xi^{(l)}(t, \cdot)$ belongs to $\mathcal{S}_\Upsilon^\Gamma$,

for small enough ε and uniformly in $t \in \mathbb{T}$.

Example 2.7. *The motivation for Definition 2.6 is to be able to handle (rough) stochastic volatility models, ubiquitous in mathematical finance, where linear growth of all the coefficients may not hold. Consider for example the rough Bergomi model [6]*

$$\begin{cases} X_t^{(1)} &= -\dfrac{1}{2}\displaystyle\int_0^t \exp\left(X_s^{(2)}\right)\mathrm{d}s + \int_0^t \exp\left(\frac{1}{2}X_s^{(2)}\right)\mathrm{d}B_s, \\ X_t^{(2)} &= y_0 - at^{2H} + \displaystyle\int_0^t (t-s)^{H-\frac{1}{2}}\mathrm{d}W_s. \end{cases}$$

with $y_0, a \in \mathbb{R}$, $H \in (0, \frac{1}{2})$, B *and* W *are correlated Brownian motions, and we dropped the dependence in* ε. *Here* X *admits* $X^{(2)}$ *as autonomous subsystem with* $\Upsilon = \{2\}$, $\Gamma(x_2) = 1 + \mathrm{e}^{x_2/2}$.

The following set of assumptions, inspired from [22], completes our framework:

H1. X_0^ε converges to $x_0 \in \mathbb{R}^d$ as ε tends to zero.

H2. For all $\varepsilon > 0$ small enough, the coefficients b_ε and σ_ε are measurable maps on $\mathbb{T} \times \mathbb{R}^d$ and converge pointwise to b and σ as ε goes to zero. Moreover, $b(t, \cdot)$ and $\sigma(t, \cdot)$ are continuous on $\mathbb{R}^d$, uniformly in $t \in \mathbb{T}$.

H3. Either **a)** or **b)** holds:

a) For all $\varepsilon > 0$ small enough, b_ε and σ_ε have linear growth uniformly in ε and in $t \in \mathbb{T}$.

b) The process X^ε admits an autonomous $\mathcal{S}_\Upsilon^\Gamma$-subsystem.

H4. The SVE (2.1) is exact for small enough $\varepsilon > 0$.

H2 ensures that, on compact subsets of $\mathbb{T} \times \mathbb{R}^d$, the convergence of b_ε and σ_ε is uniform and that b and σ are uniformly continuous. **H1, H2, H3a** are standard and easily verifiable. **H3b** is unusual but includes a large number of functions; Assumption 3.6 will complete it to indicate the role of Γ such as to include Example 2.7. The main restrictions arise from **H4**, although the latter is satisfied if, for instance, the coefficients b_ε and σ_ε are locally Lipschitz continuous for small enough $\varepsilon > 0$. This requirement was also relaxed in [67] to the one-dimensional case where $K(t) = t^{-\alpha}$, for $\alpha \in (0, \frac{1}{2})$ and $\sigma(x) = x^\gamma$, for $\gamma \in (\frac{1}{2(1-\alpha)}, 1]$, which is clearly not Lispchitz continuous. Although **H4** can seem constraining, to the best of our knowledge there currently exists no pathwise LDP for SDEs where pathwise uniqueness fails.

3. Large and moderate deviations

As discussed in the introduction, our goal is to provide pathwise large and moderate deviations for the general convolution stochastic Volterra system (2.1). The classical Freidlin-Wentzell approach, used in [44], has limitations regarding the behaviour of the coefficients, and the rate function is often rather cumbersome to write. We follow here instead the weak convergence approach developed by Dupuis and Ellis [36]. We first introduce the reader to their abstract setting, and refine the large deviations result by Budhiraja and Dupuis [13] to our general setup. We then show how this abstract framework applies to the small-noise stochastic Volterra system (2.1), first proving pathwise large deviations, and then the moderate deviations counterpart.

3.1. **Weak convergence approach: the abstract setting.** Given a family of Borel-measurable functions $\{\mathcal{G}^\varepsilon\}_{\varepsilon>0}$ from $\mathcal{W}^m$ to $\mathcal{W}^d$, we enquire about the large deviations behaviour of the family of random variables $\{\mathcal{G}^\varepsilon(W)\}_{\varepsilon>0}$, where W is a standard Brownian motion on the filtered probability space above. For each $N > 0$, the spaces of bounded stochastic and deterministic controls

$$\mathcal{S}_N := \left\{ v \in L^2 : \int_0^T |v_s|^2 \mathrm{d}s \le N \right\} \qquad \text{and} \qquad \mathcal{A}_N := \left\{ v \in \mathcal{A} : v \in \mathcal{S}_N \text{ almost surely} \right\}, \tag{3.1}$$

with $\mathcal{A}$ introduced in (1.4), are equipped with the weak topology on $L^2(\mathbb{T} \times \Omega)$ such that they are closed and even compact (by Banach-Alaoglu-Bourbaki theorem). Budhiraja and Dupuis [13] assume, for any sequence $\{v^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{A}_N$ converging weakly to $v \in \mathcal{A}_N$, the existence of a limit in distribution of $\mathcal{G}^\varepsilon \left(W + \frac{1}{\varepsilon} \int_0^\cdot v_s^\varepsilon \mathrm{d}s \right)$ which is uniquely characterised by v. However, such uniqueness may fail when the coefficients of the system (1.1) (in particular the diffusion coefficient σ) are not locally Lipschitz, as is the case for the Feller diffusion for example (in this case without singular kernel, a dedicated analysis was carried out in [27, 35] using the Freidlin-Wentzell approach). We relax here this uniqueness assumption by replacing the limiting trajectory by a perturbed version.

Let $\phi \in \mathcal{W}^d$. For a sequence of controls $\{v^n\}_{n\in\mathbb{N}} \subset L^2$, we consider the abstract sets $\mathcal{G}^0_{v^n} \subset \mathcal{W}^d$ for all $n \in \mathbb{N}$, together with the conditions

$$\phi \in \bigcap_{n\in\mathbb{N}} \mathcal{G}^0_{v^n} \qquad \text{and} \qquad \frac{1}{2}\int_0^T |v^n_s|^2 \mathrm{d}s \le I(\phi) + \frac{1}{n}, \quad \text{for all } n \in \mathbb{N}, \tag{3.2}$$

where the functional $I : \mathcal{W}^d \to \mathbb{R}_+$ is given by

$$I(\phi) := \inf \left\{ \frac{1}{2} \int_0^T |v_s|^2 \mathrm{d}s : \phi \in \mathcal{G}^0_v, v \in L^2 \right\}. \tag{3.3}$$

The sequence of controls satisfying condition (3.2) exists if $I(\phi) < \infty$ and approaches the infimum in (3.3). We will say that ϕ is *uniquely characterised* if there exists a sequence $\{v^n\}_{n\in\mathbb{N}} \subset L^2$ satisfying (3.2) and such that for any $n \ge 0$, there exists $m \ge n$ such that $\mathcal{G}^0_{v^m}$ is a singleton; in that case $\mathcal{G}^0_{v^m} = \{\phi\}$. In particular, if there exists $\widetilde{v} \in L^2$ which attains the infimum in (3.3) and $\mathcal{G}^0_{\widetilde{v}} = \{\phi\}$ then ϕ is uniquely characterised.

Assumption 3.1. For any $\delta > 0$ and any $\phi \in \mathcal{W}^d$ such that $I(\phi) < \infty$, there exists ϕ^δ uniquely characterised such that $\left\|\phi - \phi^\delta\right\|_{\mathbb{T}} \le \delta$ and $\left|I(\phi) - I(\phi^\delta)\right| \le \delta$.

Remark 3.2. *This assumption is reminiscent of* [35, Proposition 3.3]*, where the authors resolve the non-uniqueness issue in the diffusion case. A similar problem is also at the core of* [19, Lemma 5.1] *in an infinite-dimensional setting.*

Our abstract large deviations result is the following, extending [13, Theorem 4.4], at least when the underlying Hilbert space is $\mathbb{R}^m$, to the non-uniqueness case:

Theorem 3.3. *Consider $N > 0$ and a family $\{v^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{A}_N$ converging in distribution to $v \in \mathcal{A}_N$. Assume that there exists a random set $\mathcal{G}^0_v \subset \mathcal{W}^d$ such that:*

(i) $\mathcal{G}^\varepsilon\left(W_\cdot + \varepsilon^{-1}\int_0^\cdot v^\varepsilon_s \mathrm{d}s\right)$ *converges in distribution to an element of $\mathcal{G}^0_v$.*
(ii) The functional I defined by (3.3) *has compact level sets.*
(iii) Assumption 3.1 holds.

Then the family $\{\mathcal{G}^\varepsilon(W)\}_{\varepsilon>0}$ satisfies the Laplace principle and, by equivalence, the Large Deviations Principle with rate function I and speed ε^{-1}.

Remark 3.4. *In the large deviations literature, a rate function is sometimes called 'good' if it has compact level sets. All the rate functions in the present paper satisfy this requirement (item (ii) above takes care of that), therefore we drop the adjective 'good'.*

We defer the proof to Appendix A.1; the lower bound can be tackled as in [13], and we therefore concentrate on the upper bound. The idea is that the Laplace principle (1.2) upper bound involves an infimum so deriving it only requires a δ-optimal path. Hence a perturbation will also do the trick, provided one knows how to handle the control associated to it. In [13, Theorem 4.4], unique characterisation of the limiting element in (i) is granted, and the set $\mathcal{G}^0_v$ is a singleton that takes the form $\mathcal{G}^0(\int_0^\cdot v_s \mathrm{d}s)$, where they view $\mathcal{G}^0$ as a map. In that case Assumption 3.1 is clearly satisfied since ϕ^δ can be taken as ϕ itself.

3.2. **Application to stochastic Volterra systems.** We now show how the abstract setting developed above in Section 3.1 applies to the small-noise stochastic Volterra system (2.1) and why pathwise uniqueness is so fundamental. If **H4** holds, define the functional $\mathcal{G}^\varepsilon$ as the Borel-measurable map associating the multidimensional Brownian motion W to the solution of the stochastic Volterra equation (2.1), that is: $\mathcal{G}^\varepsilon(W) = X^\varepsilon$. For any control $v \in \mathcal{A}_N$, $N > 0$ (introduced in (3.1)) and any $\varepsilon > 0$, the process $\widetilde{W} := W + \vartheta_\varepsilon^{-1}\int_0^\cdot v_s \mathrm{d}s$ is a $\widetilde{\mathbb{P}}$-Brownian motion by Girsanov's theorem, where

$$\frac{\mathrm{d}\widetilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}} := \exp\left\{-\frac{1}{\vartheta_\varepsilon}\sum_{i=1}^m \int_0^T v_s^{(i)} \mathrm{d}W_s^{(i)} - \frac{1}{2\vartheta_\varepsilon^2}\int_0^T |v_s|^2 \mathrm{d}s\right\}.$$

Hence the shifted version $X^{\varepsilon,v} := \mathcal{G}^\varepsilon(\widetilde{W})$ appearing in Theorem 3.3(i) is the strong unique solution of (2.1) under $\widetilde{\mathbb{P}}$, with X^ε and W replaced by $X^{\varepsilon,v}$ and $\widetilde{W}$. Because $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent, $X^{\varepsilon,v}$ is also the unique strong solution of the controlled equation

$$X_t^{\varepsilon,v} = X_0^\varepsilon + \int_0^t K(t-s)\Big[b_\varepsilon(s, X_s^{\varepsilon,v}) + \sigma_\varepsilon(s, X_s^{\varepsilon,v})v_s\Big]\mathrm{d}s + \vartheta_\varepsilon\int_0^t K(t-s)\sigma_\varepsilon(s, X_s^{\varepsilon,v})\mathrm{d}W_s. \tag{3.4}$$

Under appropriate conditions, and using the notations set in **H1, H2**, we heuristically observe that taking ε to zero, the system (3.4) reduces to the deterministic Volterra equation

$$\phi_t = x_0 + \int_0^t K(t-s)\Big[b(s,\phi_s) + \sigma(s,\phi_s)v_s\Big]\mathrm{d}s, \tag{3.5}$$

solutions of which precisely corresponds to the set $\mathcal{G}_0^v$ in Theorem 3.3.

Example 3.5. *To illustrate the need for a set $\mathcal{G}^0_v$ rather than a singleton, consider the Feller diffusion*

$$X_t = x_0 + \kappa\int_0^t (\theta - X_s)\mathrm{d}s + \int_0^t \sqrt{X_s}\mathrm{d}W_s,$$

for $t \in \mathbb{T}$, with $x_0, \kappa, \theta > 0$. Letting $t \mapsto \varepsilon t$ and denoting $X_t^\varepsilon := X_{\varepsilon t}$ yields, by scaling,

$$X_t^\varepsilon = x_0 + \kappa\varepsilon \int_0^t (\theta - X_s^\varepsilon)\,\mathrm{d}s + \sqrt{\varepsilon}\int_0^t \sqrt{X_s^\varepsilon}\mathrm{d}W_s,$$

which is exactly (2.1) with $d = 1$, $K \equiv 1$, $\vartheta_\varepsilon = \sqrt{\varepsilon}$, $b_\varepsilon(x) \equiv \kappa\varepsilon(\theta - x)$, $\sigma_\varepsilon(x) \equiv \sqrt{x}$. For $v \in L^2$, taking limits as ε tends to zero in the corresponding controlled equation (3.4) yields (3.5), or

$$\phi_t = x_0 + \int_0^t \sqrt{\phi_s} v_s \mathrm{d}s, \quad t \in \mathbb{T}.$$

Uniqueness does not hold in general because of the non-Lipschitz coefficient, and thus $\mathcal{G}_v^0$ corresponds to the set of non-negative solutions. Consider for example $x_0 = 1$, $\mathbb{T} = [0, 4]$ and the control

$$v_t := \begin{cases} -1, & \text{if } t \in [0, 2) \\ 1, & \text{if } t \in [2, 4]. \end{cases}$$

The function $\phi_t := \frac{(t-2)^2}{4}$ is clearly a solution, but so is φ equal to ϕ on $[0, 2]$ and null on $[2, 4]$. The square root function is indeed locally Lipschitz away from zero, and uniqueness can thus be guaranteed as long as the solution remains positive. The perturbation $\phi^\delta := \phi + \delta t$ is now the unique solution to

$$\phi_t^\delta = 1 + \int_0^t \sqrt{\phi_s^\delta} v_s^\delta \mathrm{d}s,$$

for all $t \in [0, 4]$, where $v^\delta := \dot{\phi}^\delta / \sqrt{\phi^\delta}$. The infimum in (3.3) is attained by v^δ and $\mathcal{G}_{v^\delta}^0 = \{\phi^\delta\}$, thus ϕ^δ is uniquely characterised. Furthermore, [35, Proposition 3.3] *shows that ϕ^δ satisfies Assumption 3.1.*

In [9], the authors were also confronted to a limiting equation with multiple solutions. Instead of perturbing the path ϕ, they perturb the control in a way that the resulting equation has a unique solution which is precisely ϕ, i.e. $\mathcal{G}_{v^\delta}^0 = \{\phi\}$. This approach may seem more natural; however, it is not always obvious how to perturb the control ensuring uniqueness of the ODE, while our formulation makes it more straightforward. Before stating the main large and moderate deviations results for small-noise stochastic Volterra equations, we introduce the following assumption, monitoring the moments of the controlled equation:

Assumption 3.6. Let $X^{\varepsilon,v}$ be the exact solution to (3.4). If **H3a** holds then the present assumption is satisfied. If instead **H3b** holds, then there exists $\varepsilon_0 > 0$ such that, for any $p \geq 1$ and $N > 0$,

$$\sup\left\{\mathbb{E}\left[\left|\Gamma\big((X_t^{\varepsilon,v})^{(\Upsilon)}\big)\right|^p\right], t \in \mathbb{T}, v \in \mathcal{A}_N, \varepsilon \in (0, \varepsilon_0)\right\} < \infty, \tag{3.6}$$

$$\sup\left\{\left|\Gamma\big((\phi_t)^{(\Upsilon)}\big)\right|, t \in \mathbb{T}, v \in \mathcal{S}_N, \phi \in \mathcal{G}_v^0\right\} < \infty. \tag{3.7}$$

Remark 3.7. *In the following,* **H3b** *will always be complemented by Assumption 3.6.*

3.3. **Large Deviations.** Armed with the abstract setting in Section 3.1, and its application to the stochastic Volterra system (2.1) in Section 3.2, we can at last show large deviations for the latter:

Theorem 3.8 (Large Deviations)**.** *Under* **H1 - H4***, Assumptions 2.3, 3.1 and 3.6, the family $\{X^\varepsilon\}_{\varepsilon>0}$, unique solution of (2.1), satisfies a Large Deviations Principle with rate function (3.3) and speed $\vartheta_\varepsilon^{-1}$, where $\mathcal{G}_v^0$ is the set of solutions of the limiting equation (3.5).*

Remark 3.9. *We recall that Assumption 3.1 is immediately satisfied if the limiting equation (3.5) has a unique solution. Also, it is only necessary to check Assumption 3.6 if* **H3a** *does not hold.*

3.3.1. *Technical preliminary results.* The proof will rely on the following results: Lemma 3.10 (proved in Section A.2) shows the moment bound of the controlled process, Lemma 3.12 (proved in Section A.3) demonstrates the tightness and Lemma 3.14 (proved in Section A.4) deals with the compactness of the level sets of the rate function. They will then allow the use of Theorem 3.3.

Lemma 3.10 (LDP Moment bound)**.** *Under* **H1 - H4***, Assumptions 2.3 and 3.6, for all* $p \geq 2$*,* $N > 0$ *and* $v \in \mathcal{A}_N$*, there exists a constant* $\overline{c} > 0$ *independent of* ε, v, t *such that for all* $\varepsilon > 0$ *small enough*

$$\sup_{t\in\mathbb{T}} \mathbb{E}\left[|X_t^{\varepsilon,v}|^p\right] \leq \overline{c}. \tag{3.8}$$

Remark 3.11. *This bound also holds for any solution* ϕ *of* (3.5) *using the bound* (3.7)*. Therefore we also obtain* $\sup\{\|\phi\|_{\mathbb{T}}, v \in \mathcal{S}_N, \phi \in \mathcal{G}_v^0\} \leq \overline{c}'$*.*

Lemma 3.12 (LDP Tightness)**.** *Consider* **H1 - H4***, Assumptions 2.3 and 3.6. Let* $p > 2\vee 2/\gamma$*,* $N > 0$ *and a family* $\{v^\varepsilon\}_{\varepsilon>0}$ *in* $\mathcal{A}_N$*. Then* $X^{\varepsilon,v^\varepsilon}$ *admits a version which is Hölder continuous on* $\mathbb{T}$ *of any order* $\alpha < \gamma/2 - 1/p$*, uniformly for all* $\varepsilon > 0$*. Denoting again this version by* $X^{\varepsilon,v^\varepsilon}$*, one has for all* $\varepsilon > 0$ *small enough*

$$\mathbb{E}\left[\left(\sup_{0\leq s<t\leq T} \frac{\left|X_t^{\varepsilon,v^\varepsilon} - X_s^{\varepsilon,v^\varepsilon}\right|}{|t-s|^\alpha}\right)^p\right] \leq \overline{C}, \tag{3.9}$$

for all $\alpha \in [0, \gamma/2 - 1/p)$*, where* $\overline{C}$ *is a constant independent of* $\varepsilon, v^\varepsilon, t$*. Moreover, the family of random variables* $\{X^{\varepsilon,v^\varepsilon}\}_{\varepsilon>0}$ *is tight in* $\mathcal{W}^d$*.*

Remark 3.13. *This lemma entails that for all* $N > 0, v \in \mathcal{S}_N$ *any solution to* (3.5) *also has Hölder continuous paths of the same order.*

The following lemma proves Theorem 3.3(ii) and its proof can be found in Appendix A.4.

Lemma 3.14 (LDP Compactness)**.** *Under* **H2, H3***, Assumptions 2.3 and 3.6, the functional* I *in* (3.3) *has compact level sets.*

We now have all the ingredients to prove the Large Deviations Principle.

3.3.2. *Proof of Theorem 3.8.* Fix $N > 0$. Consider a family $\{v^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{A}_N$ converging in distribution to $v \in \mathcal{A}_N$. We take an arbitrary subsequence $\{v_{\varepsilon_n}\}_{n\in\mathbb{N}}$ and prove convergence along a subsequence thereof. If every subsequence has a converging subsequence then the sequence converges.

For $\varepsilon > 0$ small enough, the SVE (2.1) is exact by **H4** and we showed that its controlled counterpart (3.4) also has a unique strong solution $X^{\varepsilon,v^\varepsilon}$. Lemma 3.12 shows that the family $\{X^{\varepsilon_n,v^{\varepsilon_n}}\}_{n\geq 0}$ is tight in $\mathcal{W}^d$. Moreover, the trajectories of v^{ε_n} belong to a compact space with respect to the weak topology so the family of controls is tight as a sequence of $\mathcal{S}_N$-valued random variables. Since these are both Polish spaces, the family $\left\{X^{\varepsilon_n,v^{\varepsilon_n}}, v^{\varepsilon_n}\right\}_{n\geq 0}$ is tight in $\mathcal{W}^d \times \mathcal{S}_N$. Hence there exists a subsequence, denoted hereafter $\left\{X^n, v^n\right\}$, that converges weakly to a $\mathcal{W}^d \times \mathcal{S}_N$-valued random variable (X^0, v) defined on a possibly different probability space $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ as n tends to $+\infty$. We also denote $\varepsilon_n, X_0^n, b_n, \sigma_n$ along this subsequence. We follow the technique in [22] to identify the limit. For $t \in \mathbb{T}$, define $\Phi_t : \mathcal{S}_N \times \mathcal{W}^d \to \mathbb{R}$ as

$$\Phi_t(f,\omega) := \left|\omega_t - x_0 - \int_0^t K(t-s)\Big[b(s,\omega_s) + \sigma(s,\omega_s)f_s\Big]\mathrm{d}s\right| \wedge 1.$$

Clearly, Φ_t is bounded and we show that it is also continuous. Indeed, let $\omega^n \to \omega$ in $\mathcal{W}^d$ and $f^n \to f$ in $\mathcal{S}_N$ with respect to the weak topology. **H2** implies the existence of continuous moduli of continuity ρ_b and ρ_σ for both coefficients on compact subsets (see Definition 2.1). Since the paths $\omega^n, n \geq 1$

and ω are continuous, these moduli are available. Then, using Cauchy-Schwarz inequality and the fact that $|x \wedge 1 - y \wedge 1| \le |x-y|$ for all $x, y > 0$,

$$\begin{aligned}|\Phi_t(f,\omega) - \Phi_t(f^n,\omega^n)| &\le |\omega_t - \omega_t^n| + \int_0^t |K(t-s)|\,|b(s,\omega_s) - b(s,\omega_s^n)|\,\mathrm{d}s \\ &\quad + \int_0^t |K(t-s)|\,\big|\big(\sigma(s,\omega_s) - \sigma(s,\omega_s^n)\big)f_s^n + \sigma(s,\omega_s)\big(f_s - f_s^n\big)\big|\,\mathrm{d}s \\ &\le \|\omega - \omega^n\|_{\mathbb{T}} + \|\rho_b(|\omega-\omega^n|)\|_{\mathbb{T}}\,\|K\|_1 + \|\rho_\sigma(|\omega-\omega^n|)\|_{\mathbb{T}}\,\|K\|_2\,\|f^n\|_2 \\ &\quad + \|\sigma(\cdot,\omega)\|_{\mathbb{T}}\int_0^t |K(t-s)|\,|f_s - f_s^n|\,\mathrm{d}s.\end{aligned}$$

Since $K(t-\cdot) \in L^2$ and f_n tends to f weakly in L^2 then the last integral converges to zero as n goes to infinity. Moreover $\lim_{n\uparrow\infty} \|\omega - \omega^n\|_{\mathbb{T}} = 0$, $\|f^n\|_2 \le \sqrt{N}$ for all $n \ge 0$ and $\|K\|_2 + \|\sigma(\cdot,\omega)\|_{\mathbb{T}} < \infty$, this proves that Φ_t is continuous, and therefore

$$\lim_{n\uparrow\infty} \mathbb{E}\left[\Phi_t(v^n, X^n)\right] = \mathbb{E}^0\left[\Phi_t(v, X^0)\right].$$

We now prove that the left-hand side is actually equal to zero. We start with the observation that, using BDG inequality,

$$\begin{aligned}\mathbb{E}\left[\Phi_t(v^n,X^n)\right] &\le |X_0^n - x_0| + \int_0^t |K(t-s)|\,\mathbb{E}\big[\,|b_n(s,X_s^n) - b(s,X_s^n)|\,\big]\mathrm{d}s \\ &\quad + \int_0^t |K(t-s)|\,\mathbb{E}\left[|\sigma_n(s,X_s^n) - \sigma(s,X_s^n)|\,|v_s^n|\right]\mathrm{d}s \\ &\quad + \vartheta_{\varepsilon_n}\mathbb{E}\left[\int_0^t |K(t-s)\sigma_n(s,X_s^n)|^2\,\mathrm{d}s\right]^{\frac{1}{2}}. \end{aligned} \tag{3.10}$$

The bounds (A.3) and (A.5) show how to control the last term under **H3a** and **H3b** respectively, hence there exists $C_1 > 0$ independent of t and n such that $\mathbb{E}\left[\int_0^t |K(t-s)\sigma_n(s,X_s^n)|^2\,\mathrm{d}s\right]^{\frac{1}{2}} \le C_1$.

However the convergence of b_n, σ_n only occurs on compact subsets so we use a localisation argument. For all $n \ge 0$, $M > 0$ we introduce

$$A_n^M := \Big\{\omega \in \Omega, \|X^n(\omega)\|_{\mathbb{T}} > M\Big\}.$$

The uniform (in $n \in \mathbb{N}$) Hölder regularity of X^n, encompassed by (3.9), entails the existence, for all $p > 2 \vee 2/\gamma$, of $C_2(p), C_3(p) > 0$ independent of n such that

$$\mathbb{E}\left[\sup_{t\in\mathbb{T}} |X_t^n|^p\right] \le C_2(p)\big(\,|X_0^n|^p + T^{p\alpha}\big) \le C_3(p),$$

for some $0 < \alpha < \gamma/2 - 1/p$ and where X_0^n is uniformly bounded by $2|x_0|$ for n large enough. Markov's inequality then implies that

$$\lim_{M\uparrow\infty}\sup_{n\in\mathbb{N}} \mathbb{P}\big(A_n^M\big) \le \lim_{M\uparrow\infty}\sup_{n\in\mathbb{N}} \frac{C_3(p)}{M^p} = 0. \tag{3.11}$$

Moreover, for all $n \in \mathbb{N}$, $\omega \in \Omega \setminus A_n^M$, and $t \in \mathbb{T}$, $|X_t^n(\omega)|$ is bounded by M, which means

$$|b_n(t, X_t^n(\omega)) - b(t, X_t^n(\omega))| \le \|b_n(t,\cdot) - b(t,\cdot)\|_M\,,$$

which tends to zero uniformly on $\mathbb{T}$ as n goes to infinity (and likewise for σ_n) from **H2**. Define then

$$I_n := \int_0^t |K(t-s)|\,\Big(\,|b_n(s,X_s^n) - b(s,X_s^n)| + |\sigma_n(s,X_s^n) - \sigma(s,X_s^n)|\,|v_s^n|\,\Big)\mathrm{d}s$$

and observe that, using Jensen's and Cauchy-Schwarz inequalities, the growth condition on the coefficients from **H3** and the moment bounds on X^n from (3.8), there exists $C_4 > 0$ independent of n such that

$$\begin{aligned}\mathbb{E}\big[\,|I_n|^2\,\big] &\leq 2t\int_0^t |K(t-s)|^2\,\mathbb{E}\big[\,|b_n(s,X_s^n)-b(s,X_s^n)|^2\,\big]\mathrm{d}s\\ &\quad + 2N\int_0^t |K(t-s)|^2\,\mathbb{E}\left[|\sigma_n(s,X_s^n)-\sigma(s,X_s^n)|^2\right]\mathrm{d}s\\ &\leq 2\,\|K\|_2^2 \sup_{s\leq t}\left\{t\mathbb{E}\big[\,|b_n(s,X_s^n)-b(s,X_s^n)|^2\,\big]+N\mathbb{E}\left[|\sigma_n(s,X_s^n)-\sigma(s,X_s^n)|^2\right]\right\}\\ &\leq C_4. \qquad (3.12)\end{aligned}$$

Let us fix $\epsilon > 0$ and choose $M_\epsilon > 0$ large enough such that $\sup_{n\in\mathbb{N}}\mathbb{P}\big(A_n^{M_\epsilon}\big) \leq \epsilon^2/C_4$; this choice is possible because of (3.11). Therefore, using the bound (3.12) and Cauchy-Schwarz inequality to separate I_n and $\mathbb{1}_{A_n^{M_\varepsilon}}$, one obtains

$$\begin{aligned}\lim_{n\uparrow\infty}\mathbb{E}[I_n] &= \lim_{n\uparrow\infty}\mathbb{E}\Big[I_n\big(\mathbb{1}_{A_n^{M_\varepsilon}}+\mathbb{1}_{\Omega\setminus A_n^{M_\varepsilon}}\big)\Big]\\ &\leq \lim_{n\uparrow\infty}\left\{\sqrt{C_4\mathbb{P}\big(A_n^{M_\varepsilon}\big)}+\|K\|_1\,\|\|b_n-b\|_M\|_{\mathbb{T}}+\sqrt{N}\,\|K\|_2\,\|\|\sigma_n-\sigma\|_M\|_{\mathbb{T}}\right\}\leq\epsilon.\end{aligned}$$

It follows from (3.10) that

$$\lim_{n\uparrow\infty}\mathbb{E}\big[\Phi_t(v^n,X^n)\big]\leq\lim_{n\uparrow\infty}\Big\{\,|X_0^n-x_0|+\mathbb{E}[I_n]+\vartheta_{\varepsilon_n}C_1\Big\}\leq\epsilon,$$

hence $\lim_{n\uparrow\infty}\mathbb{E}\left[\Phi_t(v^n,X^n)\right]=0$ since $\epsilon>0$ was chosen arbitrarily. The equality $\mathbb{E}^0\left[\Phi_t(v,X^0)\right]=0$ implies that X^0 satisfies (3.5) $\mathbb{P}^0$-almost surely, for all $t\in\mathbb{T}$. Since X^0 has continuous paths, it satisfies (3.5) for all $t\in\mathbb{T}$, $\mathbb{P}^0$-almost surely. We have proved weak convergence and compactness of level sets of the rate function from Proposition 3.14, therefore Theorem 3.3 yields the claim and concludes the proof.

3.4. **Moderate Deviations.** Let h_ε tend to infinite such that $\vartheta_\varepsilon h_\varepsilon$ tends to zero as ε goes to zero and define $\overline{X}$ to be the limit in law of X^ε, which we identified in the previous subsection as a solution of the Volterra equation

$$\overline{X}_t = x_0+\int_0^t K(t-s)b(s,\overline{X}_s)\mathrm{d}s. \qquad (3.13)$$

Then the MDP for $\{X^\varepsilon\}_{\varepsilon>0}$ is equivalent to the LDP for the family $\{\eta^\varepsilon\}_{\varepsilon>0}$ defined as

$$\eta^\varepsilon := \frac{X^\varepsilon-\overline{X}}{\vartheta_\varepsilon h_\varepsilon}=\frac{\mathcal{G}^\varepsilon(W)-\overline{X}}{\vartheta_\varepsilon h_\varepsilon}=:\mathcal{T}^\varepsilon(W),$$

where $\mathcal{T}^\varepsilon:\mathcal{W}^m\to\mathcal{W}^d$ are Borel-measurable maps for each $\varepsilon>0$. Therefore η^ε satisfies the following SVE for all $\varepsilon>0$, and is its unique solution if **H4** holds.

$$\eta_t^\varepsilon=\frac{X_0^\varepsilon-x_0}{\vartheta_\varepsilon h_\varepsilon}+\int_0^t K(t-s)\frac{b_\varepsilon\big(s,\overline{X}_s+\vartheta_\varepsilon h_\varepsilon\eta_s^\varepsilon\big)-b(s,\overline{X}_s)}{\vartheta_\varepsilon h_\varepsilon}\mathrm{d}s+\int_0^t K(t-s)\frac{\sigma_\varepsilon\big(s,\overline{X}_s+\vartheta_\varepsilon h_\varepsilon\eta_s^\varepsilon\big)}{h_\varepsilon}\mathrm{d}W_s.$$

Similarly to the LDP case we are interested in a certain shift of the driving Brownian motion, controlled by $v\in\mathcal{A}$. For all $\varepsilon>0$, let

$$\eta^{\varepsilon,v}:=\mathcal{T}^\varepsilon\left(W+h_\varepsilon\int_0^\cdot v_s\mathrm{d}s\right)=\frac{\mathcal{G}^\varepsilon\big(W+h_\varepsilon\int_0^\cdot v_s\mathrm{d}s\big)-\overline{X}}{\vartheta_\varepsilon h_\varepsilon}. \qquad (3.14)$$

For convenience we introduce the family $\{\Theta^{\varepsilon,v}\}_{\varepsilon>0}$ defined, for all $\varepsilon > 0$, $v \in \mathcal{A}$, $t \in \mathbb{T}$, by

$$\begin{aligned}\Theta_t^{\varepsilon,v} :=& \mathcal{G}^\varepsilon\left(W + h_\varepsilon \int_0^\cdot v_s \mathrm{d}s\right)(t)\\ =& X_0^\varepsilon + \int_0^t K(t-s)\Big[b_\varepsilon(s,\Theta_s^{\varepsilon,v}) + \vartheta_\varepsilon h_\varepsilon \sigma_\varepsilon(s,\Theta_s^{\varepsilon,v})v_s\Big]\mathrm{d}s + \vartheta_\varepsilon \int_0^t K(t-s)\sigma_\varepsilon(s,\Theta_s^{\varepsilon,v})\mathrm{d}W_s.\end{aligned}$$

This sequence satisfies the bound (3.8) and converges weakly towards $\overline{X}$ since $\vartheta_\varepsilon h_\varepsilon \sigma_\varepsilon$ tends to zero as ε tends to zero. Finally the process defined by (3.14) satisfies

$$\begin{aligned}\eta_t^{\varepsilon,v} =& \frac{X_0^\varepsilon - x_0}{\vartheta_\varepsilon h_\varepsilon} + \int_0^t K(t-s)\frac{b_\varepsilon\big(s,\overline{X}_s + \vartheta_\varepsilon h_\varepsilon \eta_s^{\varepsilon,v}\big) - b(s,\overline{X}_s)}{\vartheta_\varepsilon h_\varepsilon}\mathrm{d}s\\ &+ \int_0^t K(t-s)\sigma_\varepsilon\big(s,\overline{X}_s + \vartheta_\varepsilon h_\varepsilon \eta_s^{\varepsilon,v}\big)v_s\mathrm{d}s + \frac{1}{h_\varepsilon}\int_0^t K(t-s)\sigma_\varepsilon\big(s,\overline{X}_s + \vartheta_\varepsilon h_\varepsilon \eta_s^{\varepsilon,v}\big)\mathrm{d}W_s.\end{aligned}\tag{3.15}$$

For all $v \in \mathcal{A}$, we define $\mathcal{T}_v^0$ to be the solution of the limiting equation

$$\psi_t = \int_0^t K(t-s)\Big[\nabla b(s,\overline{X}_s)\psi_s + \sigma(s,\overline{X}_s)v_s\Big]\mathrm{d}s.\tag{3.16}$$

The form of the limit equation is dramatically simpler than for the LDP and much easier to compute. Moreover $\mathcal{T}_v^0$ is well defined because the linearity of the equation and Assumption 2.3 grant uniqueness for free, provided ∇b exists. Hence we will need the following assumptions:

H5. For each $t \in \mathbb{T}$, the function $b(t,\cdot)$ is continuously differentiable and b is Lipschitz continuous.
H6. There exists $\delta > 0$ such that $\sigma(t,\cdot)$ is locally δ-Hölder continuous, uniformly for all $t \in \mathbb{T}$.
H7. $\lim_{\varepsilon\downarrow 0}\big(\vartheta_\varepsilon h_\varepsilon\big)^{-1}|X_0^\varepsilon - x_0| = 0$.
H8. There exist $\varepsilon_0 > 0$, a sequence $\{\nu_\varepsilon\}_{\varepsilon>0}$ with $\lim_{\varepsilon\downarrow 0}\nu_\varepsilon(\vartheta_\varepsilon h_\varepsilon)^{-1} = 0$ and a function $\Xi : \mathbb{R}^d \to \mathbb{R}$ such that $|b_\varepsilon(t,x) - b(t,x)| \le \nu_\varepsilon \Xi(x)$ for all $t \in \mathbb{T}$, $\varepsilon \in (0,\varepsilon_0)$, where for all $p \ge 1$, $N > 0$,

$$\sup\left\{\mathbb{E}\Big[\big|\Xi\big(\Theta_t^{\varepsilon,v}\big)\big|^p\Big], \varepsilon \in (0,\varepsilon_0), v \in \mathcal{A}_N, t \in \mathbb{T}\right\} < \infty.\tag{3.17}$$

Remark 3.15. **H5** *entails that* (3.13) *has a unique solution and yields the bound* $\big\|\nabla b(\cdot,\overline{X})\big\|_{\mathbb{T}} < \infty$ *by continuity.* **H7** *implies* **H1**. *We have already proved the moments of all orders of* $\Theta^{\varepsilon,v}$ *are bounded in Lemma 3.8 hence* (3.17) *is satisfied if, for instance,* Ξ *is of polynomial growth. This is however not sufficient for the applications we have in mind where* Ξ *is of exponential growth.*

The main theorem of this section is the following.

Theorem 3.16 (Moderate Deviations). *Under* **H2 - H8**, *Assumptions 2.3 and 3.6, the family* $\{\eta^\varepsilon\}_{\varepsilon>0}$ *satisfies a Large Deviations Principle (equivalently* $\{X^\varepsilon\}_{\varepsilon>0}$ *satisfies a Moderate Deviations Principle) with speed* h_ε *and rate function*

$$\Lambda(\psi) := \inf\left\{\frac{1}{2}\int_0^T |v_t|^2\,\mathrm{d}t : v \in L^2, \psi = \mathcal{T}_v^0\right\},\tag{3.18}$$

and $\Lambda(\psi) = +\infty$ *if this set is empty.*

The proof of the moderate deviations theorem follows a similar structure to that of Theorem 3.8, making use of Theorem 3.3. It will rely on moment bounds in Lemma 3.17 (proved in Section B.1), tightness in Lemma 3.18 (proved in Section B.2), weak convergence in Lemma 3.19 (proved in Section B.3), and finally compactness of the level sets in Lemma 3.20.

Lemma 3.17 (MDP Moment bound). *Under* **H2 - H5, H7, H8**, *Assumptions 2.3 and 3.6, for all* $p \ge 2$, $N > 0$ *and* $v \in \mathcal{A}_N$, *there exists* $\widehat{c} > 0$ *independent of* ε, v, t *such that for* $\varepsilon > 0$ *small enough*

$$\sup_{t\in\mathbb{T}}\mathbb{E}\left[|\eta_t^{\varepsilon,v}|^p\right] \le \widehat{c}.$$

Lemma 3.18 (MDP tightness)**.** *Let $p > 2 \vee 2/\gamma$, $N > 0$ and a family $\{v^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{A}_N$. Under* **H2 - H5, H7, H8***, Assumptions 2.3 and 3.6, $\eta^{\varepsilon,v^\varepsilon}$ admits a version which is Hölder continuous on $\mathbb{T}$ of any order $\alpha < \gamma/2 - 1/p$, uniformly for all $\varepsilon > 0$. Denoting again this version by $\eta^{\varepsilon,v^\varepsilon}$, one has for all $\varepsilon > 0$ small enough,*

$$\mathbb{E}\left[\left(\sup_{0\le s<t\le T} \frac{\left|\eta_t^{\varepsilon,v^\varepsilon} - \eta_s^{\varepsilon,v^\varepsilon}\right|}{|t-s|^\alpha}\right)^p\right] \le \widehat{C}, \tag{3.19}$$

for all $\alpha \in [0, \gamma/2 - 1/p)$, where $\widehat{C}$ is a constant independent of $\varepsilon, v^\varepsilon, s, t$. Moreover, the family $\{\eta^{\varepsilon,v^\varepsilon}\}_{\varepsilon>0}$ is tight in $\mathcal{W}^d$.

Lemma 3.19 (MDP weak convergence)**.** *Let $N > 0$, a family $\{v^\varepsilon\}_{\varepsilon>0}$ such that, for all $\varepsilon > 0$, $v^\varepsilon \in \mathcal{A}_N$ and v^ε converges in distribution to $v \in \mathcal{A}_N$, and ψ the unique solution of* (3.16)*. Under* **H2 - H8***, Assumptions 2.3 and 3.6, $\eta^{\varepsilon,v^\varepsilon}$ converges in distribution to ψ as ε goes to zero.*

We recall that $\eta^{\varepsilon,v^\varepsilon} = \mathcal{T}^\varepsilon\big(W + h_\varepsilon \int_0^\cdot v_s \mathrm{d}s\big)$ and $\psi = \mathcal{T}_v^0$, hence the weak convergence we just proved deals with Theorem 3.3 *(i)*. Item *(ii)* is dealt with in the following lemma.

Lemma 3.20 (MDP compactness)**.** *Under* **H2, H3, H5***, Assumptions 2.3 and 3.6, the functional Λ defined by* (3.18) *has compact level sets.*

Proof. Noticing that $\nabla b(t, \overline{X}_t)$ and $\sigma(t, \overline{X}_t)$ are uniformly bounded on $\mathbb{T}$ by continuity, this lemma amounts to a particular case of Lemma 3.14. □

Theorem 3.3(iii) is immediate by uniqueness of (3.16), therefore all the conditions are met and Theorem 3.16 follows as a direct application of Theorem 3.3.

3.5. **Extension to non-convolution kernels.** The analysis undertaken in this paper is based, both for notational convenience and with a view towards application, on convolution kernels. Different assumptions were studied in the literature, in particular Decreusefond [32] considered the properties of the map $f \mapsto \int_0^\cdot K(\cdot, s) f(s)\mathrm{d}s$ in order to include the fractional Brownian motion in his setting.

3.5.1. *Setting.* We call a kernel a map $K : \mathbb{T}^2 \to \mathbb{R}$ for which both $\int_0^t K(t,s)^2 \mathrm{d}s$ and $K(t,s)$ are finite for all $t \in \mathbb{T}$ and $s \neq t$. The associated space is defined as

$$\mathcal{K} := \left\{u : \mathbb{T} \to \mathbb{R}, \{\mathcal{F}_t\}\text{-progressively measurable, such that } \mathbb{E}\int_0^t \left[K(t,s)u(s)\right]^2 \mathrm{d}s < \infty, \text{ for all } t \in \mathbb{T}\right\}.$$

Hence, for all $u \in \mathcal{A}_K$ the stochastic integral

$$\widetilde{M}_t^K(u) := \int_0^t K(t,s)u(s)\mathrm{d}W_s$$

is well defined for all $t \in \mathbb{T}$ in the Itô sense. For any $\alpha \in (0,1)$, we denote the Riemann-Liouville integral I^α and derivative D^α, as

$$(\mathrm{I}^\alpha f)(t) := \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1} f(s)\mathrm{d}s, \qquad (\mathrm{D}^\alpha f)(t) := \frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{I}^{1-\alpha} f)(t), \quad \text{for } f \in L^1, t \in \mathbb{T}. \tag{3.20}$$

Define $\mathcal{I}_{\alpha,p} := \mathrm{I}^\alpha(L^p)$ equipped with the norm $\|f\|_{\mathcal{I}_{\alpha,p}} := \|D^\alpha f\|_{L^p}$ if $f \in \mathrm{I}^\alpha(L^p)$ and infinity otherwise. If $\alpha > \frac{1}{p}$, then $\mathcal{I}_{\alpha,p} \subset \mathcal{C}_0^{\alpha-\frac{1}{p}}$, the space of $(\alpha - \frac{1}{p})$-Hölder continuous functions null at time 0. Let K denote the linear map associated to $K(t,s)$ by

$$\mathrm{K}f(t) := \int_0^t K(t,s)f(s)\mathrm{d}s, \tag{3.21}$$

and introduce, for $x \in (0,\infty) \setminus \{2\}$,

$$\theta(x) := \frac{2x}{2-x}. \tag{3.22}$$

Given the space inclusions above, the following assumption implies precise Hölder regularity for the integral (3.21):

Assumption 3.21. There exist $\eta \in (1,2)$ and $\alpha > 1/\theta(\eta)$ for which K is continuous from L^2 to $\mathcal{I}_{\alpha+\frac{1}{2},2}$ and from L^η to $\mathcal{I}_{\alpha,\theta(\eta)}$.

Example 3.22. *The operators associated to the following kernels satisfy Assumption 3.21:*

- *The Riemann-Liouville kernel*

$$J_H(t,s) = \frac{(t-s)_+^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}, \qquad \text{with } H \in (0,1),$$

  satisfies this assumption with $\alpha = H$ and any $\eta < 2$ [32, Theorem 4.1].
- *The fractional Brownian motion kernel*

$$K_H(t,s) = \frac{(t-s)_+^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} F\left(H-\frac{1}{2}, \frac{1}{2}-H, H+\frac{1}{2}, 1-\frac{t}{s}\right),$$

  where F is the Gauss Hypergeometric function, also satisfies this assumption with the same parameters as above [32, Theorem 4.2].

Decreusefond's main result yields the Hölder regularity of the stochastic Volterra integral [32, Theorem 3.1]:

Theorem 3.23. *Under Assumption 3.21, let $u \in \mathcal{K} \cap L^{\theta(\eta)}(\Omega \times \mathbb{T})$. Then $\widetilde{M}^K(u)$ has a measurable version $M^K(u)$ which is γ-Hölder continuous for all $\gamma < \alpha - 1/\theta(\eta)$.*

From now on, we only consider the measurable version of the stochastic integral. Although this theorem was proved in a one-dimensional setting, it also covers multi-dimensional stochastic Volterra integrals by considering its components individually and summing them.

3.5.2. *Large and moderate deviations.* For all $\varepsilon > 0$ consider the stochastic Volterra equation

$$X_t^\varepsilon = X_0^\varepsilon + \int_0^t K(t,s) b_\varepsilon(s, X_s^\varepsilon)\mathrm{d}s + \vartheta_\varepsilon \int_0^t K(t,s)\sigma_\varepsilon(s, X_s^\varepsilon)\mathrm{d}W_s, \tag{3.23}$$

which was studied in [30] without the ε-dependence. To complete the non-convolution setup we also need the following condition.

Assumption 3.24. There exists $q > 2$ such that

$$\sup_{t\in\mathbb{T}} \left\{ \int_0^t |K(t,s)|^{-\theta(q)}\, \mathrm{d}s \right\} < \infty, \tag{3.24}$$

with $\theta(q)$ introduced in (3.22).

Let $p > q > 2$, then $2 < -\theta(p) < -\theta(q)$ and Hölder's and Jensen's inequalities yield

$$\begin{aligned}\left[\int_0^t |K(t,s)f(s)|^2\, \mathrm{d}s\right]^{\frac{p}{2}} &\le \left[\int_0^t |K(t,s)|^{-\theta(p)}\, \mathrm{d}s\right]^{\frac{p-2}{2}} \int_0^t |f(s)|^p\, \mathrm{d}s \\ &\le t^{\frac{p-q}{q}} \left[\int_0^t |K(t,s)|^{-\theta(q)}\, \mathrm{d}s\right]^{\frac{p}{-\theta(q)}} \int_0^t |f(s)|^p\, \mathrm{d}s.\end{aligned}$$

This replaces the Gronwall-type inequality derived for convolution kernels in Lemma A.1. Hence replacing Assumption 2.3 by the condition (3.24) one recovers the moments bounds of Lemmata 3.10 and 3.17 for the process $X^{\varepsilon,v}$ and for any $p \ge 1$. Setting in particular $p = \theta(\eta)$ from Theorem 3.23

and for $\xi \in \{b_\varepsilon, \sigma_\varepsilon\}$, we have $\xi(X^{\varepsilon,v}) \in \mathcal{K} \cap L^p(\Omega \times \mathbb{T})$. Therefore Assumption 3.21, Theorem 3.23 and Assumption 3.24 yield the almost sure Hölder regularity of the following processes defined on $\mathbb{T}$:

$$\int_0^\cdot K(\cdot, s)\xi(X_s^{\varepsilon,v})\mathrm{d}s \qquad \text{and} \qquad \int_0^\cdot K(\cdot, s)\xi(X_s^{\varepsilon,v})\,\mathrm{d}W_s.$$

Hence we recover the tightness of Lemmata 3.12 and 3.18 under this new set of assumptions. Notice that we can consider kernels consisting of both convolution and non-convolution components. Finally we extend the LDP and MDP results without further modifications:

Theorem 3.25. *The conclusions of Theorems 3.8 and 3.16 stand for $\{X^\varepsilon\}_{\varepsilon>0}$ defined by (3.23) when each component of the kernel K satisfies either Assumption 2.3 or Assumptions 3.21 and 3.24.*

4. Application to rough volatility

We now show how our results (Theorems 3.8, 3.16 and 3.25) apply to a large class of models recently developed in mathematical finance. Originally proposed by Comte and Renault [26] with financial econometrics applications in mind, rough volatility models were rediscovered later in the context of option pricing in [2, 6, 47, 51], developed and extended widely, and have now become the new standards of volatility modelling. They usually take the following form:

$$\begin{cases} X_t &= -\dfrac{1}{2}\displaystyle\int_0^t \Sigma(Y_s)\mathrm{d}s + \int_0^t \sqrt{\Sigma(Y_s)}\mathrm{d}B_s, \\ Y_t &= y_0 + \displaystyle\int_0^t K_1(t-s)\mathfrak{b}(s, Y_s)\mathrm{d}s + \int_0^t K_2(t-s)\zeta(Y_s)\mathrm{d}W_s, \end{cases} \tag{4.1}$$

where $K_1, K_2 \in L^2$ and B and W are two standard Brownian motions with $\mathrm{d}\langle B, W\rangle_t = \rho \mathrm{d}t$, for some correlation parameter $\rho \in (-1,1)$. We further define $\overline{\rho} := \sqrt{1-\rho^2}$, and set $X_0 = 0$ without loss of generality. Here X denotes the logarithm of a stock price process, and Y its instantaneous volatility. We adopt a slight abuse of notation, as X previously denoted the multidimensional system, but writing now X as the log-stock price is consistent with the mathematical finance literature and should not create any confusion. We summarise in Table 1 the most common rough volatility models used in mathematical finance, indicating where their asymptotic behaviours were covered, and where our framework not only encompasses those, but fills the gaps so far missing[1]. The detailed analysis of these cases is then provided in Section 4.2 in the small-time case, and in Section 4.3 for their tail behaviours.

4.1. **Small-time rescaling (general).**

In the small-time case, we need to assume some scaling behaviour for the kernel functions. We say that a function $f : \mathbb{R} \to \mathbb{R}^{d\times d}$ is homogeneous of degree $\alpha \in \mathbb{R}$ if $f(\lambda x) = \lambda^\alpha f(x)$ holds for all $x, \lambda \in \mathbb{R}$.

Assumption 4.1. K_1 and K_2 are homogeneous of degrees $\varpi \in (-\frac{1}{2}, \frac{1}{2}]$ and $H - \frac{1}{2} \in (-\frac{1}{2}, \frac{1}{2}]$.

Since K_1 is homogeneous of degree ϖ, then

$$\begin{aligned} \int_0^h K_1(t)^2\mathrm{d}t &= \int_0^h K_1(1)^2 t^{2\varpi}\mathrm{d}t \leq \frac{K_1(1)}{1+2\varpi}h^{1+2\varpi}, && \text{for any } h > 0, \\ \int_0^T \big(K_1(t+h) - K_1(t)\big)^2\mathrm{d}t &= K_1(1)^2 \int_0^T \big((t+h)^\varpi - t^\varpi\big)^2\mathrm{d}t = \mathcal{O}\left(h^{1+2\varpi}\right), && \text{for } h \text{ small enough,} \end{aligned}$$

[1] As discussed below, our application to the rough Heston model is conditional on the latter to have a unique pathwise solution, a problem that remains open so far.

Models	Rough Stein-Stein				Rough Bergomi		Multifactor rough Bergomi		Rough Heston		
	Small-time		Tail		Small-time		Small-time		Small-time [42]		Tail
	LDP [56]	MDP	LDP [56]	MDP	LDP [58]	MDP	LDP [61]	MDP	LDP	MDP	LDP
$(X^\varepsilon, Y^\varepsilon)$	IF	IF	IF	-	IF	IF	IF	IF	IF	IF	IF
X^ε	OP	IF	OP	-	OP	IF	OP	OP	OP	IF	OP
X_ε	OP	CF	OP	-	OP	CF	OP	OP	OP	CF	OP
$\widehat{\sigma}$	OP	CF	OP	-	OP	CF	OP	OP	OP	CF	OP
Y^ε	IF	IF	IF	IF	IF	IF	IF	IF	IF	IF	IF
Y_ε	CF	CF	IF	IF	CF	CF	CF	CF	IF	CF	IF

TABLE 1. Summary of rough volatility results and form of the rate functions (CF=closed-form; IF=integral form; OP=optimisation problem; shadowed cells are new contributions from this paper). $\widehat{\sigma}$ corresponds to the implied volatility, defined precisely after Remark 4.4.

and so Assumption 2.3 is satisfied with $\gamma = 1+2\varpi \in (0,2]$, and likewise for K_2 with $\gamma = 2H$. Under this assumption, the rescalings $X^\varepsilon_t := \varepsilon^{H-\frac{1}{2}} X_{\varepsilon t}$ and $Y^\varepsilon_t := Y_{\varepsilon t}$ turn (4.1) into

$$\left\{\begin{array}{ll} X^\varepsilon_t & = -\dfrac{\varepsilon^{H+\frac{1}{2}}}{2}\displaystyle\int_0^t \Sigma(Y^\varepsilon_s)\mathrm{d}s + \varepsilon^H \int_0^t \sqrt{\Sigma(Y^\varepsilon_s)}\mathrm{d}B_s, \\ Y^\varepsilon_t & = y_0 + \varepsilon^{1+\varpi}\displaystyle\int_0^t K_1(t-s)\mathfrak{b}(\varepsilon s, Y^\varepsilon_s)\mathrm{d}s + \varepsilon^H \int_0^t K_2(t-s)\zeta(Y^\varepsilon_s)\mathrm{d}W_s, \end{array}\right. \tag{4.2}$$

so that we are precisely in the framework of (2.1) with $d=3$, $\vartheta_\varepsilon = \varepsilon^H$,

$$K(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & K_1(t) & K_2(t) \\ 0 & 0 & 0 \end{pmatrix}, \quad b_\varepsilon(t,(x,y)) = \begin{pmatrix} -\frac{1}{2}\varepsilon^{H+\frac{1}{2}}\Sigma(y) \\ \varepsilon^{1+\varpi}\mathfrak{b}(t,y) \\ 0 \end{pmatrix}, \quad \sigma_\varepsilon(t,(x,y)) = \begin{pmatrix} \rho\sqrt{\Sigma(y)} & \overline{\rho}\sqrt{\Sigma(y)} & 0 \\ 0 & 0 & 0 \\ \zeta(y) & 0 & 0 \end{pmatrix},$$

where the third component is meaningless but allows us to handle the two different kernels. Note that σ_ε does not depend on ε but integrates the correlation. The controlled equation (3.4) for the second component reads

$$Y^{\varepsilon,v}_t = y_0 + \varepsilon^{1+\varpi}\int_0^t K_1(t-s)\mathfrak{b}(s, Y^{\varepsilon,v}_s)\mathrm{d}s + \varepsilon^H \int_0^t K_2(t-s)\zeta(Y^{\varepsilon,v}_s)\mathrm{d}W_s + \int_0^t K_2(t-s)\zeta(Y^{\varepsilon,v}_s)v_s\mathrm{d}s,$$

for each $t \in \mathbb{T}, \varepsilon > 0$ and $v \in \mathcal{A}$. Note that the dynamics of X^ε do not feed back into Y^ε and that $\Sigma \in \mathcal{S}^{|\Sigma|}_{\{2\}}$ in the sense of Definition 2.5. The following assumption stands throughout this section:

Assumption 4.2 (Small-time assumptions)**.**

- Σ, ζ and $\mathfrak{b}(t,\cdot)$ are continuous on $\mathbb{R}$, uniformly in $t \in \mathbb{T}$;
- $\mathfrak{b}(t,\cdot)$ and ζ are of linear growth, uniformly in $t \in \mathbb{T}$;
- Σ is either of linear growth or such that for all $p \geq 1$, $N > 0$ and $\varepsilon > 0$ small enough,

$$\sup_{t\in\mathbb{T}, v\in\mathcal{A}_N} \mathbb{E}\Big[\big|\Sigma\big(Y^{\varepsilon,v}_t\big)\big|^p\Big] < \infty; \tag{4.3}$$

- the equation for Y^ε in (4.2) is exact for small enough $\varepsilon > 0$.

These conditions ensure that **H2** holds with limit coefficients $b \equiv (0,0,0)^\top$ and $\sigma = \sigma_\varepsilon$. Furthermore, Y^ε is an autonomous subsystem in the sense of Definition 2.6 and **H3** and the bound (3.6) holds. An exact solution of the system (4.1) exists since X^ε is explicit from Y^ε, and **H4** is satisfied.

4.1.1. *Large deviations.* For more explicit results, we now fix $K_2(t) := \frac{1}{\Gamma(H+\frac{1}{2})}t^{H-\frac{1}{2}}$, a common setup in rough volatility models. For each control $v \in \mathcal{S}_N$ with $N > 0$, the limit equation (3.5) of the volatility in the large deviations regime reads

$$\varphi_t = y_0 + \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}\zeta(\varphi_s)v_s\mathrm{d}s, \qquad \text{for } t \in \mathbb{T}. \tag{4.4}$$

From Cauchy-Schwarz inequality, the linear growth condition of ζ and the continuity of Σ, we obtain that $|\Sigma(\varphi_t)|$ is uniformly bounded in $t \in \mathbb{T}$ and in $v \in \mathcal{S}_N$, hence (3.7) holds. Therefore Assumption 3.6 and **H1 - H4** follow from Assumption 4.2. Mimicking the fractional integral notation from Section 3.5, we introduce for convenience the notations

$$\mathrm{I}_x^{H+\frac{1}{2}}(f) := x + \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} f_s\mathrm{d}s \qquad \text{and} \qquad \mathrm{I}_x^{H+\frac{1}{2}}(L^1) := \left\{\mathrm{I}_x^{H+\frac{1}{2}}(f), f \in L^1\right\},$$

and the fractional derivative D is defined in (3.20). From now on, to simplify the statements, we write $Z^\varepsilon \sim \mathrm{LDP}(I, \varepsilon^{-1})$ to express that the family of random variables $\{Z^\varepsilon\}_{\varepsilon>0}$ satisfies an LDP with rate function I and speed ε^{-1}, as ε tends to zero.

Proposition 4.3 (Large deviations)**.** *Under Assumptions 3.1 for (4.4), 4.1 and 4.2, the following hold:*

(L1) $(X^\varepsilon, Y^\varepsilon) \sim \mathrm{LDP}\left(I, \varepsilon^{-H}\right)$, *where* $I : \mathcal{W}^2 \to \mathbb{R}_+$ *is given by*

$$I(\phi,\varphi) = \frac{1}{2\overline{\rho}}\int_0^T \left(\frac{\dot{\phi}_t^2}{\Sigma(\varphi_t)}\mathbb{1}_{\Sigma(\varphi_t)\neq 0} - \frac{2\rho\dot{\phi}_t \mathrm{D}^{H+\frac{1}{2}}(\varphi-y_0)_t}{\zeta(\varphi_t)\sqrt{\Sigma(\varphi_t)}}\mathbb{1}_{\zeta(\varphi_t)\Sigma(\varphi_t)\neq 0} + \left|\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi-y_0)_t}{\zeta(\varphi_t)}\right|^2 \mathbb{1}_{\zeta(\varphi_t)\neq 0}\right)\mathrm{d}t,$$

if $\phi \in \mathcal{AC}_0$ *and* $\varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$*, and infinity otherwise.*

(L2) $X^\varepsilon \sim \mathrm{LDP}\left(I^X, \varepsilon^{-H}\right)$*, where*

$$I^X(\phi) = \inf\left\{I(\phi,\varphi) : \varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)\right\},$$

if $\phi \in \mathcal{AC}_0$ *and infinity otherwise;*

(L3) $\varepsilon^{H-\frac{1}{2}}X_\varepsilon \sim \mathrm{LDP}\left(I_1^X, \varepsilon^{-H}\right)$ *where* $I_1^X(x) = \inf\left\{I^X(\phi) : \phi_1 = x\right\}$ *for all* $x \in \mathbb{R}$*;*

(L4) $Y^\varepsilon \sim \mathrm{LDP}\left(I^Y, \varepsilon^{-H}\right)$*, where*

$$I^Y(\varphi) = \frac{1}{2}\int_0^T \left|\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi-y_0)(t)}{\zeta(\varphi_t)}\right|^2 \mathbb{1}_{\zeta(\varphi_t)\neq 0}\mathrm{d}t$$

if $\varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$ *and infinity otherwise;*

(L5) $Y_\varepsilon \sim \mathrm{LDP}\left(I_1^Y, \varepsilon^{-H}\right)$*, with* $I_1^Y(y) = \inf\left\{I^Y(\varphi) : \varphi_1 = y\right\}$ *for* $y \in \mathbb{R}$.

While **(L1)**, **(L2)** and **(L4)** deal with pathwise large deviations, **(L3)** and **(L5)** are one-dimensional large deviations statements, about the marginal distributions of X and Y. In this small-time behaviour case, we recover the same scaling as in [42, 43].

Proof of Proposition 4.3.

(L1) As discussed above, the assumptions of Theorem 3.8 are satisfied, so that the three-dimensional process $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon)$, where $Z^\varepsilon \equiv 0$ for all $\varepsilon > 0$, satisfies an LDP with rate function

$$J(\phi,\varphi,\psi) = \inf\left\{\frac{1}{2}\int_0^T \left(u_t^2 + v_t^2 + w_t^2\right)\mathrm{d}t : \quad u, v, w \in L^2, \right.$$
$$\left. \phi_t = \int_0^t \sqrt{\Sigma(\varphi_s)}\left[\overline{\rho}u_s + \rho v_s\right]\mathrm{d}s, \quad \varphi = \mathrm{I}_{y_0}^{H+\frac{1}{2}}\left(\zeta(\varphi)v\right), \quad \psi \equiv 0\right\}.$$

The map $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon) \mapsto (X^\varepsilon, Y^\varepsilon)$ is continuous so that the contraction principle [33, Theorem 4.2.1] yields an LDP for $(X^\varepsilon, Y^\varepsilon)$ with rate function $\inf\left\{J(\phi,\varphi,\psi) : \psi \equiv 0\right\} = J(\phi,\varphi,0)$, which corresponds to

$$I(\phi,\varphi) = \inf\left\{\frac{1}{2}\int_0^T \left(u_t^2+v_t^2\right)\mathrm{d}t : u,v\in L^2, \phi_t = \int_0^t \sqrt{\Sigma(\varphi_s)}\left[\overline{\rho}u_s+\rho v_s\right]\mathrm{d}s,\ \varphi = \mathrm{I}_{y_0}^{H+\frac{1}{2}}(\zeta(\varphi)v)\right\}.$$

We can identify the unique controls for each $\phi \in \mathcal{AC}_0$, $\varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$ by reverting the integrals:

$$u_t = \frac{1}{\overline{\rho}}\left(\frac{\dot{\phi}_t}{\sqrt{\Sigma(\varphi_t)}} - \rho v_t\right)\mathbb{1}_{\Sigma(\varphi_t)\neq 0}, \quad v_t = \frac{1}{\zeta(\varphi_t)}\mathrm{D}^{H+\frac{1}{2}}(\varphi-y_0)_t\mathbb{1}_{\zeta(\varphi_t)\neq 0}, \quad \text{for all } t\in\mathbb{T},$$

because whenever $\zeta(\varphi_t)=0$, although v is not uniquely determined by φ, the optimal choice of control (the one minimising the cost) is $v_t = 0$, and likewise for u when $\Sigma(\varphi_t)=0$, see [27, Remark 2.3] for more details. Plugging these into the rate function yields the claim. Note that if $\phi \notin \mathcal{AC}_0$ or $\varphi \notin \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$ then they cannot satisfy the equations and therefore the infimum takes place over an empty set.

(L2) Since the map $(X^\varepsilon, Y^\varepsilon) \mapsto X^\varepsilon$ is continuous, the claim follows from the contraction principle.

(L3) Projecting the pathwise large deviations **(L2)** onto the last coordinate point $t=1$ is equivalent to applying the contraction principle, and the claim follows immediately.

(L4) A direct application of Theorem 3.8 yields an LDP with rate function

$$I^Y(\varphi) = \inf\left\{\frac{1}{2}\int_0^T v_t^2\mathrm{d}t : v\in L^2, \varphi = \mathrm{I}_{y_0}^{H+\frac{1}{2}}(\zeta(\varphi)v)\right\};$$

Inverting it as above ends the proof and **(L5)** follows from the contraction principle.

□

Remark 4.4. *The form of I above is reminiscent of the Freidlin-Wentzell rate function after application of the contraction principle from the Brownian motion LDP. The weak convergence approach allows to relax the assumptions of continuity (essentially Lipschitz continuity of the coefficients) to simple well-posedness of the equation.*

4.1.2. *Moderate deviations.* We now show how our moderate deviations results apply to the rough volatility model (4.1). Let $h_\varepsilon = \varepsilon^{-\beta}$ for any $\beta\in(0,H)$, and define the two-dimensional process

$$\eta^\varepsilon := \frac{1}{\vartheta_\varepsilon h_\varepsilon}\left(X^\varepsilon, Y^\varepsilon - y_0\right) = \frac{1}{\varepsilon^{H-\beta}}\left(X^\varepsilon, Y^\varepsilon-y_0\right).$$

The case $\beta=0$ corresponds to the Central Limit Theorem, whereas $\beta=H$ is the LDP regime, so that MDP precisely corresponds to some interpolation between the two. Regarding the assumptions note that $y_0^\varepsilon = y_0$ and b_ε clearly tends to zero as ε goes to zero, hence it is trivial that $b\equiv 0$ is continuously differentiable and Lipschitz continuous, thus **H5** and **H7** hold.

Now let Assumption 4.2 hold. One notices that $\left|b_\varepsilon^{(1)}(t,(x,y)) - b^{(1)}(t,(x,y))\right| \le \varepsilon^{H+\frac{1}{2}}\left|\Sigma(y)\right|$ and $\left|b_\varepsilon^{(2)}(t,(x,y)) - b^{(2)}(t,(x,y))\right| \le \varepsilon^{1+\varpi}C_L(1+|y|)$ by linear growth. For **H8** to hold, one then requires that $\varepsilon^{1+\varpi-(H-\beta)}$ and $\varepsilon^{H+\frac{1}{2}-(H-\beta)}$ both tend to zero as ε goes to zero. Moreover the bound (4.3) corresponds to (3.17).

Assumption 4.5 (Moderate deviations assumptions)**.**

- The parameters H, ϖ and β are such that $(1+\varpi)\wedge(H+\frac{1}{2}) > H-\beta$;
- There exists $\delta>0$ such that Σ and ζ are locally δ-Hölder continuous.

Notice that the first inequality is always satisfied if $H\le\frac{1}{2}$. Therefore, Assumptions 4.1, 4.2, 4.5 imply **H1 - H8** and Assumptions 2.3 and 3.6. Similarly to the LDP case, and recalling the definition of MDP from the introduction, we write $Z^\varepsilon \sim \mathrm{MDP}(\Lambda, l_\varepsilon)$ if in fact $\varepsilon^{\beta-H}(Z^\varepsilon - \overline{Z}) \sim \mathrm{LDP}(\Lambda, l_\varepsilon)$, for any $l_\varepsilon > 0$ converging to zero as ε tends to zero, where $\overline{Z}$ is the limit in distribution of Z^ε. We also denote the subset of $\mathcal{W}^d$ of absolutely continuous functions by $\mathcal{AC}$, and $\mathcal{AC}_0 := \{\phi \in \mathcal{AC}, \phi_0 = 0\}$, recall that $K_2(t) = t^{H-\frac{1}{2}}/\Gamma(H+\frac{1}{2})$ and refer to (3.20) for the definition of the Riemann-Liouville fractional derivative.

Proposition 4.6. *Under Assumptions 4.1, 4.2 and 4.5, the following moderate deviations hold:*

(M1) $(X^\varepsilon, Y^\varepsilon) \sim \mathrm{MDP}\left(\Lambda, \varepsilon^{-\beta}\right)$, *where* $\Lambda : \mathcal{W}^2 \to \mathbb{R}_+$ *is given by*

$$\Lambda(\phi,\varphi) = \frac{1}{2\overline{\rho}}\int_0^T \left(\frac{\dot{\phi}_t^2}{\Sigma(y_0)}\mathbb{1}_{\Sigma(y_0)\neq 0} - \frac{2\rho\dot{\phi}_t \mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)_t}{\zeta(y_0)\sqrt{\Sigma(y_0)}}\mathbb{1}_{\zeta(y_0)\Sigma(y_0)\neq 0} + \left|\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi-y_0)_t}{\zeta(y_0)}\right|^2 \mathbb{1}_{\zeta(y_0)\neq 0}\right)\mathrm{d}t,$$

if $\phi \in \mathcal{AC}_0$ *and* $\varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$, *and infinity otherwise;*

(M2) $X^\varepsilon \sim \mathrm{MDP}\left(\Lambda^X, \varepsilon^{-\beta}\right)$, *where* $\Lambda^X(\phi) = \dfrac{1}{2\Sigma(y_0)}\displaystyle\int_0^T \dot{\phi}_t^2 \mathrm{d}t$, *if* $\phi \in \mathcal{AC}_0$ *and infinity otherwise;*

(M3) $\varepsilon^{H-\frac{1}{2}} X_\varepsilon \sim \mathrm{MDP}\left(\Lambda_1^X, \varepsilon^{-\beta}\right)$, *with* $\Lambda_1^X(x) = \dfrac{x^2}{2\Sigma(y_0)}$, *for* $x \in \mathbb{R}$*;*

(M4) $Y^\varepsilon \sim \mathrm{MDP}\left(\Lambda^Y, \varepsilon^{-\beta}\right)$, *where* $\Lambda^Y : \mathcal{W} \to \mathbb{R}_+$ *is given by*

$$\Lambda^Y(\varphi) = \frac{1}{2}\int_0^T \left|\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t)}{\zeta(y_0)}\right|^2 \mathbb{1}_{\zeta(y_0)\neq 0}\mathrm{d}t, \qquad \textit{if } \varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1) \textit{ and infinity otherwise};$$

(M5) $Y_\varepsilon \sim \mathrm{MDP}(\Lambda_1^Y, \varepsilon^{-\beta})$, *where* $\Lambda_1^Y(y) = \frac{1}{2}y^2$ *for* $y \in \mathbb{R}$.

Proof.

(M1) As discussed above, the assumptions of Theorem 3.16 are satisfied, thus it yields an MDP with rate function

$$\Lambda(\phi,\varphi) = \left\{\frac{1}{2}\int_0^T \left(u_t^2 + v_t^2\right)\mathrm{d}t : u, v \in L^2, \phi_t = \int_0^t \sqrt{\Sigma(y_0)}\left[\overline{\rho}u_s + \rho v_s\right]\mathrm{d}s, \varphi = \mathrm{I}_{y_0}^{H+\frac{1}{2}}(\zeta(y_0)v)\right\},$$

and inverting it as in the proof of Proposition 4.3 gives the claim.

(M2) The contraction principle implies a MDP for X^ε holds with rate function $\Lambda^X(\phi) = \inf\left\{\Lambda(\phi,\varphi) : \varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)\right\}$. If $\zeta(y_0) = 0$, it is easy to check that $\Lambda^X(\phi) = \Lambda(\phi, 0)$. Otherwise, let $\psi \in \mathcal{AC}_0$ such that $\dot{\psi} := \frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)}{\zeta(y_0)}\mathbb{1}_{\zeta(y_0)\neq 0}$, then the rate function translates to

$$\Lambda^X(\phi) = \inf\left\{\frac{1}{2\overline{\rho}^2}\int_0^T \left(\frac{\dot{\phi}_t^2}{\Sigma(y_0)}\mathbb{1}_{\Sigma(y_0)\neq 0} - \frac{2\rho\dot{\phi}_t\dot{\psi}_t}{\sqrt{\Sigma(y_0)}}\mathbb{1}_{\Sigma(y_0)\neq 0} + \dot{\psi}_t^2\right)\mathrm{d}t : \psi \in \mathcal{AC}\right\},$$

which can be solved as a variational problem as in [59, Corollary 2.4]. The corresponding Euler-Lagrange equation reads $\ddot{\psi} = \rho\ddot{\phi}/\sqrt{\Sigma(y_0)}$ hence $\dot{\psi} = \rho\dot{\phi}/\sqrt{\Sigma(y_0)}$ because $\dot{\psi}_0 = 0$ by definition. Plugging into the above equation finishes the proof.

(M3) The rate function is given by contraction principle as

$$\Lambda_1^X(x) = \inf\left\{\Lambda^X(\phi) : \phi \in \mathcal{AC}_0, \phi_1 = x\right\} = \inf\left\{\frac{1}{2\Sigma(y_0)}\int_0^T \dot{\phi}_t^2\mathrm{d}t : \phi \in \mathcal{AC}_0, \phi_1 = x\right\}.$$

Setting $T = 1$ the optimal path under the constraint $\phi_1 = x$ is $\phi_t = xt$ by the Euler-Lagrange equation. Again, plugging it into the rate function ends the proof.

(M4) Theorem 3.16 gives an MDP for Y^ε with rate function

$$\Lambda^Y(\varphi) = \inf\left\{\frac{1}{2}\int_0^T v_t^2\mathrm{d}t : v \in L^2, \varphi = \mathrm{I}_{y_0}^{H+\frac{1}{2}}(\zeta(y_0)v)\right\}.$$

Inverting it yields **(M4)**.

(M5) By contraction principle: $\Lambda^Y_1(y) := \inf\left\{\Lambda^Y(\varphi), \varphi_1 = y\right\}$. Then setting ψ as in **(M2)** it boils down to the same optimisation problem as for **(M3)**.

□

As in the large deviations results, **(M1)**, **(M2)** and **(M4)** correspond to pathwise statements, whereas **(M3)** and **(M5)** are finite-dimensional results about the marginal distributions. For the log-stock price, **(M3)** corresponds precisely to the Moderately Out-of-The-Money regime presented and justified in [46] (for diffusion volatility models), based on the observation that the range of observable strikes grows with maturity.

4.1.3. *Implied volatility asymptotics.* We can easily deduce from the above results the asymptotic behaviour of the implied volatility, a standard norm for quoting option prices. For each maturity $t \geq 0$ and log-moneyness $k \in \mathbb{R}$, the implied volatility $\widehat{\sigma}(t,k)$ is the unique non-negative solution to $C_{\mathrm{BS}}\big(t,k,\widehat{\sigma}(t,k)\big) = C(t,k)$, where C_{BS} corresponds to the price of a European Call option under the Black-Scholes model, and C a given Call option price (for example in a rough volatility model). This notion is only well defined if the underlying stock price is a true martingale, which we have not assumed so far, and may require additional conditions on the coefficients. This will be the case though in all our examples below, but for now, with the current level of generality, we assume it:

Assumption 4.7. The process $\exp(X^\varepsilon)$ in (4.2) is a true martingale for all small enough $\varepsilon > 0$.

Small-time implied volatility asymptotics can be derived from Properties 4.3 and 4.6 in a similar fashion. The explicit form of the MDP rate function allows a closed form expression.

Corollary 4.8. *Let Assumption 4.7 hold.*

(LDP) *Under the same assumptions as Proposition 4.3,*

$$\lim_{t\downarrow 0}\widehat{\sigma}\big(t, kt^{\frac{1}{2}-H}\big)^2 = \begin{cases} \dfrac{k^2}{2\inf_{x\geq k} I^X_1(x)}, & \text{if } k>0,\\[2mm] \dfrac{k^2}{2\inf_{x\leq k} I^X_1(x)}, & \text{if } k<0. \end{cases}$$

(MDP) *Under the same assumptions as Proposition 4.6 and for any $\beta \in (0,H)$, $k \neq 0$,*

$$\lim_{t\downarrow 0}\widehat{\sigma}\big(t, kt^{\frac{1}{2}-\beta}\big)^2 = \Sigma(y_0).$$

Proof.

(LDP) Consider the case $k>0$. Proposition 4.3**(L3)** translates into

$$\lim_{t\downarrow 0} t^{2H}\log\mathbb{P}(t^{H-\frac{1}{2}}X_t \geq k) = -\inf_{x\geq k} I^X_1(x).$$

Meanwhile in the Black-Scholes model with constant volatility $\sigma>0$ the log-price process satisfies $X_t = -\frac{\sigma^2 t}{2} + \sigma B_t$ for all $t\in\mathbb{T}$, and simple Gaussian computations yield the large deviations behaviour

$$\lim_{t\downarrow 0} t^{2H}\log\mathbb{P}(X_t \geq kt^{\frac{1}{2}-H}) = -\frac{k^2}{2\sigma^2}.$$

The claim then follows directly from [49, Corollary 7.1], and by symmetry for the case $k<0$.

(MDP) Following the same arguments as above, we obtain

$$\lim_{t\downarrow 0}\widehat{\sigma}\big(t, kt^{\frac{1}{2}-\beta}\big)^2 = \begin{cases} \dfrac{k^2}{2\inf_{x\geq k} \Lambda^X_1(x)}, & \text{if } k>0,\\[2mm] \dfrac{k^2}{2\inf_{x\leq k} \Lambda^X_1(x)}, & \text{if } k<0. \end{cases}$$

Plugging in the expression of Λ^X_1 from **(M3)** finishes the proof. □

This concludes the presentation of the general results for rough volatility models. The next sections display the diversity of the models found in the literature and how large and moderate deviations principle apply to them.

4.2. **Small-time rescaling (examples).**

4.2.1. *Rough Stein-Stein.* The rough Stein-Stein, suggested in [56] is an extension of the classical Stein-Stein volatility model [73] to the fractional setting. It corresponds to (4.1) with $K_1 \equiv 1$ (hence $\varpi = 0$), $K_2(t) = \frac{1}{\Gamma(H+\frac{1}{2})} t^{H-\frac{1}{2}}$, $H \in (0, \frac{1}{2})$, $\Sigma(y) = y^2$, $\mathfrak{b}(y) = \kappa(\theta - y)$, $\kappa, \theta > 0$ and $\zeta(y) \equiv \xi > 0$. The coefficients are Lipschitz continuous and well-behaved, hence Assumptions 4.2 and 4.5 are easily checkable and the limit equation (4.4) has a unique solution, hence Propositions 4.3 and 4.6 apply. Note that because ζ is constant, one can solve **(L5)** in closed-form using the Euler-Lagrange equation in a similar way as in the proof of Proposition 4.6. Furthermore, since Y is Gaussian, therefore its exponential moments are finite and Novikov's condition [60, Section 3.5.D] ensures that Assumption 4.7 holds. Therefore, Corollary 4.8 yields the small-time behaviour of the implied volatility. Notice that the LDP and MDP for this model still hold when replacing the Riemann-Liouville kernel with the standard fractional Brownian motion by virtue of Theorem 3.25. The pathwise LDP for this model was first derived in [56] albeit with the different scaling $X^\varepsilon_t := \varepsilon^{H-\frac{1}{2}+2\beta} X_{\varepsilon t}$ and $Y^\varepsilon_t := \varepsilon^\beta Y_{\varepsilon t}$, for $\beta > 0$.

4.2.2. *Rough Bergomi.* The rough Bergomi model as presented in [6, 58] reads

$$\begin{cases} X_t &= -\dfrac{1}{2} \displaystyle\int_0^t V_s \mathrm{d}s + \int_0^t \sqrt{V_s} \mathrm{d}B_s, \\ V_t &= V_0 \exp\left(\displaystyle\int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \mathrm{d}W_s - a t^{2H} \right), \end{cases}$$

with $V_0 > 0$ and $a \in \mathbb{R}$. Proving an LDP is quite intricate because the exponential does not satisfy the linear growth bound but we circumvented this issue by introducing the notion of autonomous system, illustrated in Example 2.7 and completed by Assumption 3.6 and **H3b**. Hence, with $Y := \log(V)$, the system (X, Y) fits into (4.1) where $K_1(t) = t^{2H-1}$, $K_2(t) = \frac{1}{\Gamma(H+\frac{1}{2})} t^{H-\frac{1}{2}}$, for $H \in (0, \frac{1}{2})$, $\Sigma(y) = \exp(y)$, $\mathfrak{b}(y) = -a/(2H)$, $a > 0$, $\zeta(y) \equiv 1$ and $y_0 := \log(V_0)$. The bound (4.3) is then satisfied since $\int_0^t (t-s)^{H-\frac{1}{2}} \mathrm{d}W_s$ is a Gaussian process hence its exponentional moments are bounded uniformly in $t \in \mathbb{T}$, and for each $N > 0$, and $v \in \mathcal{A}_N$:

$$\int_0^t (t-s)^{H-\frac{1}{2}} v_s \mathrm{d}s \le N \frac{t^{2H}}{2H},$$

almost surely, by Cauchy-Schwarz inequality. Therefore $\sup_{t\in\mathbb{T}} \mathbb{E}\big[\exp\big(Y^{\varepsilon,v}_t\big)\big]$ is finite, yielding the claim. Moreover, the volatility equation is explicit so we shall not be concerned with uniqueness and the rest of Assumptions 4.2 and 4.5 is straightforward to check. This implies that Propositions 4.3 and 4.6 apply. Again, Theorem 3.25 guarantees that the LDP and MDP still hold when K_2 is replaced with the non-convolution fractional Brownian motion kernel. Gassiat [50] showed that, if $\rho \le 0$, then the stock price process is a true martingale, ensuring that Assumption 4.7 holds, and implied volatility asymptotics thus follow from Corollary 4.8.

4.2.3. *Rough Heston.* As introduced in [41] the rough Heston model fits into the framework of (4.1) with $K_1(t) = K_2(t) = \frac{1}{\Gamma(H+\frac{1}{2})} t^{H-\frac{1}{2}}$, for $H \in (0, \frac{1}{2})$, $\Sigma(y) = y$, $\mathfrak{b}(s, y) = \kappa\big(\theta(s) - y\big)$, $\kappa, \theta > 0$ and $\zeta(y) = \xi\sqrt{y}$, $\xi \in \mathbb{R}^d$. Linear growth and local Hölder continuity of the coefficients clearly holds. The weak existence and uniqueness was proved in [1], however the square-root coefficient brings an issue for pathwise uniqueness of the SVE. We assume here that there exists a set $\mathcal{U}$ of coefficients $(H, \kappa, \theta_., \xi, \rho, y_0)$ such that pathwise uniqueness indeed stands. The only known result so far is due to [74] in the smooth case $H = \frac{1}{2}$. We also recall that pathwise uniqueness was proved for $\zeta(y) = y^\gamma$

where $\gamma > \frac{1}{2H+1}$ in [67], but does not encompass the square root case. Therefore Assumption 4.2 holds in those cases. On a heuristic note remark that, even if pathwise uniqueness fails, there is a unique candidate for $\mathcal{G}^\varepsilon$ since there exists a unique strong solution until the first hitting time of zero. The issue is it may not satisfy the SVE anymore after that time, but should be consistent for small-time LDP.

Moreover, uniqueness of the limit equation (4.4) only holds up to first hitting time of zero. Hence we will make use of the uniqueness relaxation presented in Section 3.1 and the same arguments as in Example 3.5 to prove that Assumption 3.1 holds. The suggested rate function (3.3) reads now

$$
\begin{aligned}
I(\phi,\varphi) =& \Bigg\{ \frac{1}{2}\int_0^t \left(u_t^2+v_t^2\right)\mathrm{d}t : \quad u,v\in L^2, \\
& \qquad \phi_t = \int_0^t \sqrt{\varphi_s}\left[\overline{\rho}u_s+\rho v_s\right]\mathrm{d}s, \quad \varphi_t = y_0 + \xi\int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}\sqrt{\varphi_s}v_s\mathrm{d}s \Bigg\} \\
=& \int_0^T \frac{\mathbb{1}_{\varphi_t>0}}{2\overline{\rho}^2\varphi_t}\left(\left|\dot{\phi}_t\right|^2 - 2\rho\dot{\phi}_t \frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi-y_0)_t}{\xi} + \left|\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi-y_0)_t}{\xi}\right|^2\right)\mathrm{d}t,
\end{aligned} \tag{4.5}
$$

if the integral is well defined, $\phi\in\mathcal{AC}_0$, $\varphi\in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$, and $I=+\infty$ otherwise.

Lemma 4.9. *Let $v\in L^2$ and φ satisfying the Volterra equation (4.4) with $K_2(t)=\frac{t^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}$, $H\in(0,\frac{1}{2})$, $y_0>0, \zeta(y)=\xi\sqrt{y}, \xi\in\mathbb{R}$. Then the set $\mathcal{D}:=\{t\in[0,T]:\varphi_t>0\}$ has Lebesgue measure T.*

Proof. We follow some arguments in the proof of [1, Theorem 3.6]. Let us drop the subscript in the kernel and write it K for clarity, and introduce its resolvent of the first kind $L(\mathrm{d}t) := \frac{t^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2}-H)}\mathrm{d}t$ [53, Definition 5.5.1]. Moreover, for $h>0$, define $\Delta_h K(t) := K(t+h)$ and for every measurable function f on $\mathbb{R}_+$ and measure g on $\mathbb{R}_+$, $(f*g)(t) := \int_0^t f(t-s)\mathrm{d}g(s)$. It is proved, in [1, Equation (3.9)], that $\Delta_h K * L$ is non-decreasing but in fact in this special (rough) case, it is strictly increasing. Indeed, the authors show that in the general case, for all $0\le s\le t\le T$,

$$
(\Delta_h K * L)(t) - (\Delta_h K * L)(s) = \int_0^h K(h-u)\big(L(s+\mathrm{d}u)-L(t+\mathrm{d}u)\big),
$$

which is positive because $K>0$ and L is decreasing. Furthermore K is decreasing and $L>0$ thus

$$
0 < (\Delta_h K * L)(t) < (K*L)(t) = 1,
$$

where the equality holds by definition. Let $\varphi_t = y_0 + \int_0^t K(t-s)\xi\sqrt{\varphi_s}v_s\mathrm{d}s =: y_0 + (K*z)(t)$, where z is trivially a semimartingale, hence from [1, Equation (2.15)]:

$$
y_0 + (\Delta_h K * \mathrm{d}z)(t) = \Big(1-(\Delta_h K*L)(t)\Big)y_0 + \Big(\Delta_h K * L\Big)(0)\varphi_t + \Big(\mathrm{d}(\Delta_h K*L)*\varphi\Big)(t),
$$

which is strictly positive because because $y_0,\varphi\ge 0$ and the two lines before. Now let us suppose there exists an interval $[t,t+h]\subset\mathbb{T}$ on which $\varphi=0$. Then

$$
\varphi_{t+h} = y_0 + \int_0^t K(t+h-s)\sqrt{\varphi_s}v_s\mathrm{d}s = y_0 + (\Delta_h K*\mathrm{d}z)(t) > 0,
$$

which is a contradiction. Hence no such interval exists and the claim follows. □

Remark 4.10. *This argument works for any rough kernel but not for the diffusion case $H=\frac{1}{2}$. We refer to* [35, Proposition 3.3] *for the latter.*

We can now prove the following:

Lemma 4.11. *Let $H\in(0,\frac{1}{2})$, then the functional I satisfies Assumption 3.1.*

Proof. Note that any solution φ of the second Volterra equation in (4.5) is non-negative. Let (ϕ,φ) be such that $\Lambda(\phi,\varphi)$ is finite. Then, for each $\delta>0$, define $\varphi^\delta_t := \varphi_t + \delta t^{H+\frac{1}{2}}$ such that φ^δ is strictly positive. Therefore from definition (3.20):

$$\begin{aligned}\mathrm{D}^{H+\frac{1}{2}}(\varphi^\delta - y_0)(t) - \mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t (t-s)^{-H-\frac{1}{2}}(\varphi^\delta_s - \varphi_s)\mathrm{d}s\\ &= \delta\frac{\mathrm{d}}{\mathrm{d}t}\int_0^t (t-s)^{-H-\frac{1}{2}} s^{H+\frac{1}{2}}\mathrm{d}s = \delta\,\frac{\pi\left(H+\frac{1}{2}\right)}{\cos(\pi H)}.\end{aligned}$$

Now define the control

$$v_t := \frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t)}{\sqrt{\varphi_t}}\mathbb{1}_{\varphi_t>0},$$

and v belongs to L^2 since $\Lambda(\phi,\varphi)$ is finite. Then for each $\delta>0$ the control v^δ defined as

$$v^\delta_t := \frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi^\delta - y_0)(t)}{\sqrt{\varphi^\delta_t}} \le \frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi^\delta - y_0)(t)}{\sqrt{\varphi_t}}$$

is also in L^2 because $\mathbb{1}_{\varphi_t>0} = 1$ almost everywhere. Furthermore, for all $t\in\mathcal{D}$, $\lim_{\delta\downarrow 0}\left(\varphi^\delta_t\right)^{-1} = (\varphi_t)^{-1}$ and therefore $\lim_{\delta\downarrow 0} v^\delta_t = v_t$. Let $P(\varphi)$ denote the term between brackets in (4.5) divided by $2\overline{\rho}^2$, which is non-negative by design since it corresponds to $\varphi(u^2+v^2)$. Therefore, by Lemma 4.9,

$$I(\phi,\varphi) - I(\phi,\varphi^\delta) = \int_{\mathcal{D}}\left(\frac{1}{\varphi_t} - \frac{1}{\varphi^\delta_t}\right)P(\varphi)_t\mathrm{d}t + \int_{\mathcal{D}}\frac{1}{\varphi^\delta_t}\Big(P(\varphi)_t - P(\varphi^\delta)_t\Big)\mathrm{d}t,$$

where the first integrand is smaller than $P(\varphi)_t/\varphi_t$ for all $t\in\mathcal{D}$ and this upper bound belongs to L^1 by assumptions. Hence the dominated convergence theorem implies that the first integral goes to zero. From the calculations above we deduce that $P(\varphi)_t - P(\varphi^\delta)_t$ tends to zero uniformly as δ goes to zero, hence the second integrand converges pointwise and, for δ small enough, is dominated by $P(\varphi)/\varphi$. A second application of DCT yields convergence of the integral, and the claim follows. □

Therefore the large and moderate deviations from Propositions 4.3 and 4.6 apply if the coefficients belong to $\mathcal{U}$ and $H\in(0,\frac{1}{2}]$. Observe that Proposition 4.6**(M3)** agrees with [42, Section 3.5], although the routes taken differ significantly. El Euch and Rosenbaum [40, Appendix B] showed that Assumption 4.7 is satisfied, and the implied volatility behaviour thus follows from Corollary 4.8.

4.2.4. *Multifactor rough Bergomi.* Let W be an $\mathbb{R}^{m+1}$-Brownian motion, $\boldsymbol{Z} := \left(Z^{(1)},\cdots,Z^{(m)}\right)^\dagger$ where

$$Z^{(j)}_t := \int_0^t K^{(j)}(t-s)\mathrm{d}W^{(j)}_s,$$

and we allow $K^{(j)}, 1\le j\le m$, to be homogeneous of different degrees $H_j-\frac{1}{2}$ with $H_j\in(0,1]$. Assume without loss of generality that the H_j are ordered by increasing values, then we will design the rescaling at the speed ε^{-H_1}. Denote $m^\star := \inf\{j=2,..,m : H_j > H_1\}$. Let L be an m-dimensional square matrix and Y an m-dimensional process defined for all $t\in\mathbb{T}$ by

$$Y_t := y_0 + \mathrm{L}\boldsymbol{Z}_t - at^{2H_1},$$

where $y_0, a\in\mathbb{R}^m$ and the log price reads

$$X_t = -\frac{1}{2}\int_0^t\sum_{i=1}^m \exp\left(Y^{(i)}_s\right)\mathrm{d}s + \int_0^t\sum_{i=1}^m\exp\left(\frac{1}{2}Y^{(i)}_s\right)\mathrm{d}B_s.$$

The rescaling $X_t^\varepsilon = \varepsilon^{H_1-\frac{1}{2}} X_{\varepsilon t}, Y_t^\varepsilon = Y_{\varepsilon t}$ yields

$$\begin{cases} X_t^\varepsilon = -\dfrac{\varepsilon^{H_1+\frac{1}{2}}}{2}\displaystyle\int_0^t \sum_{i=1}^m \exp(Y_s^{\varepsilon,(i)})\mathrm{d}s + \varepsilon^{H_1}\int_0^t \sum_{i=1}^m \exp\left(\frac{1}{2}Y_s^{\varepsilon,(i)}\right)\mathrm{d}B_s, \\ Y_t^{\varepsilon,(i)} = y_0^{(i)} + \displaystyle\sum_{j=1}^m \varepsilon^{H_j}\mathrm{L}_{i,j}Z_t^{(j)} - a_i(\varepsilon t)^{2H_1} = y_0^{(i)} + \varepsilon^{H_1}\sum_{j=1}^m \mathrm{L}^\varepsilon_{i,j}Z_t^{(j)} - a_i(\varepsilon t)^{2H_1}, \quad \text{for all } 1\le i\le m, \end{cases}$$

where $B = \overline{\rho}W^{(m+1)} + \sum_{i=1}^m \rho_i W^{(i)}$, $\overline{\rho}^2 + \sum_{i=1}^m \rho_i^2 = 1$, and $\mathrm{L}^\varepsilon_{i,j} := \varepsilon^{H_j-H_1}\mathrm{L}_{i,j}$ goes to zero if $H_j > H_1$, i.e. if $j > m^\star$. It means that the roughest component(s) (the one(s) with H_1) will outweigh the others, and only the former will make a contribution to the rate function.

Although similar to its one-dimensional counterpart, this model does not fit into the framework of (4.1). Regarding the assumptions of Theorems 3.8, we only check **H3b** and Assumption 3.6 because the others are standard and similar to the one-dimensional case. Clearly $(Y^{\varepsilon,(1)},\cdots,Y^{\varepsilon,(m)})$ is an autonomous subsystem. As a Gaussian process, $\mathrm{L}\boldsymbol{Z}$ has exponential moments of all orders and for all $N>0$, $1\le j\le m$ and $v^j \in \mathcal{S}_N$:

$$\int_0^t K^{(j)}(t-s)v_s^{(j)}\mathrm{d}s \le \sqrt{N}\|K^{(j)}\|_2 \quad \text{almost surely},$$

thus $\exp\left(Y^{\varepsilon,(i),v}\right) \in L^p(\Omega)$ for all $p\ge 1$. Therefore the bound (3.6) and Assumption 3.6 are satisfied. This estimate also checks that (3.17) and thus **H8** stand. Since $\vartheta = \varepsilon^{H_1}$, we define for the moderate deviations regime $h_\varepsilon = \varepsilon^\beta, \beta\in(0,H_1)$.

Corollary 4.12. *Denoting* $Y^\varepsilon := (Y^{\varepsilon,(1)},\cdots,Y^{\varepsilon,(m)})$, *we have:*

- $(X^\varepsilon, Y^\varepsilon) \sim \mathrm{LDP}\left(I, \varepsilon^{-H_1}\right)$ *where for all* $\phi\in\mathcal{W}$, $\varphi\in\mathcal{W}^m$*:*

$$\begin{aligned} I(\phi,\varphi) = \inf\Bigg\{ &\frac{1}{2}\int_0^T \left(u_t^2+v_t^2\right)\mathrm{d}t : \quad u\in L^2(\mathbb{R}), v\in L^2(\mathbb{R}^m), \\ &\phi_t = \int_0^t \left[\sum_{i=1}^m \mathrm{e}^{\varphi_s^{(i)}/2}\right]\left[\overline{\rho}u_s + \sum_{j=1}^m \rho_j v_s^{(j)}\right]\mathrm{d}s, \quad \varphi_t^{(i)} = y_0^{(i)} + \int_0^t \sum_{j=1}^{m^\star} K^{(j)}(t-s)\mathrm{L}_{i,j}v_s^{(j)}\mathrm{d}s \Bigg\}. \end{aligned}$$

- $(X^\varepsilon, Y^\varepsilon) \sim \mathrm{MDP}\left(\Lambda, \varepsilon^{-\beta}\right)$ *where for all* $\phi\in\mathcal{W}$ *and* $\varphi\in\mathcal{W}^m$*,*

$$\begin{aligned} \Lambda(\phi,\varphi) = \inf\Bigg\{ &\frac{1}{2}\int_0^T \left(u_t^2+v_t^2\right)\mathrm{d}t : \quad u\in L^2(\mathbb{R}), v\in L^2(\mathbb{R}^m), \\ &\phi_t = \int_0^t \left[\sum_{i=1}^m \mathrm{e}^{y_0^{(i)}/2}\right]\left[\overline{\rho}u_s + \sum_{j=1}^m \rho_j v_s^{(j)}\right]\mathrm{d}s, \quad \varphi_t^{i} = y_0^{(i)} + \int_0^t \sum_{j=1}^{m^\star} K^{(j)}(t-s)\mathrm{L}_{i,j}v_s^{(j)}\mathrm{d}s \Bigg\}. \end{aligned} \tag{4.6}$$

Proof. The LDP is a direct application of Theorem 3.8 and the MDP of Theorem 3.16. □

If L is lower triangular (i.e. $\mathrm{L}_{i,j}=0$ for all $i<j$), for instance if it arises from the Cholesky decomposition of a covariance matrix, then for all $\phi\in\mathcal{W}$, $\varphi\in\mathcal{W}^m$ one can derive the vector v recursively. Note that if $m^\star < m$, φ may not be attainable by the restrained number of controls $\{v^{(j)}, 1\le j\le m^\star\}$.

Example 4.13. *Consider the case $m=2$. Let $K^1(t) = K^2(t) = \frac{1}{\Gamma(H+\frac{1}{2})}t^{H-\frac{1}{2}}$ for $H\in(0,\frac{1}{2})$ and L be lower triangular (i.e. $L_{12}=0$). Then v and u are explicit from* (4.6)*:*

$$v^{(1)} = \frac{1}{L_{11}}\mathrm{D}^{H+\frac{1}{2}}\left(\varphi^{(1)} - y_0^{(1)}\right), \qquad v^{(2)} = \frac{1}{L_{22}}\mathrm{D}^{H+\frac{1}{2}}\left(\varphi^{(2)} - y_0^{(2)}\right) - \frac{L_{21}}{L_{22}}v^{(1)},$$

$$u = \frac{1}{\overline{\rho}}\left[\dot{\phi}\left\{\exp\left(\frac{y_0^{(1)}}{2}\right) + \exp\left(\frac{y_0^{(2)}}{2}\right)\right\}^{-1} - \rho_1 v^{(1)} - \rho_2 v^{(2)}\right].$$

Remark 4.14. *We can similarly consider multidimensional versions of the other models presented in this paper and derive large and moderate deviation principles. We only work out the computations for the multifactor rough Bergomi model because it is the most relevant in the literature.*

4.3. **Tail rescaling.** We now investigate tail rescalings, which generally have the form $X^\varepsilon = \varepsilon X$, such that an LDP provides asymptotic estimates on $\mathbb{P}(X^\varepsilon \geq 1) = \mathbb{P}(X \geq \varepsilon^{-1})$. The MDP for the whole system is not available in this case because $\overline{Y} := \lim_{\varepsilon\downarrow 0} Y^\varepsilon \equiv 0$ hence the limit equation for X^ε, arising from (3.16), would be independent of the control. Note that the theory does not break down but the rate function is trivial (equals zero at zero and $+\infty$ everywhere else). Furthermore, the exponential function prevents the study of such a rescaling in the rough Bergomi model.

4.3.1. *Rough Stein-Stein.* This model was defined in Section 4.2.1, but with the rescaling $Y_t^\varepsilon := \varepsilon Y_t$ and $X_t^\varepsilon := \varepsilon^2 X_t$, the system becomes

$$\left\{\begin{array}{ll} X_t^\varepsilon & = -\dfrac{1}{2}\displaystyle\int_0^t (Y_s^\varepsilon)^2 \mathrm{d}s + \varepsilon \int_0^t Y_s^\varepsilon \mathrm{d}B_s, \\ Y_t^\varepsilon & = \varepsilon y_0 + \displaystyle\int_0^t \kappa\left(\varepsilon\theta - Y_s^\varepsilon\right)\mathrm{d}s + \varepsilon\int_0^t \xi \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}\mathrm{d}W_s, \end{array}\right. \tag{4.7}$$

where the coefficients are identical to the small-time case. Although the rescaling is diferent, Assumption 2.3 and **H1 - H4** are easily satisfied in a similar way, the limit equation (3.5) has a unique solution, and therefore Theorem 3.8 applies.

Corollary 4.15. *The following hold:*

(L1) $(X^\varepsilon, Y^\varepsilon) \sim \mathrm{LDP}\left(I, \varepsilon^{-1}\right)$ *with*

$$I(\phi,\varphi) = \frac{1}{2}\int_0^T \left(u_t^2 + v_t^2\right)\mathrm{d}t, \quad \textit{where} \quad \left\{\begin{array}{ll} u & = \dfrac{1}{\overline{\rho}}\left(\dfrac{\dot{\phi}}{\varphi} + \dfrac{1}{2}\varphi - \rho v\right)\mathbb{1}_{\varphi\neq 0}, \\ v & = \dfrac{1}{\xi}\left(\mathrm{D}^{H+\frac{1}{2}}(\varphi) + \kappa \mathrm{I}^{\frac{1}{2}-H}(\varphi)\right), \end{array}\right.$$

if $\phi, \varphi \in \mathcal{AC}_0$ *and infinity otherwise.*

(L2) $X^\varepsilon \sim \mathrm{LDP}\left(I^X, \varepsilon^{-1}\right)$ *with* $I^X(\phi) = \inf\left\{I(\phi,\varphi) : \varphi \in \mathcal{AC}_0\right\}$ *if* $\phi \in \mathcal{AC}_0$ *and infinity otherwise.*

(L3) *For each* $t \in \mathbb{T}$, $\varepsilon^2 X_t \sim \mathrm{LDP}\left(I_t^X, \varepsilon^{-1}\right)$, *where* $I_t^X(x) = \inf\left\{I^X(\phi) : \phi_t = x\right\}$.

Proof. For **(L1)**, Theorem 3.8 entails that the rate function is

$$\begin{aligned} I(\phi,\varphi) = \inf\Bigg\{ & \frac{1}{2}\int_0^T \left(u_t^2 + v_t^2\right)\mathrm{d}t : \quad u, v \in L^2, \\ & \phi_t = -\frac{1}{2}\int_0^t \varphi_s^2 \mathrm{d}s + \int_0^t \varphi_s\left[\overline{\rho}u_s + \rho v_s\right]\mathrm{d}s,\ \varphi_t = -\int_0^t \kappa\,\varphi_s \mathrm{d}s + \int_0^t \xi\,\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} v_s \mathrm{d}s \Bigg\}. \end{aligned}$$

Inverting the integrals to obtain the unique controls yields the claim. Similarly to the small-time case, **(L2)** follows from the contraction principle, and one only needs to fix $t \in \mathbb{T}$ to prove **(L3)**. ☐

One can prove an LDP for Y^ε in a similar way; a more interesting problem is the moderate deviations setting. Recall that an MDP for the couple $(X^\varepsilon, Y^\varepsilon)$ would have a trivial rate function because the limit equation of X^ε is independent of the control. However, since the diffusion coefficient of Y^ε is constant equal to ξ, one can obtain an MDP for Y^ε. More surprisingly, the limit equations in the large deviations (3.5) and moderate deviations (3.16) regimes coincide, which leads to identical rate functions. Notice that $\vartheta_\varepsilon = \varepsilon$ in this example, and therefore let $h_\varepsilon = \varepsilon^{-\beta}$ where $\beta \in (0,1)$.

Corollary 4.16. $Y^\varepsilon \sim \mathrm{MDP}\left(\Lambda^Y, \varepsilon^{-\beta}\right)$ *where,*

$$\Lambda^Y(\varphi) = \frac{1}{2\xi^2}\int_0^T \left(\mathrm{D}^{H+\frac{1}{2}}(\varphi)(t) + \kappa \mathrm{I}^{\frac{1}{2}-H}(\varphi)(t)\right)^2 \mathrm{d}t,$$

if $\varphi \in \mathcal{AC}_0$ *and infinity otherwise.*

Proof. From (4.7), $b_\varepsilon(y) := \kappa(\varepsilon\theta - y)$ converges to $b(y) = -\kappa y$ and the diffusion coefficient is constant, hence **H2 - H6** are easily satisfied. Moreover, $b_\varepsilon - b \equiv \varepsilon\kappa\theta$ and $\varepsilon^{1-(H-\beta)}$ tends to zero thus **H7** and **H8** also hold. Theorem 3.16 thus yields an MDP with rate function

$$\Lambda^Y(\varphi) = \inf\left\{\frac{1}{2}\int_0^T v_t^2 \mathrm{d}t : v \in L^2, \quad \varphi_t = -\int_0^t \kappa\varphi_s \mathrm{d}s + \xi\int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} v_s \mathrm{d}s\right\}.$$

Inverting the integral yields the claim. □

4.3.2. *Rough Heston.* Introduced in Section 4.2.3, with the rescaling $Y_t^\varepsilon := \varepsilon^2 Y_t$ and $X_t^\varepsilon := \varepsilon^2 X_t$, we obtain

$$\begin{cases} X_t^\varepsilon &= -\dfrac{1}{2}\displaystyle\int_0^t Y_s^\varepsilon \mathrm{d}s + \varepsilon\int_0^t \sqrt{Y_s^\varepsilon}\mathrm{d}B_s, \\ Y_t^\varepsilon &= \varepsilon^2 y_0 + \displaystyle\int_0^t \kappa\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}\left(\varepsilon^2\theta - Y_s^\varepsilon\right)\mathrm{d}s + \varepsilon\int_0^t \xi\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}\sqrt{Y_s^\varepsilon}\mathrm{d}W_s. \end{cases}$$

Clearly **H1 - H3** hold and we recall that $\mathcal{U}$ is the set of coefficients such that pathwise uniqueness, and hence **H4**, hold. We appeal to the uniqueness relaxation in the same way as the small-time case to prove the following result, which extends [27, Theorem 1.1] to the rough case.

Corollary 4.17. *If the rough Heston coefficients belong to $\mathcal{U}$ and $H \in (0, \frac{1}{2})$, then the following hold*

(L1) $(X^\varepsilon, Y^\varepsilon) \sim \mathrm{LDP}\left(I, \varepsilon^{-1}\right)$ *where*

$$I(\phi,\varphi) = \frac{1}{2}\int_0^T \left(u_t^2 + v_t^2\right)\mathrm{d}t, \quad \textit{where} \begin{cases} u &= \dfrac{1}{\overline{\rho}}\left(\dfrac{\dot\phi}{\sqrt{\varphi}} + \dfrac{1}{2}\sqrt{\varphi} - \rho v\right)\mathbb{1}_{\varphi>0}, \\ v &= \dfrac{1}{\xi\sqrt{\varphi}}\left(\mathrm{D}^{H+\frac{1}{2}}(\varphi) + \kappa\varphi\right)\mathbb{1}_{\varphi>0}, \end{cases}$$

where $\phi \in \mathcal{AC}_0$ and $\varphi \in \mathrm{I}_0^{H+\frac{1}{2}}(L^1)$ and infinity otherwise.

(L2) $X^\varepsilon \sim \mathrm{LDP}\left(I^X, \varepsilon^{-1}\right)$ *with $I^X(\phi) = \inf\left\{I(\phi,\varphi) : \varphi \in \mathrm{I}_0^{H+\frac{1}{2}}\right\}$ if $\phi \in \mathcal{AC}_0$ and infinity otherwise.*

(L3) *For each $t \in \mathbb{T}$, $\varepsilon^2 X_t \sim \mathrm{LDP}\left(I_t^X, \varepsilon^{-1}\right)$, where $I_t^X(x) = \inf\left\{I^X(\phi) : \phi_t = x\right\}$.*

Proof. The proof is similar to the small-time case. The potential rate function for the couple is

$$I(\phi,\varphi) = \inf\left\{\frac{1}{2}\int_0^T \left(u_t^2 + v_t^2\right)\mathrm{d}t : \quad u, v \in L^2, \right.$$
$$\left.\phi_t = -\frac{1}{2}\int_0^t \varphi_s \mathrm{d}s + \int_0^t \sqrt{\varphi_s}\left[\overline{\rho}u_s + \rho v_s\right]\mathrm{d}s, \quad \varphi_t = \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}\Big(-\kappa\varphi_s + \xi\sqrt{\varphi_s}v_s\Big)\mathrm{d}s\right\}.$$

The same arguments that were used to prove Lemmata 4.9 and 4.11 in the small-time case can be applied again here. They entail that Assumption 3.1 holds and hence Theorem 3.8 applies, and the form of the rate function in **(L1)** follows by inverting the relationships between (u, v) and (ϕ, φ). **(L2)** and **(L3)** follow from the same steps as in Corollary 4.15. □

4.3.3. *Implied volatility asymptotics.* We can also obtain implied volatility asymptotics since, by the same arguments as before, $\exp(X^\varepsilon)$ is a martingale in both the rough Stein-Stein and rough Heston models.

Corollary 4.18. *In both the rough Stein-Stein and rough Heston models, for each $t \in \mathbb{T}$, the implied volatility $\widehat{\sigma}$ satisfies*

$$\lim_{k\uparrow\infty} \frac{\widehat{\sigma}(t,k)^2 t}{k} = \frac{1}{2}\left(\inf_{y\geq 1} I_t^X(y)\right)^{-1},$$

*where I_t^X is the respective rate function, given in Corollaries 4.15***(L3)** *and 4.17***(L3)***.*

Proof. Mapping ε^2 to $1/k$ we have from Corollaries 4.15 and 4.17 respectively that, for each $t \in \mathbb{T}$,

$$\lim_{k\uparrow\infty} \frac{1}{k} \log \mathbb{P}(X_t \geq k) = -\inf_{y\geq 1} I_t^X(y).$$

In the Black-Scholes model with constant volatility $\sigma > 0$, we can directly compute

$$\lim_{k\uparrow\infty} \frac{1}{k} \log \mathbb{P}(X_t \geq k) = -\frac{1}{2\sigma^2 t},$$

and, similarly to the small-time case, the proof follows from [49, Corollary 7.1]. □

Appendix A. Technical large deviations proofs

A.1. **Abstract relaxation: Proof of Theorem 3.3.** The proof follows [13, Theorem 4.4]. The lower bound stands as it is except that at the very end, when taking the infimum, the condition becomes $\phi \in \mathcal{G}_v^0$ instead of $\phi = \mathcal{G}_v^0$. Thus it suggests the potential rate function I in (3.3). We prove the Laplace principle upper bound, for all $F \in \mathcal{C}_b(\mathcal{W}^d : \mathbb{R})$:

$$\limsup_{\varepsilon\downarrow 0} -\varepsilon^2 \log \mathbb{E}\left[\mathrm{e}^{-F\circ\mathcal{G}^\varepsilon(W)/\varepsilon^2}\right] \leq \inf_{\psi\in\mathcal{W}^d}\{I(\psi) + F(\psi)\}.$$

Assume that the right-hand side is finite otherwise there is nothing to prove. Fix $\epsilon > 0$ and let $\phi \in \mathcal{W}^d$ such that

$$I(\phi) + F(\phi) \leq \inf_{\psi\in\mathcal{W}^d}\{I(\psi) + F(\psi)\} + \epsilon.$$

Since F is continuous, there exists $\delta \in (0, \epsilon)$ such that $\left|F(\phi^1) - F(\phi^2)\right| \leq \epsilon$ when $\left\|\phi^1 - \phi^2\right\|_{\mathbb{T}} \leq \delta$, for all $\phi^1, \phi^2 \in \mathcal{W}^d$. If ϕ is uniquely characterised then the proof is the same as in [13]. Otherwise, by Theorem 3.3(iii), we can choose ϕ^δ uniquely characterised such that $\left\|\phi - \phi^\delta\right\|_{\mathbb{T}} \leq \delta$ and $\left|I(\phi) - I(\phi^\delta)\right| \leq \delta$, which implies $\left|I(\phi) + F(\phi) - I(\phi^\delta) - F(\phi^\delta)\right| \leq 2\epsilon$. Hence, combining inequalities we obtain

$$I(\phi^\delta) + F(\phi^\delta) \leq \inf_{\psi\in\mathcal{W}^d}\{I(\psi) + F(\psi)\} + 3\epsilon.$$

Moreover, there exist $\{v^n\}_{n\in\mathbb{N}}$ in L^2 such that (3.2) is satisfied with ϕ^δ and $m \geq 1/\epsilon$ such that

$$\mathcal{G}_{v^m}^0 = \{\phi^\delta\} \quad \text{and} \quad \frac{1}{2}\int_0^T |v_t^m|^2 \,\mathrm{d}t \leq I(\phi^\delta) + \frac{1}{m} \leq I(\phi^\delta) + \epsilon,$$

and therefore the remainder of the upper bound proof unfolds identically.

Along the subsequence $\{\varepsilon_n\}_{n\geq 0}$, $\mathcal{G}^{\varepsilon_n}(W + \varepsilon_n^{-1}\int_0^\cdot v_t^m \mathrm{d}t)$ converges in distribution to ϕ^δ. Using the variational representation formula (1.3) and the convergence we obtain

$$\begin{aligned}
\limsup_{n\uparrow\infty} -\varepsilon_n^2 \log \mathbb{E}\left[\exp\left\{-\frac{F\circ\mathcal{G}^{\varepsilon_n}(W)}{\varepsilon_n^2}\right\}\right] &= \limsup_{n\uparrow\infty} \inf_{v\in\mathcal{A}} \mathbb{E}\left[\frac{1}{2}\int_0^T |v_t|^2 \,\mathrm{d}t + F\circ\mathcal{G}^{\varepsilon_n}\left(W + \varepsilon_n^{-1}\int_0^\cdot v_t \mathrm{d}t\right)\right] \\
&\leq \limsup_{n\uparrow\infty} \mathbb{E}\left[\frac{1}{2}\int_0^T |v_t^m|^2 \,\mathrm{d}t + F\circ\mathcal{G}^{\varepsilon_n}\left(W + \varepsilon_n^{-1}\int_0^\cdot v_t^m \mathrm{d}t\right)\right] \\
&= \frac{1}{2}\int_0^T |v_t^m|^2 \,\mathrm{d}t + F(\phi^\delta) \\
&\leq I(\phi^\delta) + F(\phi^\delta) + \epsilon \leq \inf_{\psi\in\Omega}\{I(\psi) + F(\psi)\} + 4\epsilon.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary this concludes the proof.

A.2. **LDP moment bounds: Proof of Lemma 3.10.** Let us fix $p\ge 2$, $N>0$, $v\in\mathcal{A}_N$, $\varepsilon>0$ and $t\in\mathbb{T}$. Let $\tau_n:=\inf\{t\ge 0: |X_t^{\varepsilon,v}|\ge n\}\wedge T$ for all $n\in\mathbb{N}$. For clarity we write $b_s^n:=b_\varepsilon(s,X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n})$ and $\sigma_s^n:=\sigma_\varepsilon(s,X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n})$. We start by assuming that all the coefficients satisfy the linear growth condition **H3a**. We fix $n\in\mathbb{N}$ and observe that, almost surely:

$$\begin{aligned} |X_t^{\varepsilon,v}|^p\,\mathbb{1}_{t\le\tau_n} &\le 4^{p-1}\left[|X_0^\varepsilon|^p+\left|\int_0^t K(t-s)b_s^n\mathrm{d}s\right|^p+\left|\int_0^t K(t-s)\sigma_s^n v_s\mathrm{d}s\right|^p+\vartheta_\varepsilon^p\left|\int_0^t K(t-s)\sigma_s^n\mathrm{d}W_s\right|^p\right]\\ &=:4^{p-1}\Big[|X_0^\varepsilon|^p+\mathrm{I}_n+\mathrm{II}_n+\mathrm{III}_n\Big], \end{aligned} \tag{A.1}$$

because if $t>\tau_n$ then the left-hand side is zero while the right-hand side is non-negative, and if $t\le\tau_n$ then $s\le\tau_n$ for all $s\in[0,t]$ and the τ_n dependence vanishes on both sides of the inequality. For ε small enough we can bound $|X_0^\varepsilon|$ by $2\,|X_0|$ and ϑ_ε by 1 and we will do so repetitively in the sequel. Using Hölder's and Jensen's inequalities, we obtain the following estimates almost surely:

$$\mathrm{I}_n\le\left[\int_0^t|K(t-s)|^{\frac{4}{p}}\,|b_s^n|^2\,\mathrm{d}s\right]^{\frac{p}{2}}\left[\int_0^t|K(t-s)|^{2-\frac{4}{p}}\,\mathrm{d}s\right]^{\frac{p}{2}}\le t^{\frac{p}{2}}\,\|K\|_2^{p-2}\int_0^t|K(t-s)|^2\,|b_s^n|^p\,\mathrm{d}s, \tag{A.2}$$

and

$$\mathrm{II}_n\le N^{\frac{p}{2}}\left[\int_0^t|K(t-s)|^{2-4/p}\,|K(t-s)|^{\frac{4}{p}}\,|\sigma_s^n|^2\,\mathrm{d}s\right]^{\frac{p}{2}}\le N^{\frac{p}{2}}\,\|K\|_2^{p-2}\int_0^t|K(t-s)|^2\,|\sigma_s^n|^p\,\mathrm{d}s, \tag{A.3}$$

where we also used that $\int_0^T|v_s|^2\,\mathrm{d}s\le N$ almost surely. Notice that $\{\int_0^u K(t-s)\sigma_s^n\mathrm{d}W_s, u\in\mathbb{T}\}$ is a continuous local martingale for fixed $t\in\mathbb{T}$ and is bounded in $L^2(\Omega)$. Hence, using Burkholder-Davis-Gundy (BDG) inequality and similar calculations as (A.3) there exists $C_p>0$ such that

$$\mathbb{E}[\mathrm{III}_n]\le C_p\,\|K\|_2^{p-2}\int_0^t|K(t-s)|^2\,\mathbb{E}\,|\sigma_s^n|^p\,\mathrm{d}s.$$

From the linear growth condition on b_ε and σ_ε (uniform in $\varepsilon>0$) we deduce that there exists $C_1>0$ independent of ε,v,n,t such that, for all $n\in\mathbb{N}$, $f_t^n:=\mathbb{E}\left[|X_t^{\varepsilon,v}|^p\,\mathbb{1}_{t\le\tau_n}\right]$ satisfies the inequality

$$f_t^n\le C_1+C_1\int_0^t|K(t-s)|^2\,f_s^n\mathrm{d}s. \tag{A.4}$$

The following lemma (Lemma A.1) yields a uniform bound in both $n\in\mathbb{N}$ and $t\in\mathbb{T}$ for f_t^n. Taking the limit as n goes to infinity and using Fatou's lemma concludes the first part of the proof.

Lemma A.1. *Let $f:\mathbb{T}\to\mathbb{R}$ and K a kernel satisfying Assumption 2.3. If there exists $c\ge 0$ such that*

$$f(t)\le c+c\int_0^t|K(t-s)|^2\,f(s)\mathrm{d}s,\qquad\text{for all } t\in\mathbb{T},$$

then f is uniformly bounded on $\mathbb{T}$ by a constant depending only on $c,\|K\|_2,T$. If $c=0$ then $f=0$.

Proof of Lemma A.1. By definition $|K(t-s)|^2=\sum_{i,j=1}^d|K_{i,j}(t-s)|^2$, and $\widetilde{K}(t,s):=c\,|K(t-s)|^2\,\mathbb{1}_{s\le t}$ is a Volterra kernel in the sense of [53, Definition 9.2.1]. Following the same arguments as in the proof of [1, Lemma 3.1], the generalised Gronwall lemma [53, Theorem 9.8.2] yields the bound

$$f(t)\le c-c\int_0^t R'(s)\mathrm{d}s\le c-c\int_0^T R'(s)\mathrm{d}s,$$

where R' is the (non-positive) resolvent of type L^1 of $-K'$ [1, Equation (2.11)], proving the lemma. □

If only **H3b** and Assumption 3.6 hold with an autonomous sub-system Υ, then by the previous calculations for all $l\in\Upsilon$, the components $(X^{\varepsilon,v})^{(l)}$ satisfy the bound (3.8) because their coefficients have linear growth. Then turning our attention to the components $(X^{\varepsilon,v})^{(j)}$, $j \notin \Upsilon$ and using (3.6) and Hölder's inequality as in (A.3)

$$
\begin{aligned}
&\mathbb{E}\left[\left(\int_0^t \left|K_j(t-s)\Big(\sigma^n\big(s, X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n}\big)\Big)^{(j)}\right|^2 \mathrm{d}s\right)^{p/2}\right] \\
&\le C_\Upsilon^p \mathbb{E}\left[\left(\int_0^t |K_j(t-s)|^2 \left(1 + \left|X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n}\right|_{\Upsilon^c} + \left|\Gamma\big((X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n})^{(\Upsilon)}\big)\right|\right)^2 \mathrm{d}s\right)^{p/2}\right] \\
&\le C_\Upsilon^p \mathbb{E}\left[\int_0^t |K_j(t-s)|^2 \left(1 + |X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n}|_{\Upsilon^c} + \left|\Gamma\big((X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n})^{(\Upsilon)}\big)\right|\right)^p \mathrm{d}s\right]\left(\int_0^t |K_j(t-s)|^2\,\mathrm{d}s\right)^{p/2-1} \\
&\le 3^{p-1} C_\Upsilon^p \|K\|_2^{p-2} \int_0^t |K_j(t-s)|^2\, \mathbb{E}\left[1 + \left|X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n}\right|_{\Upsilon^c}^p + \left|\Gamma\big((X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n})^{(\Upsilon)}\big)\right|^p\right]\mathrm{d}s
\end{aligned}
$$

(A.5)

$$
\le C_2 + C_2 \int_0^t |K_j(t-s)|^2\, \mathbb{E}\left[|X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n}|_{\Upsilon^c}^p\right]\mathrm{d}s,
$$

for some $C_2 > 0$. Applying the same calculations to the other terms and summing all the coefficients we fall back on (A.4). Taking the limit and applying Fatou's lemma again conclude the proof.

A.3. **LDP tightness: Proof of Lemma 3.12.** Let us fix $p > 2 \vee 2/\gamma$, $N > 0$, $v \in \mathcal{A}_N$ and $\varepsilon > 0$. For clarity we will write $b_u := b_\varepsilon(u, X_u^{\varepsilon,v})$ and $\sigma_u := \sigma_\varepsilon(u, X_u^{\varepsilon,v})$ for all $u \in \mathbb{T}$. In a first step we assume that all the coefficients satisfy the linear growth condition **H3a**. Then, for all $0 \le s < t \le T$, using Cauchy-Schwarz and BDG inequalities as in the previous proof we obtain:

$$
\begin{aligned}
\mathbb{E}\left[\left|X_t^{\varepsilon,v^\varepsilon} - X_s^{\varepsilon,v^\varepsilon}\right|^p\right] \le\ & 6^{p-1}\mathbb{E}\left[\left|\int_0^s \big(K(t-u) - K(s-u)\big) b_u \mathrm{d}u\right|^p\right] \\
&+ 6^{p-1}\mathbb{E}\left[\left|\int_s^t K(t-u) b_u \mathrm{d}u\right|^p\right] \\
&+ 6^{p-1}N^{p/2}\mathbb{E}\left[\left(\int_0^s \left|\big(K(t-u) - K(s-u)\big)\sigma_u\right|^2 \mathrm{d}u\right)^{p/2}\right] \\
&+ 6^{p-1}N^{p/2}\mathbb{E}\left[\left(\int_s^t |K(t-u)\sigma_u|^2\,\mathrm{d}u\right)^{p/2}\right] \\
&+ 6^{p-1}\vartheta_\varepsilon^p C_p\mathbb{E}\left[\left(\int_0^s \left|\big(K(t-u) - K(s-u)\big)\sigma_u\right|^2 \mathrm{d}u\right)^{p/2}\right] \\
&+ 6^{p-1}\vartheta_\varepsilon^p C_p\mathbb{E}\left[\left(\int_s^t |K(t-u)\sigma_u|^2\,\mathrm{d}u\right)^{p/2}\right].
\end{aligned}
$$

Analogous calculations to the proof of Lemma 3.10, bounds on $\sup_{t\in\mathbb{T},\varepsilon>0} \mathbb{E}\left[\left|X_t^{\varepsilon,v^\varepsilon}\right|^p\right]$, linear growth from **H3a** and Assumption 2.3 lead to

$$
\begin{aligned}
\mathbb{E}\left[\left|X_t^{\varepsilon,v^\varepsilon} - X_s^{\varepsilon,v^\varepsilon}\right|^p\right] &\le C_1\left(\left|\int_0^s \big(K(t-u) - K(s-u)\big)^2 \mathrm{d}u\right|^{p/2} + \left|\int_s^t K(t-u)^2 \mathrm{d}u\right|^{p/2}\right) \\
&= C_1\left(\left|\int_0^s \big(K(u+t-s) - K(u)\big)^2 \mathrm{d}u\right|^{p/2} + \left|\int_0^{t-s} K(u)^2 \mathrm{d}u\right|^{p/2}\right) \\
&\le C_2 (t-s)^{\gamma p/2},
\end{aligned}
$$

for some $C_1, C_2 > 0$ independent of ε, v, s, t. Again, if there are components such that only **H3b** holds with Assumption 3.6 then following the example of (A.5) yields the same result. The Kolmogorov continuity theorem then asserts that $X^{\varepsilon,v^\varepsilon}$ admits a version which is Hölder continuous on $\mathbb{T}$ of any order $\alpha < \gamma/2 - 1/p$, uniformly in $\varepsilon > 0$ because C_2 does not depend on ε, and which satisfies (3.9). Furthermore, Aldous theorem [10, Theorem 16.10] states that the sequence $\{X^{\varepsilon,v^\varepsilon}\}_{\varepsilon>0}$ is tight.

A.4. **LDP compactness: Proof of Lemma 3.14.** We prove that for all $N > 0$, the sublevel sets

$$L_N := \{\phi \in \mathcal{W}^d : I(\phi) \leq N\}$$

of the map $I : \mathcal{W}^d \to \mathbb{R}$ given by (3.3) or

$$I(\phi) = \inf\left\{\frac{1}{2}\int_0^T |v_s|^2\,\mathrm{d}s : v \in L^2, \phi_t = x_0 + \int_0^t K(t-s)\big[b(s,\phi_s) + \sigma(s,\phi_s)v_s\big]\mathrm{d}s\right\} \tag{A.6}$$

are compact. Fix $N > 0$ and consider an arbitrary sequence $\mathcal{J} := \{\phi^n : n \in \mathbb{N}\} \subset L_N$, we will show that there exists a converging subsequence the limit of which belongs to L_N. Interestingy enough, the proof parallels, in a deterministic context, the proofs of bound, Hölder continuity and convergence of $X^{\varepsilon,v}$.

Relative compactness. According to Arzelà-Ascoli's theorem, the family $\mathcal{J}$ is relatively compact in $\mathcal{W}^d$ if and only if $\{\phi_t^n\}$ is bounded uniformly in $n \in \mathbb{N}$ and in $t \in \mathbb{T}$ and $\mathcal{J}$ is equicontinuous. Moreover, for all $n \in \mathbb{N}$ and all $t \in \mathbb{T}$, there exists $v^n \in L^2$ such that $\frac{1}{2}\int_0^T |v_t^n|^2\,\mathrm{d}t \leq N$ and $\phi^n \in \mathcal{G}^0_{v^n}$, which means $v^n \in \mathcal{S}_{2N}$ and

$$\phi_t^n = x_0 + \int_0^t K(t-s)\Big[b(s,\phi_s^n) + \sigma(s,\phi_s^n)v_s^n\Big]\mathrm{d}s.$$

Hence Remarks 3.11 and 3.13 grant the uniform bound and equicontinuity respectively. Therefore $\mathcal{J}$ is relatively compact which entails that L_N is relatively compact for any $N > 0$.

Closure. Let $\{\phi^n\}_{n\in\mathbb{N}}$ be a converging sequence of L_N and denote its limit by $\phi \in \mathcal{W}^d$. The controls v^n associated to ϕ^n through (A.6) belong to $\mathcal{S}_{2N}$ which is a compact space with respect to the weak topology. Hence there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ such that v^{n_k} converges weakly in L^2 to a limit $v \in \mathcal{S}_{2N}$ and $\lim_{k\uparrow\infty} \phi^{n_k} = \phi$. Now let us prove that $\phi \in L_N$. For clarity we replace n_k by n from now on. The convergence as n goes to $+\infty$ and the continuity of the paths entail

$$\sup_{n\in\mathbb{N}}\sup_{t\in\mathbb{T}}\big(|\phi_t^n| + |\phi_t|\big) < +\infty,$$

such that the paths lie in compact subsets of $\mathbb{R}^d$ and **H2** asserts that uniform continuity of the coefficients b and σ hold. Therefore they admit continuous moduli of continuity that we respectively name ρ_b and ρ_σ. Using Cauchy-Schwarz inequality and **H3** we get that for all $t \in \mathbb{T}$:

$$\left|\int_0^t K(t-s)b(s,\phi_s^n)\mathrm{d}s - \int_0^t K(t-s)b(s,\phi_s)\mathrm{d}s\right| \leq \|K\|_{L^1}\big\|\rho_b\big(|\phi_\cdot^n - \phi_\cdot|\big)\big\|_{\mathbb{T}}$$

$$\begin{aligned}
&\left|\int_0^t K(t-s)\sigma(s,\phi_s^n)v_s^n\mathrm{d}s - \int_0^t K(t-s)\sigma(s,\phi_s)v_s\mathrm{d}s\right| \\
&\quad\leq \int_0^t \big|K(t-s)\big(\sigma(s,\phi_s^n) - \sigma(s,\phi_s)\big)v_s^n\big|\,\mathrm{d}s + \int_0^t |K(t-s)\sigma(s,\phi_s)(v_s^n - v_s)|\,\mathrm{d}s \\
&\quad\leq \|K\|_2\,\|v^n\|_2\,\|\rho_\sigma(|\phi_\cdot^n - \phi_\cdot|)\|_{\mathbb{T}} + \|\sigma(\phi)\|_{\mathbb{T}}\int_0^t |K(t-s)(v_s^n - v_s)|\,\mathrm{d}s,
\end{aligned}$$

which converges towards zero as n tends to infinity. Therefore, for all $t \in \mathbb{T}$

$$\begin{aligned}\phi_t = \lim_{n\uparrow\infty} \phi_t^n &= \lim_{n\uparrow\infty} \left(x_0 + \int_0^t K(t-s)\Big[b(s,\phi_s^n) + \sigma(s,\phi_s^n)v_s^n\Big]\mathrm{d}s \right) \\ &= x_0 + \int_0^t K(t-s)\Big[b(s,\phi_s) + \sigma(s,\phi_s)v_s\Big]\mathrm{d}s,\end{aligned}$$

so that $\phi \in L_N$ since $v \in \mathcal{S}_{2N}$, which concludes the proof of the closure and therefore of the compactness of L_N.

Appendix B. Technical moderate deviations proofs

B.1. **MDP moment bounds: Proof of Lemma 3.17.** Let $p \geq 2$, $N > 0$, $v \in \mathcal{A}_N$, $\varepsilon > 0$ and $t \in \mathbb{T}$. Starting from (3.15), we proceed as in (A.1) and use Cauchy-Schwarz and BDG inequalities to obtain

$$\begin{aligned}\mathbb{E}\left[|\eta_t^{\varepsilon,v}|^p\right] \leq\ & 5^{p-1}\frac{|X_0^\varepsilon - x_0|^p}{(\vartheta_\varepsilon h_\varepsilon)^p} \\ &+ 5^{p-1}\mathbb{E}\left|\int_0^t K(t-s)\frac{b_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big) - b\big(s,\Theta_s^{\varepsilon,v}\big)}{\vartheta_\varepsilon h_\varepsilon}\mathrm{d}s\right|^p \\ &+ 5^{p-1}\mathbb{E}\left|\int_0^t K(t-s)\frac{b\big(s,\Theta_s^{\varepsilon,v}\big) - b(s,\overline{X}_s)}{\vartheta_\varepsilon h_\varepsilon}\mathrm{d}s\right|^p \\ &+ 5^{p-1}N^{p/2}\mathbb{E}\left[\left(\int_0^t \left|K(t-s)\sigma_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big)\right|^2\mathrm{d}s\right)^{p/2}\right] \\ &+ \frac{5^{p-1}C_p}{h_\varepsilon^p}\mathbb{E}\left[\left(\int_0^t \left|K(t-s)\sigma_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big)\right|^2\mathrm{d}s\right)^{p/2}\right].\end{aligned} \tag{B.1}$$

The first term converges by **H7** and is thus bounded. Notice that **H8** entails

$$\mathbb{E}\left|\int_0^t K(t-s)\frac{b_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big) - b\big(s,\Theta_s^{\varepsilon,v}\big)}{\vartheta_\varepsilon h_\varepsilon}\mathrm{d}s\right|^p \leq \left(\frac{\nu_\varepsilon}{\vartheta_\varepsilon h_\varepsilon}\right)^p \mathbb{E}\left|\int_0^t K(t-s)\Xi\big(\Theta_s^{\varepsilon,v}\big)\mathrm{d}s\right|^p,$$

which is bounded because $\nu_\varepsilon(\vartheta_\varepsilon h_\varepsilon)^{-1}$ tends to zero and

$$\mathbb{E}\left|\int_0^t K(t-s)\Xi\big(\Theta_s^{\varepsilon,v}\big)\mathrm{d}s\right|^p \leq \|K\|_2^p\, t^{p/2-1}\sup_{s\leq T}\mathbb{E}\Big[\left|\Xi\big(\Theta_s^{\varepsilon,v}\big)\right|^p\Big] \leq C_1,$$

by Cauchy-Schwarz and Jensen's inequalities and the bound (3.17), where C_1 is a positive constant that does not depend on ε. Since $b(s,\cdot)$ is globally Lipschitz continuous uniformly on $\mathbb{T}$, there exists $C_b > 0$ such that for all $s \in \mathbb{T}$:

$$\left|b\big(s,\overline{X}_s + \vartheta_\varepsilon h_\varepsilon \eta_s^{\varepsilon,v}\big) - b(s,\overline{X}_s)\right| \leq C_b\,\vartheta_\varepsilon h_\varepsilon\,|\eta_s^{\varepsilon,v}|\ . \tag{B.2}$$

Therefore, using Cauchy-Schwarz and Jensen's inequalities in the same way as (A.2)

$$\begin{aligned}\mathbb{E}\left|\int_0^t K(t-s)\frac{b\big(s,\Theta_s^{\varepsilon,v}\big) - b(s,\overline{X}_s)}{\vartheta_\varepsilon h_\varepsilon}\mathrm{d}s\right|^p &\leq \mathbb{E}\left|\int_0^t K(t-s)C_b\ |\eta_s^{\varepsilon,v}|\ \mathrm{d}s\right|^p \\ &\leq C_b^p t^{p/2}\,\|K\|_2^p\int_0^t |K(t-s)|^2\,\mathbb{E}\big[\,|\eta_s^{\varepsilon,v}|^p\ \big]\mathrm{d}s.\end{aligned}$$

The last two terms of (B.1) are also uniformly bounded in n and ε because, similarly to (A.3),

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_0^t \left|K(t-s)\sigma_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big)\right|^2 \mathrm{d}s\right)^{p/2}\right]\\
&\le \mathbb{E}\left[\left(\int_0^t |K(t-s)|^2 \left|\sigma_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big)\right|^p \mathrm{d}s\right)\left(\int_0^t |K(t-s)|^2 \mathrm{d}s\right)^{\frac{p-2}{2}}\right]\\
&\le \|K\|_2^p\, C_L\big(1+\sup_{t\in\mathbb{T}}\mathbb{E}\left[|\Theta_t^{\varepsilon,v}|^p\right]\big),
\end{aligned}\tag{B.3}$$

by Hölder's inequality and the linear growth condition **H3a**. If the latter fails we rely on **H3b** and the same calculations as in (A.5) to obtain a similar bound.

Overall this results in the existence of a constant $C_2>0$ independent of ε, v, t such that

$$\mathbb{E}\left[|\eta_t^{\varepsilon,v}|^p\right] \le C_2 + C_2\int_0^t |K(t-s)|^2\,\mathbb{E}\left[|\eta_s^{\varepsilon,v}|^p\right]\mathrm{d}s, \qquad \text{for all } t\in\mathbb{T},$$

and Lemma A.1 yields the uniformly bound in ε and t.

B.2. **MDP tightness: Proof of Lemma 3.18.** Let $p>2\vee 2/\gamma$, $N>0$, $\varepsilon>0$, and $0\le s<t\le T$. We proceed as in Lemma 3.12; starting from (3.15), applying consecutively **H8**, then (B.2), Cauchy-Schwarz and BDG inequalities we obtain

$$\begin{aligned}
&\mathbb{E}\left[\left|\eta_t^{\varepsilon,v^\varepsilon}-\eta_s^{\varepsilon,v^\varepsilon}\right|^p\right]\\
\le& 8^{p-1}\mathbb{E}\left|\frac{\nu_\varepsilon}{\vartheta_\varepsilon h_\varepsilon}\int_0^s \big(K(t-u)-K(s-u)\big)\Xi\big(\Theta_u^{\varepsilon,v^\varepsilon}\big)\mathrm{d}u\right|^p\\
&+8^{p-1}\mathbb{E}\left|\frac{\nu_\varepsilon}{\vartheta_\varepsilon h_\varepsilon}\int_s^t K(t-u)\Xi\big(\Theta_u^{\varepsilon,v^\varepsilon}\big)\mathrm{d}u\right|^p\\
&+8^{p-1}\mathbb{E}\left|\int_0^s \big(K(t-u)-K(s-u)\big)C_b\left|\eta_u^{\varepsilon,v^\varepsilon}\right|\mathrm{d}u\right|^p\\
&+8^{p-1}\mathbb{E}\left|\int_s^t K(t-u)C_b\left|\eta_u^{\varepsilon,v^\varepsilon}\right|\mathrm{d}u\right|^p\\
&+8^{p-1}N^{p/2}\mathbb{E}\left[\left(\int_0^s \left|\big(K(t-u)-K(s-u)\big)\sigma_\varepsilon\big(u,\Theta_u^{\varepsilon,v^\varepsilon}\big)\right|^2\mathrm{d}u\right)^{p/2}\right]\\
&+8^{p-1}N^{p/2}\mathbb{E}\left[\left(\int_s^t \left|K(t-u)\sigma_\varepsilon\big(u,\Theta_u^{\varepsilon,v^\varepsilon}\big)\right|^2\mathrm{d}u\right)^{p/2}\right]\\
&+\frac{8^{p-1}C_p}{h_\varepsilon^p}\mathbb{E}\left[\left(\int_0^s \left|\big(K(t-u)-K(s-u)\big)\sigma_\varepsilon\big(u,\Theta_u^{\varepsilon,v^\varepsilon}\big)\right|^2\mathrm{d}u\right)^{p/2}\right]\\
&+\frac{8^{p-1}C_p}{h_\varepsilon^p}\mathbb{E}\left[\left(\int_s^t \left|K(t-u)\sigma_\varepsilon\big(u,\Theta_u^{\varepsilon,v^\varepsilon}\big)\right|^2\mathrm{d}u\right)^{p/2}\right].
\end{aligned}$$

In the first four terms, Cauchy-Schwarz inequality allows to separate the kernels from the random variables. For the last four terms, analogous calculations to (A.3) achieve a similar separation of kernels and random variables. Then linear growth or (3.6) and bounds on $\sup_{t\in\mathbb{T},\varepsilon>0}\mathbb{E}\left[\left|\Xi\big(\Theta_t^{\varepsilon,v^\varepsilon}\big)\right|^p\right]$

and $\sup_{t\in\mathbb{T},\varepsilon>0}\mathbb{E}\left[\left|\eta_t^{\varepsilon,v^\varepsilon}\right|^p\right]$ lead to the existence of $C_1>0$ independent of t and ε such that

$$\begin{aligned}\mathbb{E}\left[\left|\eta_t^{\varepsilon,v^\varepsilon}-\eta_s^{\varepsilon,v^\varepsilon}\right|^p\right] &\le C_1\left(\left|\int_0^s\big(K(t-u)-K(s-u)\big)^2\mathrm{d}u\right|^{p/2}+\left|\int_s^t K(t-u)^2\mathrm{d}u\right|^{p/2}\right)\\ &= C_1\left(\left|\int_0^s\big(K(u+t-s)-K(u)\big)^2\mathrm{d}u\right|^{p/2}+\left|\int_0^{t-s} K(u)^2\mathrm{d}u\right|^{p/2}\right).\end{aligned}$$

Hence Assumption 2.3 yields the existence of a constant $C_2>0$ such that

$$\mathbb{E}\left[\left|\eta_t^{\varepsilon,v^\varepsilon}-\eta_s^{\varepsilon,v^\varepsilon}\right|^p\right]\le C_2(t-s)^{\gamma p/2}.$$

Then the Kolmogorov continuity theorem asserts that $\eta^{\varepsilon,v^\varepsilon}$ admits a version which is Hölder continuous on $\mathbb{T}$ of any order $\alpha<\gamma/2-1/p$, uniformly in $\varepsilon>0$ and which satisfies (3.19). Furthermore, Aldous theorem [10, Theorem 16.10] states that the sequence $\{\eta^{\varepsilon,v^\varepsilon}\}_{\varepsilon>0}$ is tight.

B.3. **MDP weak convergence: Proof of Lemma 3.19.** We have shown in Lemma 3.18 that for any subsequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$, $\{\eta^{\varepsilon_k,v^{\varepsilon_k}}\}_{k\in\mathbb{N}}$ and $\{v^{\varepsilon_k}\}_{k\in\mathbb{N}}$ are tight as families of random variables with values in $\mathcal{W}^d$ and $\mathcal{S}_N$ respectively. Hence there exists a subsubsequence, denoted hereafter $\{\eta^k,v^k\}$, that converges weakly to some $\mathcal{W}^d\times\mathcal{S}_N$-valued limit (η^0,v) in a possibly different probability space $(\Omega^0,\mathcal{F}^0,\mathbb{P}^0)$ as n tends to $+\infty$. We also denote $\varepsilon_k,b_k,\sigma_k,X_0^k,\Theta^k$ along this subsequence.

We follow the same method as in the LDP case which comes from [22]. For all $t\in[0,T]$, let $\Psi_t:\mathcal{S}_N\times\mathcal{W}^d\to\mathbb{R}$ such that

$$\Psi_t(f,\omega):=\left|\omega_t-\int_0^t K(t-s)\left[\nabla b(s,\overline{X}_s)\omega_s+\sigma(s,\overline{X}_s)f_s\right]\mathrm{d}s\right|\wedge 1.$$

Clearly, Ψ_t is bounded and one can show its continuity along the same lines as in the LDP proof but even simpler because it is linear. Therefore

$$\lim_{k\uparrow\infty}\mathbb{E}\left[\Psi_t(v^k,\eta^k)\right]=\mathbb{E}^0\left[\Psi_t(v,\eta^0)\right],$$

and we prove that the left-hand side is actually equal to zero. By **H5** and Taylor's formula there exists a family of $\mathbb{R}^d$-valued stochastic processes $\{R^\varepsilon\}_{\varepsilon>0}$ such that

$$\frac{b\big(s,\overline{X}_s+\vartheta_\varepsilon h_\varepsilon\eta_s^{\varepsilon,v}\big)-b(s,\overline{X}_s)}{\vartheta_\varepsilon h_\varepsilon}=\nabla b\big(s,\overline{X}_s\big)\eta_s^{\varepsilon,v}+R^\varepsilon(s),\tag{B.4}$$

and a constant $C_R>0$ such that

$$|R^\varepsilon(s)|\le C_R\vartheta_\varepsilon h_\varepsilon\left|\eta_s^{\varepsilon,v}\right|^2.$$

We recall that $\left\|\nabla b(\cdot,\overline{X})\right\|_{\mathbb{T}}$ is finite from **H5** and observe that

$$\mathbb{E}\big[\,|R^\varepsilon(u)|^p\,\big]\le\big(C_R\vartheta_\varepsilon h_\varepsilon\big)^p\mathbb{E}\left[\left|\eta_s^{\varepsilon,v^\varepsilon}\right|^{2p}\right]<\infty.\tag{B.5}$$

Again starting from (3.15), we use **H8**, the Taylor estimate (B.4) and Itô isometry to get

$$
\begin{aligned}
\mathbb{E}\left[\Psi_t(\eta^k, v^k)^2\right] \leq & \frac{5\left|X_0^k - x_0\right|^2}{\vartheta_{\varepsilon_k}^2 h_{\varepsilon_k}^2} \\
& + 5\left(\frac{\nu_{\varepsilon_k}}{\vartheta_{\varepsilon_k} h_{\varepsilon_k}}\right)^2 \mathbb{E}\left|\int_0^t K(t-s)\Xi(\Theta_s^k)\mathrm{d}s\right|^2 \\
& + 5\mathbb{E}\left|\int_0^t K(t-s)R^\varepsilon(s)\mathrm{d}s\right|^2 \\
& + 5\mathbb{E}\left|\int_0^t K(t-s)\left[\sigma_k(s,\Theta_s^k) - \sigma(s,\overline{X}_s)\right]v_s^k\mathrm{d}s\right|^2 \\
& + \frac{5}{h_{\varepsilon_k}^2}\mathbb{E}\left[\int_0^t \left|K(t-s)\sigma_k(s,\Theta_s^k)\right|^2\mathrm{d}s\right] \\
=: & 5\big(\mathrm{I}_k + \mathrm{II}_k + \mathrm{III}_k + \mathrm{IV}_k + h_{\varepsilon_k}^{-2}\mathrm{V}_k\big).
\end{aligned}
$$

H7 and **H8** tell us that $\mathrm{I}_k+\mathrm{II}_k$ tends to zero as ε goes to zero while an application of Cauchy-Schwarz inequality and the bound (B.5) yields the same conclusion for III_k.

To deal with IV_k, recall that Θ^k converges in distribution towards $\overline{X}$ as k tends to infinity. Since $\overline{X}$ is deterministic, the convergence actually takes place in probability, with respect to the topology of uniform convergence. Moreover Θ^k is uniformly bounded in L^r for all $r > p$, thus the family $\{\left|\Theta^k\right|^p\}_{k\geq 0}$ is uniformly integrable. Therefore the convergence also occurs with respect to the L^p-norm.

The modulus of continuity of σ is only available on compact sets of $\mathbb{T}\times\mathbb{R}^d$ so we define a constant $M > \left\|\overline{X}\right\|_{\mathbb{T}}$ and introduce the following sets, for each $k\in\mathbb{N}$:

$$
E_k := \left\{\omega\in\Omega, \left\|\Theta^k(\omega) - \overline{X}\right\|_{\mathbb{T}} \leq M\right\},
$$

with the observation that $\lim_{k\uparrow\infty}\mathbb{P}(E_k) = 1$ thanks to the previous argument. Since $\overline{X}$ is uniformly bounded, $\left|\Theta_t^k(\omega)\right| \leq 2M$ for all $t\in\mathbb{T}, \omega\in E_k$, $k\geq 0$. Therefore using Cauchy-Schwarz inequality,

$$
\begin{aligned}
\mathbb{E}\left[\mathrm{IV}_k\right] \leq & 2N\mathbb{E}\left[\int_0^t |K(t-s)|^2\left|\sigma_k(s,\Theta_s^k) - \sigma(s,\Theta_s^k)\right|^2\mathrm{d}s(\mathbb{1}_{E_k} + \mathbb{1}_{E_k^c})\right] \\
& + 2N\mathbb{E}\left[\int_0^t |K(t-s)|^2\left|\sigma(s,\Theta_s^k) - \sigma(s,\overline{X}_s)\right|^2\mathrm{d}s(\mathbb{1}_{E_k} + \mathbb{1}_{E_k^c})\right],
\end{aligned}
$$

where we will use the localisation to obtain convergence in the first term and Hölder continuity in the second. Let us assume for the moment that linear growth **H3a** holds. We use that Θ^k is uniformly bounded by $2M$ in E_k and linear growth for both σ_k and σ to obtain

$$
\begin{aligned}
\sup_{s\in\mathbb{T}}\mathbb{E}\left[\left|\sigma_k(s,\Theta_s^k) - \sigma(s,\Theta_s^k)\right|^2\right] = & \sup_{s\in\mathbb{T}}\mathbb{E}\left[\left|\sigma_k(s,\Theta_s^k) - \sigma(s,\Theta_s^k)\right|^2\mathbb{1}_{E_k}\right] + \sup_{s\in\mathbb{T}}\mathbb{E}\left[\left|\sigma_k(s,\Theta_s^k) - \sigma(s,\Theta_s^k)\right|^2\mathbb{1}_{E_k^c}\right] \\
\leq & \|\sigma_k - \sigma\|_{2M}^2 + \sup_{s\in\mathbb{T}}\mathbb{E}\left[\mathbb{1}_{E_k^c}C_L^2\big(2 + 2|\Theta_s^k|\big)^2\right], \qquad \text{(B.6)}
\end{aligned}
$$

which tends to zero as k goes to infinity because of **H2** for the first term and because $\mathbb{P}(\Omega\setminus E_k)$ tends to zero for the second. Moreover, by **H6**, σ is locally Hölder continuous thus there exist $\delta, C_{2M} > 0$ such that

$$
\sup_{s\in\mathbb{T}}\mathbb{E}\left[\mathbb{1}_{E_k}\left|\sigma(s,\Theta_s^k) - \sigma(s,\overline{X}_s)\right|^2\right] \leq \sup_{s\in\mathbb{T}}\mathbb{E}\left[\mathbb{1}_{E_k}C_{2M}\left|\Theta_s^k - \overline{X}_s\right|^{2\delta}\right],
$$

which tends to zero. Finally, linear growth leads to

$$
\sup_{s\in\mathbb{T}}\mathbb{E}\left[\mathbb{1}_{E_k^c}\left|\sigma(s,\Theta_s^k) - \sigma(s,\overline{X}_s)\right|^2\right] \leq \mathbb{E}\left[\mathbb{1}_{E_k^c}C_L^2\big(2 + |\Theta_s^k| + \left|\overline{X}_s\right|\big)^2\right], \qquad \text{(B.7)}
$$

which also tends to zero because $\mathbb{P}(\Omega \setminus E_k)$ tends to zero. If only **H3b** with Assumption 3.6 hold then a different bound depending on (2.2) would replace those in (B.6) and (B.7), by noticing that $\overline{X} = \mathcal{G}^0(0)$. In both cases the above estimates tend to zero as k tends to infinity, hence $\mathbb{E}[\mathrm{IV}_k]$ converges towards zero. Finally, $\{\mathrm{V}_k\}_{k\in\mathbb{N}}$ is uniformly bounded across $k \geq 0$ as (B.3) shows, thus $h_{\varepsilon_k}^{-p}\mathrm{V}_k$ tends to zero. We have proved that

$$\lim_{k\uparrow\infty} \mathbb{E}\left[\Psi_t(v^k, \eta^k)\right] = 0,$$

and this entails that the limit η^0 satisfies (3.16) $\mathbb{P}^0$-almost surely, for all $t \in \mathbb{T}$. Since η^0 has continuous paths, this holds for all $t \in \mathbb{T}$, $\mathbb{P}^0$-almost surely. Since the solution is unique we conclude that $\eta^0 = \psi$. Every subsequence has a subsequence for which this convergence holds therefore $\{\eta^{\varepsilon,v^\varepsilon}\}_{\varepsilon>0}$ converges weakly towards ψ as ε goes to zero.

References

[1] E. Abi Jaber, M. Larsson, and S. Pulido. Affine Volterra processes. *Annals of Applied Probability*, 29(5):3155–3200, 2017.

[2] E. Alòs, J. A. León, and J. Vives. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics*, 11(4):571–589, 2007.

[3] O. E. Barndorff-Nielsen, M. S. Pakkanen, and J. Schmiegel. Assessing relative volatility/intermittency/energy dissipation. *Electronic Journal of Statistics*, 8(2):1996–2021, 2014.

[4] O. E. Barndorff-Nielsen and J. Schmiegel. Brownian semistationary processes and volatility/intermittency. *Radon Series Comp. Appl. Math*, 8:1–26, 2009.

[5] C. Bayer, P. K. Friz, P. Gassiat, J. Martin, and B. Stemper. A regularity structure for rough volatility. *Mathematical Finance*, pages 1–51, 2019.

[6] C. Bayer, P. K. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[7] C. Bayer, P. K. Friz, A. Gulisashvili, B. Horvath, and B. Stemper. Short-time near-the-money skew in rough fractional volatility models. Preprint, arXiv:1703.05132, 2017.

[8] M. Bennedsen, A. Lunde, and M. Pakkanen. Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, 21:931–965, 2017.

[9] S. Bhamidi, A. Budhiraja, P. Dupuis, and R. Wu. Rare event asymptotics for exploration processes for random graphs. Preprint, arXiv:1912.04714, 2019.

[10] P. Billingsey. *Convergence of probability measures*. Wiley series in probability and statistics, 1999.

[11] M. Boué and P. Dupuis. A variational representation for certain functionals of Brownian motion. *The Annals of Probability*, 26(4):1641–1659, 1998.

[12] A. Budhiraja, J. Chen, and P. Dupuis. Large deviations for stochastic partial differential equations driven by a Poisson random measure. *Stochastics Processes and Applications*, 123(2):523–560, 2013.

[13] A. Budhiraja and P. Dupuis. A variational representation for positive functionals of infinite dimensional Brownian motion. *Probability and Mathematical Statistics*, 20(1), 2001.

[14] A. Budhiraja and P. Dupuis. *Analysis and Approximation of Rare Events Representations and Weak Convergence Methods*. Springer, 2019.

[15] A. Budhiraja, P. Dupuis, and A. Ganguly. Moderate deviations principles for stochastic differential equations with jumps. *The Annals of Probability*, 44(3):1723–1775, 2016.

[16] A. Budhiraja, P. Dupuis, and A. Ganguly. Large deviations for small noise diffusions in a fast Markovian environment. *Electronic Journal of Probability*, 23(112):1–33, 2018.

[17] A. Budhiraja, P. Dupuis, and V. Maroulas. Large deviations for infinite dimensional stochastic dynamical systems. *The Annals of Probability*, 36(4):1390–1420, 2008.

[18] A. Budhiraja, P. Dupuis, and V. Maroulas. Variational representations for continuous time processes. *Annals of the Institute Henri Poincaré Probability and Statistics*, 47(3):725–747, 2011.

[19] A. Budhiraja, E. Friedlander, and R. Wu. Many-server asymptotics for join-the-shortest-queue: Large deviations and rare events. Preprint, arXiv:1904.04938, 2020.

[20] M. Cellupica and B. Pacchiarotti. Pathwise asymptotics for Volterra type rough volatility models. *Preprint, arXiv:1902.05896*, 2019.

[21] L. Chevillard. Regularized fractional Ornstein-Uhlenbeck processes and their relevance to the modeling of fluid turbulence. *Physical Review E*, 97, 2017.

[22] F. Chiarini and M. Fischer. On large deviations for small noise Itô processes. *Advances in Applied Probability*, 46(4):1126–1147, 2014.

[23] A. Chronopoulou and F. G. Viens. Estimation and pricing under long-memory stochastic volatility. *Annals of Finance*, 8:379–403, 2012.

[24] A. Chronopoulou and F. G. Viens. Stochastic volatility and option pricing with long-memory in discrete and continuous time. *Quantitative Finance*, 12:635–649, 2012.

[25] E. Comte, L. Coutin, and E. Renault. Affine fractional stochastic volatility models with application to option pricing. *Annals of Finance*, 8:337–378, 2012.

[26] E. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8:291–323, 1998.

[27] G. Conforti, S. De Marco, and J.-D. Deuschel. On small-noise equations with degenerate limiting system arising from volatility models. *In: Friz P., Gatheral J., Gulisashvili A., Jacquier A., Teichmann J. (eds) Large Deviations and Asymptotic Methods in Finance. Springer Proceedings in Mathematics & Statistics*, 2015.

[28] J. M. Corcuera, H. E., M. S. Pakkanen, and M. Podolskij. Asymptotic theory for Brownian semi-stationary processes with application to turbulence. *Stochastic Processes and their Applications*, 123(7):2552–2574, 2017.

[29] L. Coutin and L. Decreusefond. Abstract nonlinear filtering theory in the presence of fractional Brownian motion. *The Annals of Applied Probability*, 9(4):1058–1090, 1999.

[30] L. Coutin and L. Decreusefond. Volterra differential equations with singular kernels. *Proceedings of the Workshop on Mathematical Physics and Stochastic Analysis*, 2000.

[31] J. Cuchiero and J. Teichmann. Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case. Preprint, arXiv:1804.10450, 2018.

[32] L. Decreusefond. Regularity properties of some stochastic Volterra integrals with singular kernels. *Potential Analysis*, 16:139–149, 2002.

[33] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer-Verlag Berlin Heidelberg, 1998.

[34] J.-D. Deuschel and D. W. Stroock. *Large Deviations*. Academic Press Inc., 1989.

[35] C. Donati-Martin, A. Rouault, M. Yor, and M. Zani. Large deviations for squares of Bessel and Ornstein-Uhlenbeck processes. *Probability Theory and Related Fields*, 129:261–289, 2004.

[36] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley, 1997.

[37] P. Dupuis and K. Spiliopoulos. Large deviations for multiscale diffusions via weak convergence methods. *Stochastic Processes and their Applications*, 122(4):1947–1987, 2012.

[38] P. Dupuis, K. Spiliopoulos, and H. Wang. Importance sampling for multiscale diffusions. *SIAM Journal on Multiscale Modeling and Simulation*, 10(1):1–27, 2012.

[39] O. El Euch, M. Fukasawa, and M. Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, 22(2):241–280, 2018.

[40] O. El Euch and M. Rosenbaum. Perfect hedging in rough Heston models. *The Annals of Applied Probability*, 28(6):3813–3856, 2018.

[41] O. El Euch and M. Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, 29(1):3–38, 2019.

[42] M. Forde, S. Gerhold, and B. Smith. Small-time and large-time smile behaviour for the rough Heston model. Preprint, arXiv:1906.09034, 2019.

[43] M. Forde and H. Zhang. Asymptotics for rough stochastic volatility models. *SIAM J. Financial Math.*, 8(1):114–145, 2017.

[44] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer-Verlag New York, 1984.

[45] P. K. Friz, P. Gassiat, and P. Pigato. Precise asymptotics: robust stochastic volatility models. Preprint, arXiv:1811.00267, 2018.

[46] P. K. Friz, S. Gerhold, and A. Pinter. Option pricing in the moderate deviations regime. *Mathematical Finance*, 28(3), 2018.

[47] M. Fukasawa. Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics*, 15:635–654, 2011.

[48] M. Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, 17(2):189–198, 2017.

[49] K. Gao and R. Lee. Asymptotics of implied volatility to arbitrary order. *Finance and Stochastics*, 18(2):349–392, 2014.

[50] P. Gassiat. On the martingale property in the rough Bergomi model. *Electronic Communications in Probability*, 24(33), 2019.

[51] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

[52] J. Gatheral and M. Keller-Ressel. Affine forward variance models. *Finance and Stochastics*, 23:501–533, 2019.

[53] G. Gripenberg, S.-O. Londen, and O. Staffans. *Volterra integral and functional equations*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1990.

[54] H. Guennon, A. Jacquier, P. Roome, and F. Shi. Asymptotic behavior of the fractional Heston model. *SIAM Journal on Financial Mathematics*, 9(3):1017–1045, 2018.

[55] A. Gulisashvili. Large deviation principle for Volterra type fractional stochastic volatility models. *SIAM Journal on Financial Mathematics*, 9(3):1102–1136, 2018.

[56] B. Horvath, A. Jacquier, and C. Lacombe. Asymptotic behaviour of randomised fractional volatility models. *Journal of Applied Probability*, 56(2):496–523, 2019.

[57] B. Horvath, A. Jacquier, and A. Muguruza. Functional central limit theorems for rough volatility. Preprint, arXiv:1711.03078, 2017.

[58] A. Jacquier, M. Pakkanen, and H. Stone. Pathwise large deviations for the rough Bergomi model. *Journal of Applied Probability*, 55(4):1078–1092, 2018.

[59] A. Jacquier and K. Spiliopoulos. Pathwise moderate deviations for option pricing. *Mathematical Finance*, 30(2):426–463, 2020.

[60] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag New York, 1998.

[61] C. Lacombe, A. Muguruza, and H. Stone. Asymptotics for volatility derivations in multi-factor rough volatility models. Preprint, arXiv:1903.02833, 2019.

[62] Y. Li, R. Wang, N. Yao, and S. Zhang. A moderate deviation principle for stochastic Volterra equation. *Statistics and Probability Letters*, 122, 2017.

[63] N. Marie. A pathwise fractional one compartment intra-veinous bolus model. *International Journal of Statistics and Probability*, 3, 2014.

[64] R. McCrickerd and M. Pakkanen. Turbocharging Monte–Carlo pricing for the rough Bergomi model. *Quantitative Finance*, 18(11):1877–1886, 2018.

[65] M. R. Morse and K. Spiliopoulos. Moderate deviations principle for systems of slow-fast diffusions. *Asymptotic Analysis*, 105:97–135, 2017.

[66] M. R. Morse and K. Spiliopoulos. Importance sampling for slow-fast diffusions based on moderate deviations. *SIAM Journal on Multiscale Modeling and Simulation*, 18(1):315–350, 2018.

[67] L. Mytnik and T. S. Salisbury. Uniqueness for Volterra-type stochastic integral equations. Preprint arXiv:1502.05513, 2015.

[68] D. Nualart and C. Rovira. Large deviations for stochastic Volterra equations. *Bernoulli*, 6(2):339–355, 2000.

[69] F. P., G. J., G. A., J. A., and T. J. *Large Deviations and Asymptotic Methods in Finance*. Springer Proceedings in Mathematics and Statistics, 2015.

[70] S. Robertson. Sample path large deviations and optimal importance sampling for stochastic volatility models. *Stochastic Processes and their Applications*, 120(1):66–83, 2010.

[71] C. Rovira and M. Sanz-Solé. Large deviations for stochastic Volterra equations in the plane. *Potential Analysis*, 12:359–383, 2000.

[72] K. Spiliopoulos. Large deviations and importance sampling for systems of slow-fast motion. *Applied Mathematics and Optimization*, 67:123–161, 2013.

[73] E. Stein and J. Stein. Stock price distributions with stochastic volatility - an analytic approach. *Review of Financial studies*, 4:727–752, 1991.

[74] T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11(1):155–167, 1971.

Department of Mathematics, Imperial College London, and Alan Turing Institute
Email address: a.jacquier@imperial.ac.uk

Department of Mathematics, Imperial College London
Email address: a.pannier17@imperial.ac.uk