J. Appl. Prob. **56,** 496–523 (2019)
*doi:*10.1017/jpr.2019.27
© *Applied Probability Trust* 2019

ASYMPTOTIC BEHAVIOUR OF RANDOMISED FRACTIONAL VOLATILITY MODELS

BLANKA HORVATH,* ** *King's College London*
ANTOINE JACQUIER,*** **** ***** *Imperial College London and the Alan Turing Institute* AND
CHLOÉ LACOMBE,*** ******

Abstract

We study the asymptotic behaviour of a class of small-noise diffusions driven by fractional Brownian motion, with random starting points. Different scalings allow for different asymptotic properties of the process (small-time and tail behaviours in particular). In order to do so, we extend some results on sample path large deviations for such diffusions. As an application, we show how these results characterise the small-time and tail estimates of the implied volatility for rough volatility models, recently proposed in mathematical finance.

Keywords: Rough volatility; large deviations; implied volatility asymptotics

2010 Mathematics Subject Classification: Primary 41A60
Secondary 60F10; 60G15; 60G22

1. Introduction

Large deviations are used extensively in physics (thermodynamics, statistical mechanics) as well as in mathematics (information theory, stochastic analysis, mathematical finance) to estimate the exponential decay of probability measures of rare events. Varadhan [52], Schilder [49], and Freidlin and Wentzell [25] proved, with different degrees of generality, large deviations principles (LDPs; in $\mathbb{R}^n$ and on path space) for solutions of stochastic differential equations with small noise, and the monographs by Dembo and Zeitouni [16] and Deuschel-Stroock [19] provide a precise account of those advances (at least up to the mid-1990s). In the past decade, this set of techniques and results has been adopted by the mathematical finance community: finite-dimensional large deviations (in the sense of Gärtner-Ellis) have been used to prove small- and large-time asymptotics of implied volatility in affine models [22, 34], sample-path LDP (à la Freidlin–Wentzell [25]) have proved efficient for determining importance sampling changes of probability [31, 32, 48], and heat kernel expansions (following Ben Arous [9] and Bismut [11]), have led to a general understanding of small-time and tail behaviour of multi-dimensional diffusions [5, 6, 17, 18]. These asymptotics

Received 6 June 2018; revision received 20 December 2018.
* Postal address: Department of Mathematics, King's College London, London WC2R 2LS, UK.
** Email address: blanka.horvath@kcl.ac.uk
*** Postal address: Department of Mathematics, Imperial College London, London SW7 2AZ, UK.
**** Email address: a.jacquier@imperial.ac.uk
***** Postal address: Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, UK.
****** Email address: chloe.lacombe14@ic.ac.uk

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

have overall provided a deeper understanding of the behaviour of models, and, ultimately, allow for better calibration of real data; a general overview can be found in [46].

Motivated by financial applications, we derive here asymptotic small-time and tail behaviours of the solution to a generalised version of the Stein–Stein stochastic volatility model, originally proposed in [50, 51]. We in particular consider two important (in light of the recent trends in the literature-proposed models) extensions: (i) the stochastic differential equation (SDE) driving the instantaneous volatility process is started from a random distribution; this so-called 'randomised' type of model was recently proposed in [38, 39, 41], in particular to understand the behaviour of the so-called 'forward volatility'; (ii) the volatility process is driven by a fractional Brownian motion. Fractional stochastic volatility models, originally proposed by Comte and Renault [14, 15] with $H \in (1/2, 1)$, have recently been extended to the case $H \in (0, 1/2)$, and a recent flourishing activity in this area [3, 7, 26, 30, 35] has established these models as the go-to standards for estimation and calibration.

The original motivation behind randomisation of the initial starting point is rooted in financial practice, where only the initial value of the stock price process is observed directly and the instantaneous value of volatility is subject to calibration. The effect of randomisation of the initial volatility on the implied volatility surface was explored by Jacquier and Roome [37] in a simple 'random environment' setting, where the volatility component was assumed to follow constant elasticity of variance dynamics. Their results provide progress on both theoretical and practical levels: they solve a practical modelling problem in a simple tractable setting and at the same time raise awareness of the potential prowess of applying random evolution equations for financial modelling. In this paper we follow up on this direction and blend more involved approaches (proposed by [42, 43]) from the literature around random environment and random evolution equations into our financial model, where randomness also appears in the drift and diffusion coefficients of the process. On the practical level, independently from the results of [37], Mechkov [41] goes a step further in endorsing the idea of randomising the initial volatility and makes a strong case to move away from modelling hidden variables (such as stochastic volatility) in the traditional way. He argues that starting the volatility from a fixed starting point heavily underestimates the effect of the hidden variable on the slope of the implied volatility smile, and therefore 'hot start' volatility models (with random starting point) significantly outperform traditional ones. Indeed, both randomised models [37, 41] produce the desired explosion in the smile at short maturities. Jacquier and Shi [39] develop this further by providing a precise link between the rate of explosion of implied volatilities on the short end and the tail distribution of the initial distribution of the volatility process in a 'randomised' Heston model. These outputs confirm that stochastic volatility models with a random starting point constitute a class of counterexamples to the long-standing belief, formulated by Gatheral [29, Chapter 5], that jumps in the stock price process are needed to produce steep short-dated implied volatility skews. Another example of broadly different design was provided by Caravenna and Corbetta [12]. In their 'multiscaling' model, the stock price process is continuous, while the volatility process has (carefully designed) jumps, and steepness of the smile is achieved with a heavy-tail distribution of the small-time distribution of the volatility. Rough fractional volatility models (with continuous volatility paths) have recently been proposed, and are able to capture the volatility skew [3, 7, 8, 20, 23, 26, 30, 33, 36]. In this paper we analyse the combined effect of a rough fractional Brownian driver (with Hurst parameter $H \in (0, 1)$) in the volatility and a random starting point. We quantify how the tail behaviour (parametrised by a scaling coefficient $b > 0$) of the random starting point modulates the rate of explosion in the implied volatility in the presence of rough fractional volatility. Finally, in a specific simplified setting we highlight

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

how our model blends naturally into the setting of forward-start options in stochastic volatility models, whose asymptotic properties have been studied in [38]. In proving our results, we improve the large deviations literature on both SDEs with random starting points and fractional SDEs.

In Section 2 we recall some concepts that will be used in the paper and set the notation. We also introduce the model (2.5), and the main assumptions on its dynamics and on the initial random starting point. Section 3 collects the main large deviations estimates in different regimes: tail behaviour (Section 3.1) and small-time behaviour (Section 3.2). In each case we present two different scenarios consisting of an appropriately rescaled fractional model (Theorems 3.1 and 3.4) and a simplified diffusive model (Theorems 3.3 and 3.6) with more restrictive conditions on the random starting point (allowing for simpler large deviations rate functions). Section 4 displays applications to implied volatility asymptotics (Corollaries 4.1 and 4.2), and presents an application to forward-start options. Proofs can be found in the appendix.

2. Setup and notation

As outlined in the introduction, we prove pathwise large deviations for a two-dimensional system generalising the Stein–Stein model [50, 51], with random initial datum. In particular, via suitable rescaling, we determine the small-time and the large-tail behaviours of the system. Before delving into the core of the paper, let us recall some useful facts about large deviations and Gaussian processes, which shall also serve as setting the notation for the rest of the paper. Unless otherwise stated, we always work on a finite time horizon, say $[0, 1]$ without loss of generality, which we denote by $\mathcal{T}$, and we write $\mathcal{T}^* := \mathcal{T} \setminus \{0\}$. We let $\mathcal{C} := \mathcal{C}(\mathcal{T}, \mathbb{R})$ be the space of continuous functions from $\mathcal{T}$ to $\mathbb{R}$, and $\mathcal{C}_b^2$ the space of twice-differentiable functions on $\mathcal{T}$ with bounded partial derivatives up to the second order. We write $X^\varepsilon \sim \mathrm{LDP}(h_\varepsilon, I)$ when the sequence $(X^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle (Definition 2.2) on $\mathcal{C}$, as ε tends to zero with good rate function I and speed h_ε, where h_ε denotes a function satisfying $\lim_{\varepsilon\downarrow 0} h_\varepsilon = 0$. For a random variable X, we denote by $\mathrm{supp}(X)$ its support.

2.1. Large deviations and fractional Brownian motion

We use [16] as our guide through large deviations. Given a topological space $\mathcal{X}$ and its completed Borel σ-field $\mathcal{B}_\mathcal{X}$, for any $A \in \mathcal{B}_\mathcal{X}$ we denote by A^o and $\overline{A}$ respectively its interior and closure, and consider a sequence $(X^\varepsilon)_{\varepsilon>0}$ on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$.

Definition 2.1. A (good) rate function is a lower semi-continuous mapping $I : \mathcal{X} \to [0, \infty]$ such that any level set $\{x : I(x) \le z\}$ is a closed (compact) subset of $\mathcal{X}$, for $z \ge 0$.

Definition 2.2. The sequence $(X^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle (LDP) on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ as ε tends to zero, with speed h_ε and rate function I, if for any Borel subset $A \subset \mathcal{X}$, the following inequalities hold:

$$-\inf_{A^o} I(\varphi) \le \liminf_{\varepsilon\downarrow 0} h_\varepsilon \log \mathrm{P}(X^\varepsilon \in A) \le \limsup_{\varepsilon\downarrow 0} h_\varepsilon \log \mathrm{P}(X^\varepsilon \in A) \le -\inf_{\overline{A}} I(\varphi).$$

A particularly convenient tool to prove large deviations is so-called exponential equivalence, which we recall from [16, Definition 4.2.10] as follows:

Definition 2.3. On a metric space $(\mathcal{Y}, d)$, two $\mathcal{Y}$-valued sequences $(X^\varepsilon)_{\varepsilon>0}$ and $(\widetilde{X}^\varepsilon)_{\varepsilon>0}$ are called exponentially equivalent (with speed h_ε) if there exist probability spaces $(\Omega, \mathcal{B}_\varepsilon, \mathrm{P}_\varepsilon)_{\varepsilon>0}$

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

such that for any $\varepsilon > 0$, P^ε is the joint law and, for each $\delta > 0$, the set $\{\omega : (\widetilde{X}^\varepsilon, X^\varepsilon) \in \Gamma_\delta\}$ is $\mathcal{B}_\varepsilon$-measurable, and

$$\limsup_{\varepsilon \downarrow 0} h_\varepsilon \log \mathrm{P}^\varepsilon(\Gamma_\delta) = -\infty,$$

where $\Gamma_\delta := \{(\tilde{y}, y) : d(\tilde{y}, y) > \delta\} \subset \mathcal{Y} \times \mathcal{Y}$.

Theorem 2.1. *Let $(X^\varepsilon)_{\varepsilon>0}$ and $(\widetilde{X}^\varepsilon)_{\varepsilon>0}$ be two exponentially equivalent sequences (with speed h_ε) on some metric space. If $(X^\varepsilon) \sim LDP(h_\varepsilon, \Lambda^X)$ for some good rate function Λ^X, then $(\widetilde{X}^\varepsilon) \sim LDP(h_\varepsilon, \Lambda^X)$.*

The last tool we shall need repeatedly is the following contraction principle [16, Theorem 4.2.1]:

Theorem 2.2. (Contraction principle.) *Let $\mathcal{X}$, $\mathcal{Y}$ be two topological spaces, $f : \mathcal{X} \to \mathcal{Y}$ a continuous function, and $I : \mathcal{X} \to [0, \infty]$ a good rate function. For each $y \in \mathcal{Y}$, define $I'(y) := \inf\{I(x) : x \in \mathcal{X}, y = f(x)\}$. Then, if I controls the LDP associated with a family of probability measures $\{\mu_\varepsilon\}$ on $\mathcal{X}$, then I' controls the LDP associated with the family of probability measures $\{\mu_\varepsilon \circ f^{-1}\}$ on $\mathcal{Y}$ and I' is a good rate function on $\mathcal{Y}$.*

On a real, separable Banach space $(\mathcal{E}, \|\cdot\|)$, we denote by $\mathcal{B}$ the associated Borel sigma field. Letting $\mathcal{E}^*$ denote the topological dual of $\mathcal{E}$, we define a Gaussian measure as follows:

Definition 2.4. A Gaussian measure μ on $(\mathcal{E}, \|\cdot\|)$ is such that every $\vartheta^* \in \mathcal{E}^*$, when viewed as a random variable via the dual pairing $\vartheta \mapsto \langle \vartheta^*, \vartheta \rangle_{\mathcal{E}^*\mathcal{E}}$, is a real Gaussian random variable on $(\mathcal{E}, \mathcal{B}, \mu)$.

We associate a Gaussian process with a Gaussian measure in the usual way [13, Section 3.2]. Particular examples of Gaussian processes, crucial for the rest of the paper, include standard Brownian motion on the time interval $\mathcal{T}$, where $\mathcal{E} = \mathcal{C}$ equipped with the supremum norm and with the topology of uniform convergence, and $\mathcal{E}^*$ is the space of signed measures on $\mathcal{T}$. In fact, this construction applies to all (continuous) centered Gaussian processes, which are uniquely characterised by their covariance operator. A fractional Brownian motion W^H with Hurst parameter $H \in (0, 1)$ is such a Gaussian process, starting from zero, with covariance

$$\langle W_t^H, W_s^H \rangle = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t - s|^{2H}), \qquad \text{for any } s, t \in \mathcal{T}.$$

Of primary importance in understanding small-noise behaviours of Gaussian systems is the concept of reproducing kernel Hilbert spaces (RKHS), which we recall following [13, Definition 3.3]:

Definition 2.5. Let μ be a Gaussian measure on $\mathcal{E}$ and define the map $\mathcal{R} : \mathcal{E}^* \to \mathcal{E}$ by $\mathcal{R}x^* := \int_{\mathcal{E}} \langle x^*, x \rangle x \mu(\mathrm{d}x)$. The RKHS $\mathcal{H}_\mu$ of μ is the completion of the image $\mathcal{R}\mathcal{E}^*$ for the norm $\|\mathcal{R}x^*\|_{\mathcal{H}_\mu} := (\langle x^*, \mathcal{R}x^* \rangle)^{1/2}$, for all $x^* \in \mathcal{E}^*$.

To characterise the RKHS of fractional Brownian motion, the usual tool is its Volterra representation [45],

$$W_t^H = \int_0^t K^H(t, s)\, \mathrm{d}B_s, \tag{2.1}$$

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

which holds almost surely for all $t \in \mathcal{T}$, where B is a standard Brownian motion generating the same filtration as W^H, and K^H is the Volterra kernel defined, for any $s, t \in \mathcal{T}$ with $0 < s < t$, by [45, Theorem 5.2]

$$K^H(t,s) = \begin{cases} \frac{\kappa_H}{s^{H_-}}\left[(t(t-s))^{H_-} - H_- \int_s^t \frac{(u-s)^{H_-}}{u^{1-H_-}}\,\mathrm{d}u\right], & \text{if } H < \frac{1}{2}, \\ \frac{\kappa_H H_-}{s^{H_-}} \int_s^t \frac{u^{H_-}\,\mathrm{d}u}{|u-s|^{1-H_-}}, & \text{if } H > \frac{1}{2}, \\ 1, & \text{if } H = \frac{1}{2}, \end{cases} \tag{2.2}$$

with $H_\pm := H \pm \frac{1}{2}$ and $\kappa_H := \left(\frac{2H\Gamma(1-H_-)}{\Gamma(H_+)\Gamma(2-2H)}\right)^{1/2}$. For $t \in \mathcal{T}^*$, the map $K^H(t,\cdot)$ is square integrable around the origin, and the reproducing kernel Hilbert space of the fractional Brownian motion is given by

$$\mathcal{H}_{K^H} := \left\{ \int_0^t K^H(t,s) f(s)\,\mathrm{d}s,\ t \in \mathcal{T} : f \in \mathrm{L}^2(\mathcal{T}) \right\}$$

with inner product

$$\left\langle \int_0^\cdot K^H(\cdot,s)\, f_1(s)\,\mathrm{d}s, \int_0^\cdot K^H(\cdot,s) f_2(s)\,\mathrm{d}s \right\rangle_{\mathcal{H}_{K^H}} := \langle f_1, f_2 \rangle_{\mathrm{L}^2(\mathcal{T})}.$$

The notation $\mathcal{H}_{K^H}$, emphasising the link with the underlying kernel, will be useful later (in Definition 2.6) for more general kernels. In particular, the RKHS associated with (standard) Brownian motion ($H = 1/2$) is the Cameron–Martin space, and corresponds to the space of absolutely continuous functions starting at zero with square-integrable derivatives. In other words, for any $H \in (0,1) \setminus \{1/2\}$, the identity $\mathcal{H}_{K^H} = \mathbb{K}_{K^H}\mathrm{L}^2(\mathcal{T})$ holds, where $\mathbb{K}_{K^H} : \mathrm{L}^2(\mathcal{T}) \ni f \mapsto \int_0^t K^H(t,s) f(s)\,\mathrm{d}s$. This characterisation motivates the following definition:

Definition 2.6. For any strictly positive function $\Phi : \mathbb{R}_+^2 \to \mathbb{R}$ such that $\Phi(t,\cdot) \in L^2(\mathcal{T})$ for any $t \in \mathcal{T}$, the corresponding RKHS is defined as

$$\mathcal{H}_\Phi := \left\{ \int_0^t \Phi(t,s) f(s)\,\mathrm{d}s,\ t \in \mathcal{T} : f \in \mathrm{L}^2(\mathcal{T}) \right\}, \tag{2.3}$$

with inner product

$$\left\langle \int_0^\cdot \Phi(\cdot,s) f_1(s)\,\mathrm{d}s, \int_0^\cdot \Phi(\cdot,s) f_2(s)\,\mathrm{d}s \right\rangle_{\mathcal{H}_\Phi} := \langle f_1, f_2 \rangle_{\mathrm{L}^2(\mathcal{T})}.$$

Reproducing kernel Hilbert spaces, together with their inner products, turn out to provide the right spaces to characterise large deviations rate functions. In particular, for a given Gaussian Volterra process of the form $\int_0^\cdot \Phi(\cdot,s)\,\mathrm{d}B_s$, for some Volterra kernel Φ, it follows from [19, Theorem 3.4.5] that the sequence $(\varepsilon \int_0^\cdot \Phi(\cdot,s)\,\mathrm{d}B_s)_{\varepsilon>0}$ satisfies a large deviations principle with speed ε^2 and rate function

$$\Lambda_\Phi(\varphi) = \begin{cases} \frac{1}{2}\|\dot\varphi\|_{\mathcal{H}_\Phi}, & \text{if } \varphi \in \mathcal{H}_\Phi, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.4}$$

An obviously special role is played by the standard Brownian motion $H = \frac{1}{2}$, and we shall adopt the simplified notation $\mathcal{H}$ (the classical Cameron–Martin space) and Λ in place of $\mathcal{H}_{K^{1/2}}$ and $\Lambda_{K^{1/2}}$.

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

2.2. Setting and assumptions

The particular system we are interested in is

$$\mathrm{d}X_t = -\tfrac{1}{2}\sigma(Y_t)^2\mathrm{d}t + \sigma(Y_t)(\rho\mathrm{d}B_t + \overline{\rho}\mathrm{d}B_t^{\perp}), \qquad X_0 = 0, \tag{2.5a}$$

$$\mathrm{d}Y_t = (\lambda + \beta Y_t)\mathrm{d}t + \xi\,\mathrm{d}W_t^H, \qquad Y_0 \sim \Theta, \tag{2.5b}$$

where W^H is a fractional Brownian motion with Hurst parameter $H \in (0, 1)$, $(B, B^{\perp})$ is a two-dimensional standard Brownian motion, $\beta < 0$, $\lambda, \xi > 0$, $\rho \in (-1, 1)$, $\overline{\rho} := \sqrt{1-\rho^2}$, and Θ is a square-integrable continuous random variable. In order to guarantee the existence and uniqueness of a strong solution, we further assume [44, Theorem 3.1.3] that σ is Lipschitz continuous, satisfies the growth condition $|\sigma(y)| \le C(1 + |y|)$ for $y \in \mathbb{R}$, and is differentiable with a locally Hölder-continuous derivative. In order to prove our main results below, we make the following technical assumption.

Assumption $\mathcal{A}$. *There exists a measurable function $\widetilde{\sigma} : \mathbb{R} \to \mathbb{R}$, with locally Hölder-continuous derivative, such that the scaling property $\lim_{\varepsilon \downarrow 0} \varepsilon\sigma(y/\varepsilon) = \widetilde{\sigma}(y)$ holds true uniformly on $\mathbb{R}$.*

Remark 2.1. Since the map σ is Lipschitz continuous with linear growth, then so is $\widetilde{\sigma}$. The scaling property further implies that $\lim_{\varepsilon \downarrow 0} \varepsilon^b \sigma(y/\varepsilon^b) = \widetilde{\sigma}(y)$ uniformly on $\mathbb{R}$ for any $b > 0$. This will be useful for later computations.

3. Main results

The centrepiece of our analysis are large deviations estimates for suitably rescaled versions of (2.5). The first rescaling (presented in Section 3.1) is tailored to the analysis of the tail behaviour of (2.5), while the second rescaling (Section 3.2) is bespoke to its short-time asymptotic properties. In addition to these asymptotic results in the general fractional case (Theorems 3.1 and 3.4), we present two special simplified diffusive cases, where particularly tractable rate functions can be obtained (Theorems 3.3 and 3.6 respectively). For this we impose stronger conditions (Assumption 3.1) on the random starting point. This allows us to establish, following [43] in Section 3.1.2, an exponential equivalence between (2.5) and an analogous process with fixed starting point. In Section 3.2.2 we construct a third rescaling (3.8) inspired by Mellouk [42] in the short-time diffusive case under the assumption that the support of the random starting point is bounded. We shall work with rescaled versions Y^{ε} of the process Y in (2.5) (see (3.1) and (3.6) for specific examples), together with a function h_{ε} describing the speed of the large deviations estimates, for which we introduce the following assumptions.

Assumption $\mathcal{A}'$. *There is a family of continuous functions $(\sigma_n)_{n\ge 0}$ on $\mathbb{R}$ such that*

(i) *$(\sigma_n)_{n\ge 0}$ converges uniformly to $\widetilde{\sigma}$ on $\mathbb{R}$;*

(ii) *for all $\delta > 0$, $\lim_{n\uparrow\infty} \limsup_{\varepsilon\downarrow 0} h_{\varepsilon} \log \mathrm{P}(|\sigma_n(Y^{\varepsilon}) - \widetilde{\sigma}(Y^{\varepsilon})| \ge \delta) = -\infty$.*

Assumption $\mathcal{A}_b^{\Theta}$. (Tail behaviour of Θ.) $\limsup_{\varepsilon\downarrow 0} h_{\varepsilon} \log \mathrm{P}(\varepsilon^b|\Theta| > 1) = -\infty$.

In Assumptions $\mathcal{A}'$ and $\mathcal{A}_b^{\Theta}$ above, the large deviations speed h_{ε} takes the value ε^{2b} in Section 3.1 and ε^{4H+2b} in Section 3.2 respectively. The constant b (which may vary below) plays an essential role in subsequent large deviations estimates, via exponential equivalence techniques (Definition 2.3). Assumptions $\mathcal{A}$ and $\mathcal{A}'$ are naturally satisfied in the fractional

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

Stein–Stein case (where $\sigma(y) \equiv y$), but imposing them allows us to state our results for more general σ, in particular when using extended contraction principles [40, Proposition 2.3].

3.1. Tail behaviour

3.1.1. *The general case.* For $b, \varepsilon > 0$, introduce the rescaling $(X^\varepsilon, Y^\varepsilon) := (\varepsilon^{2b} X, \varepsilon^b Y)$, so that (2.5) becomes

$$\mathrm{d}X_t^\varepsilon = -\frac{\varepsilon^{2b}}{2}\sigma\Big(\frac{Y_t^\varepsilon}{\varepsilon^b}\Big)^2 \mathrm{d}t + \varepsilon^{2b}\sigma\Big(\frac{Y_t^\varepsilon}{\varepsilon^b}\Big)(\rho \mathrm{d}B_t + \overline{\rho}\mathrm{d}B_t^\perp), \qquad X_0^\varepsilon = 0, \tag{3.1a}$$

$$\mathrm{d}Y_t^\varepsilon = (\varepsilon^b \lambda + \beta Y_t^\varepsilon)\,\mathrm{d}t + \varepsilon^b \xi \mathrm{d}W_t^H, \qquad Y_0^\varepsilon \sim \varepsilon^b \Theta. \tag{3.1b}$$

The particular rescaling considered here is perfectly suited for tail behaviour, as large deviations provide estimates for $\mathrm{P}(X^\varepsilon \geq 1) = \mathrm{P}(X \geq \varepsilon^{-2b})$. Our main result is as follows, and is proved in Appendix A.1.3:

Theorem 3.1. *For any $H \in (0, 1)$, the following hold:*

(i) *for any $b > 0$ such that $\mathcal{A}_b^\Theta$ holds, $Y^\varepsilon \sim LDP(\varepsilon^{2b}, \Lambda_{F^H})$, with Λ_{F^H} in (2.4) and F^H in Lemma 3.1;*

(ii) *for any $b \geq \frac{1}{2}$ such that $\mathcal{A}$, $\mathcal{A}'$, $\mathcal{A}_b^\Theta$ hold, $X^\varepsilon \sim LDP(\varepsilon^{2b}, \widetilde{\Lambda})$, with $\widetilde{\Lambda}$ in (A.2).*

The proof of the theorem, developed later, requires a precise analysis of the RKHS of the processes under consideration, and we first state two key ingredients (proved in Appendices A.1.1 and A.1.2), which are also of independent interest. Recall from [53, Definition 2.1] the definition of the stochastic integral: for any smooth function f on $\mathcal{T}$ with bounded derivatives such that $\|f\| := \int_{\mathcal{T}} (\Gamma^*_{H,t} f(t))^2 \,\mathrm{d}t < \infty$, define, for $t \in \mathcal{T}$, $\int_0^t f(u)\,\mathrm{d}W_u^H := \int_0^t \Gamma^*_{H,t} f(u)\,\mathrm{d}Z_u$, with

$$\Gamma^*_{H,t} f(s) := -\frac{\kappa_H}{s^{H_-}} \frac{\mathrm{d}}{\mathrm{d}s} \int_s^t u^{H_-}(u-s)^{H_-} f(u)\,\mathrm{d}u, \qquad \text{for all } s \in (0, t].$$

Lemma 3.1. *For $H \in (0, 1)$ and $\beta > 0$, there exists a Brownian motion Z such that*

$$\xi \int_0^t e^{\beta(t-s)}\,\mathrm{d}W_s^H = \int_0^t F^H(t, s)\,\mathrm{d}Z_s \tag{3.2}$$

holds almost surely for $t \in \mathcal{T}$, where $F^H : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ is defined for $0 < s < t$, with κ_H in (2.2), as

$$F^H(t, s) := \begin{cases} \dfrac{\xi\kappa_H}{s^{H_-}}\Big[[t(t-s)]^{H_-} \\ \qquad + \displaystyle\int_s^t \Big\{\frac{1-2H}{2u} + \beta\Big\}[u(u-s)]^{H_-} e^{\beta(t-u)}\,du\Big], & \text{if } H < \dfrac{1}{2}, \\ \dfrac{\xi\kappa_H H_-}{s^{H_-}} \displaystyle\int_s^t \frac{u^{H_-} e^{\beta(t-u)}}{(u-s)^{1-H_-}}\,du, & \text{if } H > \dfrac{1}{2}, \\ \xi e^{\beta(t-s)}, & \text{if } H = \dfrac{1}{2}. \end{cases}$$

The case $\beta = 0$ (and $\xi = 1$) is excluded since, in that case, the lemma boils down to the classical Volterra representation of the fractional Brownian motion (2.1) and the function F^H

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

is just the kernel K^H given in (2.2). For $H \in (\frac{1}{2}, 1)$, the expression for F^H is in agreement with [53, Definition 2.1], but for $H \in (0, \frac{1}{2})$ it corrects the slightly erroneous expression therein. This function F^H allows us to fully characterise the following RKHS, with proof postponed to Section A.1.2:

Proposition 3.1. *For any $H \in (0, 1)$, $\beta > 0$, the space $\mathcal{H}_{F^H}$ is the RKHS of the Gaussian process $\xi \int_0^{\cdot} e^{\beta(\cdot - s)} \, \mathrm{d}W_s^H$.*

3.1.2. *The Millet–Nualart–Sanz approach.* In [43], Millet, Nualart, and Sanz considered a perturbed stochastic differential equation of the form

$$\mathrm{d}X_t^\varepsilon = \mathfrak{b}(\varepsilon, X_t^\varepsilon)\mathrm{d}t + \varepsilon \mathfrak{a}(X_t^\varepsilon)\mathrm{d}W_t. \tag{3.3}$$

Here, for any $\varepsilon > 0$, the functions $\mathfrak{b}(\varepsilon, \cdot) : \mathbb{R}^n \to \mathbb{R}^n$ and $\mathfrak{a} : \mathbb{R}^n \to \mathcal{M}_{(n,d)}(\mathbb{R})$ are bounded Borel measurable and uniformly Lipschitz, $\mathfrak{b}(\varepsilon, \cdot)$ converges uniformly to a function $\mathfrak{b}(\cdot)$ as ε tends to zero, W is a d-dimensional Brownian motion, and X_0^ε is an $\mathbb{R}^n$-valued square-integrable random variable. Proofs of the existence and uniqueness of a strong solution can be found in [24, Chapter 5, Theorem 2.1]. Following classical large deviations steps, consider, for any $\varphi \in \mathcal{H}$, the following controlled ordinary differential equation (ODE) on $\mathcal{T}$:

$$\dot{\psi}_t = \mathfrak{a}(g_t)\dot{\varphi}_t + \mathfrak{b}(\psi_t),$$

the solution flow of which, starting from $x_0 \in \mathbb{R}^n$, is denoted by $\mathcal{S}_{x_0}(\varphi)$. Millet, Nualart, and Sanz [43] proved a large deviations principle [43, Theorem 4.1] for the sequence $(X^\varepsilon)_{\varepsilon>0}$ under the following assumption.

Assumption 3.1. *Both $\mathfrak{a}(\cdot)$ and $\mathfrak{b}(\cdot)$ belong to $\mathcal{C}_b^2$, and there exists $x_0 \in \mathbb{R}^n$ such that, for any $\delta > 0$,*

$$\limsup_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathrm{P}(|X_0^\varepsilon - x_0| > \delta) = -\infty. \tag{3.4}$$

Theorem 3.2. *Under Assumption 3.1, $(X^\varepsilon)_{\varepsilon>0} \sim LDP(\varepsilon^2, I)$ with $I(\psi) = \inf\{\Lambda(\varphi) : \varphi \in \mathcal{H}, \psi = \mathcal{S}_{x_0}(\varphi)\}$.*

Condition (3.4) is an exponential equivalence property between the initial random variable X_0^ε and the constant x_0, and ensures that large deviations are preserved under exponentially small perturbations of the starting point. Therefore, in the standard diffusion case $H = \frac{1}{2}$, it is possible to obtain a similar result to Theorem 3.1 with a simplified rate function (albeit with stronger restrictions on the starting point) by using the approach in [43]. For this, we rewrite (3.3) to correspond to (3.1), albeit with stronger assumptions on the coefficients, with $\mathrm{W} := (W_1, W_2)'$ a Brownian motion, $\mathrm{X}^\varepsilon = (X^\varepsilon, Y^\varepsilon)$, $H = \frac{1}{2}$, $\varepsilon \to \varepsilon^b$, $\mathrm{X}_0^\varepsilon = (0, \varepsilon\Theta)$, and $\overline{\rho} := \sqrt{1 - \rho^2}$:

$$\mathfrak{b}(\varepsilon, \mathrm{X}_t^\varepsilon) = \begin{pmatrix} -\frac{1}{2}\widetilde{\sigma}(Y_t^\varepsilon)^2 \\ \varepsilon\lambda + \beta Y_t^\varepsilon \end{pmatrix} \qquad \text{and} \qquad \mathfrak{a}(\mathrm{X}_t^\varepsilon) = \begin{pmatrix} \overline{\rho}\widetilde{\sigma}(Y_t^\varepsilon) & \rho\widetilde{\sigma}(Y_t^\varepsilon) \\ 0 & \xi \end{pmatrix}.$$

The correlation between the two components of W is explicitly represented in the diffusion matrix $\mathfrak{a}$.

Theorem 3.3. *Under Assumption 3.1, the solution X^ε to (3.3) satisfies $(\mathrm{X}^\varepsilon)_{\varepsilon>0} \sim \mathrm{LDP}(\varepsilon^2, \mathbf{I})$ with $\mathbf{I}(\chi) = \inf\{\Lambda(\varphi) : \varphi \in \mathcal{H}, \chi = \mathcal{S}_{x_0}(\Psi^\rho(\varphi))\}$ and*

$$\Psi^\rho : \mathbb{R}^2 \ni z \mapsto \begin{pmatrix} \overline{\rho} & \rho \\ 0 & 1 \end{pmatrix} z, .$$

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

Proof. The proof of Theorem 3.2 relies first on proving a large deviations principle for the flow $\mathcal{S}_{x_0}$ using Schilder's Theorem [16, Theorem 5.2.3], then on extending this LDP to the original system. One can easily extend it to include a correlation parameter $\rho \in (-1, 1)$, the main difference being the rate function. Indeed, since $W_2^\perp$ and W_2 are independent, Schilder's Theorem yields that $\varepsilon(W_2^\perp, W_2)' \sim \mathrm{LDP}(\varepsilon^2, \Lambda)$. Since the map Ψ^ρ is continuous on $(\mathcal{C}, \|\cdot\|_\infty)$ and $\varepsilon \mathrm{W}' = \Psi^\rho(\varepsilon(W_2^\perp, W_2)')$, the theorem follows from the contraction principle giving an LDP for $\varepsilon \mathrm{W}'$ as ε tends to zero with speed ε^2 and good rate function

$$\Lambda^\rho(\psi) := \inf\{\Lambda(\varphi) : \varphi \in \mathcal{H},\ \psi = \Psi^\rho(\varphi)\}. \tag{3.5}$$

This completes the proof. □

3.2. Small-time behaviour

We now tackle the small-time behaviour of the process (2.5). Under the general set of assumptions $\mathcal{A}$, $\mathcal{A}'$, $\mathcal{A}_b^\Theta$, we need to introduce a particular rescaling both in time and in space in order to observe some weak convergence. This is different from the classical Itô diffusion case (with fixed starting point), where solutions of such SDEs generally converge in small time. In the Itô case, though, if the distribution of the starting point has compact support, we show in Section 3.2.2 that space rescaling is not required any longer.

3.2.1. *The general case.* The rescaling $(X_t^\varepsilon, Y_t^\varepsilon) := (\varepsilon^{2H+2b-1} X_{\varepsilon^2 t}, \varepsilon^b Y_{\varepsilon^2 t})$, for $b > 0$, turns (2.5) into

$$\mathrm{d}X_t^\varepsilon = -\frac{\varepsilon^{2H+1+2b}}{2}\sigma\left(\frac{Y_t^\varepsilon}{\varepsilon^b}\right)^2 \mathrm{d}t + \varepsilon^{2H+2b}\sigma\left(\frac{Y_t^\varepsilon}{\varepsilon^b}\right)(\rho \mathrm{d}B_t + \overline{\rho}\mathrm{d}B_t^\perp), \qquad X_0^\varepsilon = 0, \tag{3.6a}$$

$$\mathrm{d}Y_t^\varepsilon = (\varepsilon^{b+2}\lambda + \beta\varepsilon^2 Y_t^\varepsilon)\mathrm{d}t + \varepsilon^{2H+b}\xi \mathrm{d}W_t^H, \qquad Y_0^\varepsilon \sim \varepsilon^b \Theta. \tag{3.6b}$$

Our main result is as follows:

Theorem 3.4. *For any $H \in (0, 1)$,*

(i) *for any $b > 0$ such that $\mathcal{A}_b^\Theta$ holds, $Y^\varepsilon \sim LDP(\varepsilon^{4H+2b}, \Lambda_{G_0^H})$, with $\Lambda_{G_0^H}$ in (2.4);*

(ii) *if $b \geq \frac{1}{2} - 2H$ such that $\mathcal{A}$, $\mathcal{A}'$, $\mathcal{A}_b^\Theta$ hold, $X^\varepsilon \sim LDP(\varepsilon^{4H+2b}, \mathtt{I})$, with $\mathtt{I}$ in (A.5).*

The proof of (i) is similar to that of Theorem 3.1(i) and relies on proving LDP for an auxiliary process, defined in (A.1), exponentially equivalent to the original (rescaled) process Y^ε. The proof of (ii) is more involved and is postponed to Appendix A.2.

In order to state the following key result, define, for any $\varepsilon > 0$, G_ε^H as the function F^H in Lemma 3.1, replacing β by $\beta\varepsilon^2$, and for $s, t \in \mathcal{T}$ with $0 < s < t$, its pointwise limit

$$G_0^H(t, s) := \lim_{\varepsilon \downarrow 0} G_\varepsilon^H(t, s)$$

$$= \begin{cases} \dfrac{\xi \kappa_H H_-}{s^{H_-}} \displaystyle\int_s^t \frac{u^{H_-}}{(u-s)^{1-H_-}}\,\mathrm{d}u, & \text{for } H \in \left(\frac{1}{2}, 1\right), \\ \dfrac{\xi \kappa_H}{s^{H_-}}\left((t(t-s))^{H_-} - H_- \displaystyle\int_s^t (u-s)^{H_-} u^{H_- - 1}\,\mathrm{d}u\right), & \text{for } H \in \left(0, \frac{1}{2}\right), \\ \xi, & \text{for } H = \dfrac{1}{2}. \end{cases}$$

The following proposition is similar to Proposition 3.1, as $G_0^H(t, \cdot) \in \mathrm{L}^2(\mathcal{T})$ and, for all $0 < s < t$, $G_0^H(t, s) > 0$, and its proof is omitted.

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

Proposition 3.2. *For any $H \in (0, 1)$, the space $\mathcal{H}_{G_0^H}$ is the RKHS of the Gaussian process $\int_0^\cdot G_0^H(\cdot, s)\, dZ_s$.*

3.2.2. *Small-time asymptotics for bounded support in the diffusion case.* In the standard case $H = \frac{1}{2}$, the rescaling in the previous subsection is not really 'natural', in the sense that small-time weak convergence usually holds for Itô diffusions without space rescaling. In this case, using an approach introduced by Bezuidenhout [10] and further developed by Mellouk [42], we can obtain simpler large deviations estimates if the support of the initial datum Θ is bounded. A simplified version of Mellouk considers, for any $\varepsilon > 0$, the following system on $\mathcal{T}$:

$$\mathrm{d}X_t^\varepsilon = \mathfrak{b}(X_t^\varepsilon, Z)\mathrm{d}t + \varepsilon\mathfrak{a}(X_t^\varepsilon, Z)\mathrm{d}W_t, \qquad X_0^\varepsilon = x_0 \in \mathbb{R}^n, \tag{3.7}$$

where $\mathfrak{b} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $\mathfrak{a} : \mathbb{R}^n \times \mathbb{R}^m \to \mathcal{M}_{(n,d)}$ are bounded Borel measurable, uniformly Lipschitz continuous, Z is a random variable with compact support on $\mathbb{R}^m$, and W is a d-dimensional standard Brownian motion, independent of Z. The main result of the paper is a large deviations principle on $\mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^n)$, the space of α-Hölder continuous functions, for $0 \le \alpha < \frac{1}{2}$, for the sequence $(X^\varepsilon)_{\varepsilon>0}$ under the following assumptions.

Assumption 3.2.

(H1) $\mathfrak{b}(\cdot, \cdot)$ *is jointly measurable on $\mathbb{R}^n \times \mathbb{R}^m$ and there exists $C > 0$ such that, for all $x, x' \in \mathbb{R}^n$, $z, z' \in \mathbb{R}^m$,*

$$|\mathfrak{b}(x, z)| \le C(1 + |x|) \qquad \textit{and} \qquad |\mathfrak{b}(x, z) - \mathfrak{b}(x', z')| \le C(|x - x| + |z - z'|).$$

(H2) $\mathfrak{a}(\cdot, \cdot)$ *is jointly measurable on $\mathbb{R}^n \times \mathbb{R}^m$ and there exists $C > 0$ such that, for all $x, x' \in \mathbb{R}^n$, $z, z' \in \mathbb{R}^m$,*

$$\|\mathfrak{a}(x, z)\| \le C \qquad \textit{and} \qquad \|\mathfrak{a}(x, z) - \mathfrak{a}(x', z')\| \le C(|x - x'| + |z - z'|).$$

For $f \in \mathcal{H}$, $u \in \mathrm{supp}(Z)$ and $x_0 \in \mathbb{R}^n$, let $\mathcal{S}_{x_0}(f, u)$ denote the unique solution to the controlled ODE $g_t = x_0 + \int_0^t \mathfrak{b}(g_s, u_s)\, \mathrm{d}s + \int_0^t \mathfrak{a}(g_s, u_s)\dot{f}_s\, \mathrm{d}s$, for $t \in \mathcal{T}$. Let us now introduce the following definition:

Definition 3.1. Let $\alpha \in [0, \frac{1}{2})$ and $\mathcal{B}_a$ be the ball of radius a in the α-Hölder norm. The lower semi-continuous regularisation $\breve{I} : \mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^m) \to \mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^m)$ of a functional $I : \mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^m) \to \mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^m)$ is defined as

$$\breve{I}(\psi) := \lim_{a \downarrow 0} \inf\{I(\varphi) : \varphi \in \mathcal{B}_a(\psi)\}.$$

Theorem 3.5. (Theorem 2.1 in [42].) *Under Assumption 3.2, $(X^\varepsilon)_{\varepsilon>0} \sim LDP(\varepsilon^2, \breve{I}^\alpha)$, where, with Λ in (2.4),*

$$I^\alpha(\psi) := \inf\{\Lambda(\varphi) : \varphi \in \mathcal{H}, \mathcal{S}_{x_0}(\varphi, u) = \psi, \textit{ for some } u \in \textit{supp}(Z)\}.$$

Coming back to our model, the rescaling $(X_t^\varepsilon, Y_t^\varepsilon) := (X_{\varepsilon^2 t}, Y_{\varepsilon^2 t})$, equivalent to that of Section 3.2.1 with $b = 0$ and $H = \frac{1}{2}$, the small-noise system (2.5), under Assumption $\mathcal{A}$, becomes, similarly to Section 3.2.1,

$$\mathrm{d}X_t^\varepsilon = -\frac{\varepsilon^2}{2}\widetilde{\sigma}(Y_t^\varepsilon)^2\mathrm{d}t + \varepsilon\widetilde{\sigma}(Y_t^\varepsilon)(\rho\mathrm{d}B_t + \overline{\rho}\mathrm{d}B_t^\perp), \qquad X_0^\varepsilon = 0, \tag{3.8a}$$

$$\mathrm{d}Y_t^\varepsilon = (\varepsilon^2\lambda + \beta\varepsilon^2 Y_t^\varepsilon)\,\mathrm{d}t + \varepsilon\xi\,\mathrm{d}B_t, \qquad Y_0^\varepsilon \sim \Theta, \tag{3.8b}$$

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

with B a standard Brownian motion. Subtracting the initial random datum $\mathrm{X}_0^\varepsilon = \mathrm{X}_0 = (0, \Theta)'$, this system can be expressed in the form (3.7) with $\mathrm{X}^\varepsilon = (X^\varepsilon, Y^\varepsilon)$,

$$\mathfrak{b}(\varepsilon, \mathrm{X}^\varepsilon, \mathrm{X}_0) = \varepsilon^2 \begin{pmatrix} -\frac{1}{2}\widetilde{\sigma}(Y_t^\varepsilon + \Theta)^2 \\ \lambda + \beta(Y_t^\varepsilon + \Theta) \end{pmatrix}, \text{ and } a(\mathrm{X}^\varepsilon, \mathrm{X}_0) = \begin{pmatrix} \rho\widetilde{\sigma}(Y_t^\varepsilon + \Theta) & \overline{\rho}\widetilde{\sigma}(Y_t^\varepsilon + \Theta) \\ \xi & 0 \end{pmatrix}, \tag{3.9}$$

and note that $\mathfrak{b}(\varepsilon, \cdot, \cdot)$ converges to the null map as ε tends to zero. The assumptions imposed in [42] on the drift and diffusion coefficients are clearly satisfied here. While Mellouk allows the drift and diffusion to depend explicitly on external random factors, we can write our setting (dependence on a random starting point) into this framework. The large deviations estimate for the sequence $(\mathrm{X}^\varepsilon)_{\varepsilon>0} = (X^\varepsilon, Y^\varepsilon)_{\varepsilon>0}$ thus obtained is stronger than that in the previous section, as it holds on $\mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^n)$, for any $0 \le \alpha < \frac{1}{2}$. Note that the mild conditions on the coefficients [42, (H_0)–(H_2)] are easily satisfied in our case, so the only additional assumption is the boundedness of the support on Θ. We remark here that [42] is not directly applicable to the current setting but has to be extended to include ε-dependence in the drift, and we do so following the approach developed by Peithmann [47, Section 2.2.1] as inspired by Azencott [4]. In order to state LDP (proved in Section A.3), we impose the following assumption.

Assumption 3.3. *For $u \in supp(X_0)$ and $\varphi \in \mathcal{H}$, the ODE $\psi_t = \int_0^t \mathfrak{a}(\psi_s, u_s)\dot{\varphi}_s\, ds$ has a unique solution on $\mathcal{T}$, denoted by $\mathcal{S}_0(\varphi, u)$.*

Theorem 3.6. *Under Assumption 3.3, if Θ has compact support then $(\mathrm{X}^\varepsilon)_{\varepsilon>0} \sim \mathrm{LDP}(\varepsilon^2, \check{I}^\alpha)$, with Λ^ρ defined in (3.5) and*

$$I^\alpha(\psi) := \inf\{\Lambda^\rho(\varphi) : \varphi \in \mathcal{H}, \text{ such that } \mathcal{S}_0(\varphi, u) = \psi, \text{ for some } u \in supp(X_0)\}.$$

4. Applications to implied volatility asymptotics

As announced in the introduction, we unify two branches of research here, both aimed at reproducing the steepness of the implied volatility surface on the short end via models with continuous paths. While there are now numerous outputs [3, 7, 8, 23, 26, 30, 36] in the literature confirming that a fractional driving noise (with Hurst exponent $H < 1/2$) in the volatility leads to the observed steepness of the smile, recent results [37, 39] reproduce this effect by randomising the initial volatility in classical diffusive models. In this section we demonstrate how to modulate the two effects with respect to one another. In the Black–Scholes–Merton model, the price of a European Call option is $\mathcal{C}_{\mathrm{BS}}(t, \mathrm{e}^k, \Sigma)$, with associated volatility Σ. Considering a market with observed Call option prices $\mathcal{C}_{\mathrm{obs}}(t, \mathrm{e}^k)$, with maturity t and strike e^k, we denote by $\Sigma_t(k)$ the implied volatility, defined as the unique non-negative solution to $\mathcal{C}_{\mathrm{BS}}(t, \mathrm{e}^k, \Sigma_t(k)) = \mathcal{C}_{\mathrm{obs}}(t, \mathrm{e}^k)$.

4.1. General fractional case

From Theorems 3.1 and 3.4, we can deduce the asymptotic behaviour of the implied volatility for large strikes and for small maturities. We state those below, and postpone the proofs to Appendices A.4 and A.5.

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

Corollary 4.1. (Large-strike implied volatility asymptotics.) *For any $H \in (0,1)$ and any $b \geq 1/2$ such that Theorem 3.1 holds, we have the following large-strike asymptotic estimates of the implied volatility:*

$$\lim_{k\uparrow\infty} \frac{\Sigma_t^2(k)t}{k} = \frac{1}{2}\big(\inf_{y\geq 1} \widetilde{\Lambda}(\phi)|_{\phi_t=y}\big)^{-1}, \quad \text{with } \widetilde{\Lambda} \text{ as in (A.2), and for any } t \in \mathcal{T}.$$

Similarly, Theorem 3.4 yields the small-time behaviour of the implied volatility. The following corollary generalises [23, Corollary 4.10].

Corollary 4.2. (Small-time implied volatility asymptotics.) *For $H \in (0,1)$, $b \geq \frac{1}{2} - 2H$ such that Theorem 3.4 holds, the following small-time estimate is true for any $k \neq 0$:*

$$\lim_{t\downarrow 0} t^b \Sigma_t^2\big(t^{1/2-H-b}k\big) = \frac{k^2}{2}\big(\inf_{y\geq k} \mathcal{I}(\phi)|_{\phi_1=y}\big)^{-1}, \quad \text{with } \mathcal{I} \text{ as in (A.5).}$$

This implies that the implied volatility explodes with rate t^{-b}. For $b = 0$, it is identical to [23, Formula (26)].

4.2. Refined asymptotic results in the special diffusive case from Sections 3.1.2 and 3.2.2

4.2.1. *Large-strike asymptotics.* We consider here a specific case of a multidimensional diffusion, as we are only interested in studying its tail asymptotics. Let $\mathrm{X}^\zeta := (X^{\zeta,1}, \dots, X^{\zeta,n})$ be the unique strong solution in $\mathbb{R}^n$ to

$$\mathrm{d}\mathrm{X}_t^\zeta = \widetilde{\mathfrak{b}}(\mathrm{X}_t^\zeta)\mathrm{d}t + \mathfrak{a}(\mathrm{X}_t^\zeta)\mathrm{d}W_t, \qquad \mathrm{X}_0^\zeta = \zeta,$$

for some d-dimensional standard Brownian motion and some square-integrable random variable ζ, and with $\widetilde{\mathfrak{b}} : \mathbb{R}^n \to \mathbb{R}^n$ and $\mathfrak{a} : \mathbb{R}^n \to \mathcal{M}_{(n,d)}(\mathbb{R})$. Consider the following scaling assumption.

Assumption 4.1. *There exist $b_1, \dots, b_n > 0$ with $b_1 = 2$ such that $\mathrm{X}^{\varepsilon,\zeta} := (\varepsilon^{b_1}X^{\zeta,1}, \dots, \varepsilon^{b_n}X^{\zeta,n})$ satisfies*

$$\mathrm{d}\mathrm{X}_t^{\varepsilon,\zeta} = \varepsilon\widetilde{\mathfrak{b}}(\varepsilon, \mathrm{X}_t^{\varepsilon,\zeta})\mathrm{d}t + \varepsilon\mathfrak{a}(\mathrm{X}_t^{\varepsilon,\zeta})\mathrm{d}W_t, \qquad \mathrm{X}_0^\zeta = (\varepsilon^{b_1}\zeta^{(1)}, \dots, \varepsilon^{b_n}\zeta^{(n)}). \tag{4.1}$$

Furthermore, $\varepsilon\widetilde{\mathfrak{b}}(\varepsilon, \cdot/\varepsilon)$ converges uniformly to some function $\mathfrak{b}(\cdot)$ as ε tends to zero.

We can then state the main result about tail asymptotics:

Proposition 4.1. *Under Assumption 4.1, if there exists $\mathrm{x}_0 \in \mathbb{R}^n$ for which any $\delta > 0$ implies $\limsup_{\varepsilon\downarrow 0} \varepsilon^2 \log \mathrm{P}(|\zeta - \mathrm{x}_0| > \delta) = -\infty$, and if the triple $(\mathrm{x}_0, \mathfrak{b}(\cdot), \mathfrak{a}(\cdot))$ satisfies Assumption 3.1, then*

$$\lim_{\varepsilon\downarrow 0} \varepsilon^2 \log \mathrm{P}(\varepsilon X_t^{\zeta,1} \geq 1) = \lim_{\varepsilon\downarrow 0} \varepsilon^2 \log \mathrm{P}(\varepsilon X_t^{x_0,1} \geq 1), \qquad \text{for any } t \in \mathcal{T}.$$

The scaling from Assumption 4.1 may be odd at first, but reflects the fact that components of stochastic models may each act on different scales. Consider, for example, the Ornstein–Uhlenbeck process, solution to

$$\begin{aligned} \mathrm{d}X_t &= -\tfrac{1}{2}Y_t^2\mathrm{d}t + Y_t\mathrm{d}W_t, & X_0 &= \zeta, \\ \mathrm{d}Y_t &= (\lambda + \beta Y_t)\mathrm{d}t + \xi\,\mathrm{d}B_t, & Y_0 &= y_0 > 0, \end{aligned}$$

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

where W and B are correlated Brownian motions. The rescaling $(X^\varepsilon, Y^\varepsilon) := (\varepsilon^2 X, \varepsilon Y)$ (corresponding to $b_1 = 2$ and $b_2 = 1$) yields

$$\begin{aligned} \mathrm{d}X_t^\varepsilon &= -\tfrac{1}{2}(Y_t^\varepsilon)^2 \mathrm{d}t + \varepsilon Y_t^\varepsilon \mathrm{d}W_t, \quad & X_0^\varepsilon &= \varepsilon^2 \zeta, \\ \mathrm{d}Y_t^\varepsilon &= (\varepsilon\lambda + \beta Y_t^\varepsilon)\mathrm{d}t + \varepsilon\xi \mathrm{d}B_t, \quad & Y_0^\varepsilon &= \varepsilon y_0 > 0, \end{aligned}$$

namely (4.1), and the assumptions are satisfied.

The proof of Proposition 4.1 is postponed to Appendix A.6. The assumption of the random initial condition $\mathrm{X}_0^\zeta = \zeta$ being an $\mathcal{F}_0$-measurable distribution can be relaxed. Indeed, $\mathcal{F}$ is the filtration generated by the Brownian motion W. Instead, one could work with a filtration $\mathcal{F}' := (\mathcal{F}'_t)_{t\in\mathcal{T}^*}$ generated by $\mathcal{F}'_t := \sigma(\{\mathrm{W}_u, u \le t\} \cup \{\zeta\})$, for all $t \in \mathcal{T}^*$. Then the random initial point X_0^ζ has an $\mathcal{F}'_0$-measurable distribution and the results above still hold, in particular Theorem 3.2 and Proposition 4.1, on the new filtered probability space $(\Omega, \mathcal{F}', (\mathcal{F}')_{t\in\mathcal{T}}, \mathrm{P})$. In the context of implied volatility asymptotics, this result has the following meaning:

Corollary 4.3. *The wings of the smile are independent of the starting point (ζ or x_0).*

Proof. Gao and Lee [27] showed that the asymptotic behaviour of the implied volatility can be directly inferred from comparing tail probabilities to those of the Black–Scholes model. It is straightforward to see that the scaling of Proposition 4.1 is the same in Black–Scholes, and the corollary follows immediately. □

4.2.2. *Small-time asymptotics for the 'forward' Stein–Stein model.* We are interested in a 'forward' process, as defined by Jacquier and Roome [37] in the context of forward-start European options:

$$\mathbb{E}\Big(\frac{S_{t+\tau}}{S_t} - \mathrm{e}^k\Big)^+ = \mathrm{E}(\mathrm{e}^{X_{t+\tau}-X_t} - \mathrm{e}^k)^+ =: \mathrm{E}(\mathrm{e}^{X_\tau^{(t)}} - \mathrm{e}^k)^+,$$

with $(X_\tau^{(t)})_{\tau\ge 0}$ the so-called 'forward' process, defined path-wise by $X_\tau^{(t)} := X_{t+\tau} - X_t$, for some fixed $t > 0$ and for all $\tau \ge 0$. The 'forward' process $(X_\tau^{(t)})_{\tau\ge 0}$ then satisfies the following stochastic differential equation:

$$\mathrm{d}X_\tau^{(t)} = -\frac{1}{2}(Y_\tau^{(t)})^2 \mathrm{d}\tau + Y_\tau^{(t)} \mathrm{d}W_{1,\tau}, \qquad X_0^{(t)} = 0, \tag{4.2a}$$

$$\mathrm{d}Y_\tau^{(t)} = (\lambda + \beta Y_\tau^{(t)})\mathrm{d}\tau + \xi \mathrm{d}W_{2,\tau}, \qquad Y_0^{(t)} \sim \sigma_t. \tag{4.2b}$$

The stochastic differential equation for $(X_\tau^{(t)}, Y_\tau^{(t)})_{\tau\ge 0}$ is the same as that for $(X_t, Y_t)_{t\ge 0}$, albeit with an initial random distribution (δ_0, σ_t), where σ_t is Gaussian with mean $\mathrm{e}^{\beta t}(\sigma_0 + \frac{\lambda}{\beta}) - \frac{\lambda}{\beta}$ and variance $\frac{\xi^2}{2\beta}(\mathrm{e}^{2\beta t} - 1)$. We now apply the results of Section 3.2.2 to obtain small-time asymptotics for a version of the Stein–Stein 'forward' model with a generalised random starting point.

Proposition 4.2. *With the scaling $(X_\tau^\varepsilon, Y_\tau^\varepsilon) := (X_{\varepsilon^2\tau}^{(t)}, Y_{\varepsilon^2\tau}^{(t)})$ for $\varepsilon, t > 0$, the randomised Stein–Stein rescaled model (4.2) is the same as (3.8) with coefficients given in (3.9), with $\widetilde{\sigma}(y) \equiv y$, and Theorem 3.6 applies.*

We can translate this result into forward implied volatility asymptotics directly using [21, Theorem 2.4], and refer to this very paper for a precise definition of the forward implied volatility $\Sigma_{t,\tau}$:

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

Corollary 4.4. *The small-time forward smile reads* $\lim_{\tau\downarrow 0}\Sigma^2_{t,\tau}(k)=\frac{k^2}{2}\big(\inf_{y\geq k}\check{I}^{\alpha}(\phi)|_{\phi_1=y}\big)^{-1}$, *with* $\check{I}^{\alpha}$ *in Theorem 3.6.*

Appendix A. Proofs

A.1. Proof of Theorem 3.1

A.1.1. *Proof of Lemma 3.1.* From [53, Definition 2.1], for any smooth function f on $\mathcal{T}$ with bounded derivatives such that $\|f\|:=\int_{\mathcal{T}}(\Gamma^*_{H,t}f(t))^2\,\mathrm{d}t<\infty$, define, for $t\in\mathcal{T}$, $\int_0^t f(u)\,\mathrm{d}W^H_u:=\int_0^t\Gamma^*_{H,t}f(u)\,\mathrm{d}Z_u$, with

$$\Gamma^*_{H,t}f(s):=-\frac{\kappa_H}{s^{H_-}}\frac{\mathrm{d}}{\mathrm{d}s}\int_s^t u^{H_-}(u-s)^{H_-}f(u)\,\mathrm{d}u,\qquad\text{for all } s\in(0,t].$$

Introduce the function $\omega_s(t):=(t-s)^{H_-}t^{H_-}$. Applying this to $f(s):=\xi\mathrm{e}^{\beta(t-s)}$, one obtains the following:

(1) For $H<\frac{1}{2}$, using integration by parts and Leibniz' integration rule,

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}s}\int_s^t u^{H_-}(u-s)^{H_-}f(u)\,\mathrm{d}u\\&=\frac{\mathrm{d}}{\mathrm{d}s}\left\{\frac{\xi(t-s)^{H_+}t^{H_-}}{H_+}\right\}\\&\quad+\frac{\mathrm{d}}{\mathrm{d}s}\left\{-\xi\frac{H_-}{H_+}\int_s^t\frac{(u-s)^{H_+}}{u^{1-H_-}}\mathrm{e}^{\beta(t-u)}\,\mathrm{d}u+\xi\frac{\beta}{H_+}\int_s^t(u-s)^{H_+}u^{H_-}\mathrm{e}^{\beta(t-u)}\,\mathrm{d}u\right\}\\&=-\xi(t-s)^{H_-}t^{H_-}+\xi H_-\int_s^t\frac{\omega_s(u)}{u}\mathrm{e}^{\beta(t-u)}\,\mathrm{d}u-\xi\beta\int_s^t\omega_s(u)\mathrm{e}^{\beta(t-u)}\,\mathrm{d}u.\end{aligned}$$

Hence,

$$\begin{aligned}\Gamma^*_{H,t}f(s)&:=-\frac{\kappa_H}{s^{H_-}}\frac{\mathrm{d}}{\mathrm{d}s}\int_s^t\omega_s(u)f(u)\,\mathrm{d}u\\&=\frac{\xi\kappa_H}{s^{H_-}}\Big[\omega_s(t)+\beta\int_s^t\omega_s(u)\mathrm{e}^{\beta(t-u)}\,\mathrm{d}u-H_-\int_s^t\frac{\omega_s(u)}{u}\mathrm{e}^{\beta(t-u)}\,\mathrm{d}u\Big].\end{aligned}$$

(2) For $H>\frac{1}{2}$, using Leibniz' integration rule,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\int_s^t\omega_s(u)f(u)\,\mathrm{d}u&=\int_s^t\xi\frac{\mathrm{d}}{\mathrm{d}s}\omega_s(u)\mathrm{e}^{\beta(t-u)}\,\mathrm{d}u\\&=-\xi H_-\int_s^t u^{H_-}(u-s)^{H_--1}\mathrm{e}^{\beta(t-u)}\,\mathrm{d}u,\end{aligned}$$

and hence

$$\begin{aligned}\Gamma^*_{H,t}f(s)&:=-\frac{\kappa_H}{s^{H_-}}\frac{\mathrm{d}}{\mathrm{d}s}\int_s^t\omega_s(u)f(u)\,\mathrm{d}u\\&=\frac{\xi\kappa_H H_-}{s^{H_-}}\int_s^t u^{H_-}(u-s)^{H_--1}\mathrm{e}^{\beta(t-u)}\,\mathrm{d}u.\end{aligned}$$

□

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

A.1.2. *Proof of Proposition 3.1.* In order to prove that $\mathcal{H}_{F^H}$ is the RKHS of the Gaussian process $\xi \int_0^\cdot \mathrm{e}^{\beta(\cdot-s)}\,\mathrm{d}W_s^H$, one needs to show that the operator $\mathbb{K}_{F^H}$ is a bijection from $\mathrm{L}^2(\mathcal{T})$ to $\mathcal{H}_{F^H}$ and that $\mathcal{H}_{F^H}$ is dense in $\mathcal{C}$.

The operator $\mathbb{K}_{F^H}$ acting on $\mathrm{L}^2(\mathcal{T})$ defined by $(\mathbb{K}_{F^H}f)(t) := \int_0^t F^H(t,s)f(s)\,\mathrm{d}s$ satisfies $\mathbb{K}_{F^H}\mathrm{L}^2(\mathcal{T}) =: \mathcal{H}_{F^H}$ and is surjective. It is also injective on $\mathcal{H}_{F^H}$: let $f \in \mathrm{L}^2(\mathcal{T})$ be a non-zero function such that $(\mathbb{K}_{F^H}f) = 0$ on $\mathcal{T}^*$. Similar to [23], there is an interval $[t_1, t_2] \subset \mathcal{T}$ where f has constant sign. Using previous notation, F^H is defined, for $t \in \mathcal{T}^*$ and $s \in (0, t]$, as $F^H(t,s) := \xi\Gamma^*_{H,t}\mathrm{e}^{\beta(t-s)}$. The function $g(s) := \xi\int_t^s u^{H_-}(u-s)^{H_-}\mathrm{e}^{\beta(t-u)}\,\mathrm{d}u$ is increasing, hence $F^H(t,\cdot)$ is a positive function on $[0,t]$ and $\int_{t_1}^{t_2} F^H(t,s)f(s)\,\mathrm{d}s > 0$, leading to a contradiction. Thus $\mathbb{K}_{F^H}$ is injective on $\mathcal{H}_{F^H}$, hence bijective from $\mathrm{L}^2(\mathcal{T})$ to $\mathcal{H}_{F^H}$. Since $\mathbb{K}_{F^H}$ is a linear operator, one can then define $\langle \mathbb{K}_{F^H}f_1, \mathbb{K}_{F^H}f_2\rangle_{\mathcal{H}_{F^H}} := \langle f_1, f_2\rangle_{\mathrm{L}^2(\mathcal{T})}$ as an inner product.

The second part of the proof consists in showing that $\mathcal{H}_{F^H} := \mathbb{K}_{F^H}\mathrm{L}^2(\mathcal{T})$ is dense in $\mathcal{C}$. We claim that $F^H(t,\cdot)^{-1} \in \mathrm{L}^2(\mathcal{T})$. Indeed, as shown in the proof of [53, Theorem 3.1], the following hold for all $0 < s < t$:

- if $0 < H < \frac{1}{2}$, $F^H(t,s) \geq \xi\kappa_H s^{\frac{1}{2}-H}\tau^{H_-}(t-s)^{H_-} > 0$, leading to

$$\begin{aligned}\int_0^t \frac{\mathrm{d}s}{F^H(t,s)^2} &\leq \frac{1}{\xi^2\kappa_H^2 t^{2H_-}}\int_0^t \left(\frac{s}{t-s}\right)^{2H_-}\mathrm{d}s \\ &\leq \frac{1}{\xi^2\kappa_H^2 t^{4H_-}}\int_0^t s^{2H_-}\,\mathrm{d}s = \frac{t^{2-2H}}{2H\xi^2\kappa_H^2} < \infty;\end{aligned}$$

- if $\frac{1}{2} < H < 1$, $F^H(t,s) \geq \xi\kappa_H \mathrm{e}^{-\beta(s-t)}(t-s)^{H_-}$, so that

$$\begin{aligned}\int_0^t \frac{\mathrm{d}s}{F^H(t,s)^2} &\leq \frac{1}{\xi^2\kappa_H^2\mathrm{e}^{2\beta t}}\int_0^t \frac{\mathrm{e}^{2\beta s}\,\mathrm{d}s}{(t-s)^{2H_-}} \\ &\leq \frac{1}{\xi^2\kappa_H^2\mathrm{e}^{2\beta t}t^{2H_-}}\int_0^t \mathrm{e}^{2\beta s}\,\mathrm{d}s = \frac{1-\mathrm{e}^{-2\beta t}}{2\beta\xi^2\kappa_H^2 t^{2H_-}} < \infty.\end{aligned}$$

Given $a > 0$, take for $s \in [0,t]$, $f_a(s) := \frac{s^a}{F^H(t,s)} \in \mathrm{L}^2(\mathcal{T})$. Then, $\int_0^t F^H(t,s)f_a(s)\,\mathrm{d}s = \frac{t^{a+1}}{a+1}$. Hence, $\mathcal{H}_{F^H}$ contains all polynomials null at the origin, and by the Stone–Weierstrass Theorem, is dense in $\mathcal{C}$. □

A.1.3. *Proof of Theorem 3.1(i).* Let $\overline{Y}^\varepsilon$ denote the solution to the second SDE (3.1), starting from zero. The product rule for (fractional) Brownian motion [44, Theorem 3.1.4] yields, for any $t \geq 0$,

$$\overline{Y}_t^\varepsilon = -\frac{\lambda\varepsilon^b}{\beta}(1-\mathrm{e}^{\beta t}) + \xi\varepsilon^b\int_0^t \mathrm{e}^{\beta(t-u)}\,\mathrm{d}W_u^H. \tag{A.1}$$

Since $\xi\int_0^\cdot \mathrm{e}^{\beta(\cdot-u)}\,\mathrm{d}W_u^H$ is Gaussian, Lemma 3.1 combined with [19, Theorem 3.4.5] yields a large deviations principle on $\mathcal{C}$ for the sequence $(\xi\varepsilon^b\int_0^\cdot \mathrm{e}^{\beta(\cdot-s)}\,\mathrm{d}W_s^H)_{\varepsilon>0}$, with speed ε^{2b} and good rate function Λ_{F^H} as in (2.4). Since the two sequences $(\overline{Y}^\varepsilon)_{\varepsilon>0}$ and $(\xi\varepsilon^b\int_0^\cdot \mathrm{e}^{\beta(\cdot-s)}\,\mathrm{d}W_s^H)_{\varepsilon>0}$ only differ by some deterministic quantity, they are clearly

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

exponentially equivalent; namely, for any $\delta > 0$ and $t \in \mathcal{T}^*$,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon^{2b} \log \mathrm{P}\left[\left\| \overline{Y}^{\varepsilon} - \xi \varepsilon^b \int_0^{\cdot} \mathrm{e}^{\beta(\cdot - s)} \, \mathrm{d}W_s^H \right\|_{\infty} > \delta\right] = \limsup_{\varepsilon \downarrow 0} \varepsilon^{2b} \log \mathrm{P}\left[\left\| \frac{\lambda \varepsilon^b}{\beta} (\mathrm{e}^{\beta \cdot} - 1) \right\|_{\infty} > \delta\right]$$

is equal to $-\infty$ since $\beta < 0$, so that $\overline{Y}^{\varepsilon} \sim \mathrm{LDP}(\varepsilon^{2b}, \Lambda_{F^H})$. Finally, the LDP for Y^{ε} follows again by exponential equivalence:

$$\mathrm{P}(\|Y^{\varepsilon} - \overline{Y}^{\varepsilon}\|_{\infty} > \delta) = \mathrm{P}(\|\varepsilon^b \Theta \mathrm{e}^{\beta \cdot}\|_{\infty} > \delta) \leq \mathrm{P}\left(|\Theta| > \frac{\delta}{\varepsilon^b}\right)$$

for t, ε, $\delta > 0$, and the theorem follows from Assumption $\mathcal{A}_b^{\Theta}$. □

A.1.4. *Proof of Theorem 3.1(ii).* We first prove large deviations for the Gaussian drivers of the process, which we then, by means of iterated contraction principles, translate to large deviations for the whole scaled process.

(1) When $H \neq \frac{1}{2}$, Lemma B.1 in [23] implies that $(\rho B, W^H)$ is Gaussian, so that $\varepsilon^{\mathfrak{b}}(\rho B, W^H)$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, with speed ε^{2b} and good rate function

$$I_1(\varphi, \psi) := \begin{cases} \frac{1}{2} \int_{\mathcal{T}} f(s)^2 \, \mathrm{d}s, & (\varphi, \psi) \in \overline{\mathcal{H}}_H, \\ +\infty, & (\varphi, \psi) \notin \overline{\mathcal{H}}_H, \end{cases}$$

where $\overline{\mathcal{H}}_H$ is the RKHS for $(\rho B, W^H)$ defined as

$$\overline{\mathcal{H}}_H := \left\{ (\varphi, \psi) \in \mathcal{C}^2 : \varphi(\cdot) = \rho \int_0^{\cdot} f(s) \, \mathrm{d}s \right. \\ \left. \text{and } \psi(\cdot) = \int_0^{\cdot} K^H(\cdot, s) f(s) \, \mathrm{d}s, \text{ for some } f \in \mathrm{L}^2(\mathcal{T}) \right\}.$$

By the independence of $B^{\perp}$ with respect to B and W^H and using the contraction principle, $\varepsilon^{\mathfrak{b}}(\overline{\rho} B^{\perp} + \rho B, W^H)$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, with speed ε^{2b} and good rate function

$$I_3(\varphi, \psi) := \inf_{\varphi = u + w} \{I_1(u, \psi) + I_2(w)\} = \inf_{u \in \mathcal{H}} \{I_1(u, \psi) + I_2(\varphi - u)\},$$

where, with $\mathcal{H}_{\overline{\rho}}$ as in (2.3),

$$I_2(w) := \begin{cases} \frac{\overline{\rho}}{2} \int_0^1 f(s)^2 \, \mathrm{d}s, & w \in \mathcal{H}_{\overline{\rho}}, \\ +\infty, & w \notin \mathcal{H}_{\overline{\rho}}. \end{cases}$$

(2) For $H = \frac{1}{2}$, using the proof of Theorem 3.3, $\varepsilon^b(\overline{\rho} B^{\perp} + \rho B, B)'$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$ when ε tends to zero with speed ε^{2b} and good rate function Λ^{ρ} defined in (3.5).

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

We now introduce the process $\widetilde{X}^\varepsilon$ satisfying the following SDE $\mathrm{d}\widetilde{X}^\varepsilon_t = -\frac{1}{2}\widetilde{\sigma}^2(Y^\varepsilon_t)\mathrm{d}t + \varepsilon^b\widetilde{\sigma}(Y^\varepsilon_t)(\rho\mathrm{d}B_t + \overline{\rho}\mathrm{d}B^\perp_t)$, and translate this large deviations into one for the sequence $\widetilde{X}^\varepsilon$. Since the proof for the $H \neq \frac{1}{2}$ case is the same as the one for $H = \frac{1}{2}$, albeit with a different rate function, we concentrate on the former case. Since $Y^\varepsilon_t = \varepsilon^b \mathrm{e}^{\beta t}\Theta - \frac{\lambda\varepsilon^b}{\beta}(1-\mathrm{e}^{\beta t}) + \varepsilon^b\xi\int_0^t \mathrm{e}^{\beta(t-u)}\,\mathrm{d}W^H_u$, one can define a continuous map $G((0,1)\times\mathcal{C}\to\mathcal{C})$, such that $(\varepsilon^b(\rho B + \overline{\rho}B^\perp), Y^\varepsilon) = (\varepsilon^b(\rho B + \overline{\rho}B^\perp), G(\varepsilon, \varepsilon^b W^H)(t))$. Moreover, using $\mathcal{A}^\Theta_b$, $G(\varepsilon, \varepsilon^b W^H)$ is exponentially equivalent to $\overline{G}(\varepsilon^b W^H)$, with $\overline{G}(\varphi)(t) := \xi\int_0^t \mathrm{e}^{\beta(t-u)}\,\mathrm{d}\varphi_u$. The contraction principle thus yields $(\varepsilon^b(\rho B + \overline{\rho}B^\perp), Y^\varepsilon) \sim \mathrm{LDP}(\varepsilon^{2b}, I_4)$, with $I_4 : (\varphi,\psi)\mapsto \inf\{I_3(\varphi, v) \mid \psi\in\mathcal{C} : \psi = \overline{G}(v)\}$.

Under $\mathcal{A}'$, the extended contraction principle [40, Proposition 2.3] yields that $(\varepsilon^b(\rho B + \overline{\rho})B^\perp, \widetilde{\sigma}(Y^\varepsilon)) \sim \mathrm{LDP}(\varepsilon^{2b}, I_5)$ with $I_5(\varphi,\psi) := \inf\{I_4(\varphi, v) \mid \psi = \widetilde{\sigma}(v)\} = \inf\{I_3(\varphi, v) \mid \psi = \widetilde{\sigma}(\overline{G}(v))\}$. Finally, setting $b \geq \frac{1}{2}$, the sequence of semi-martingales $(\varepsilon^b W)$ is uniformly exponentially tight (UET) in the sense of [28, Definition 1.1], and the sequence $(\widetilde{\sigma}(Y^\varepsilon))$ is continuous from the right with no left limits (càdlàg; Assumption $\mathcal{A}$), and adapted to the filtration $\mathcal{F}$. Denoting by $X\cdot Y := \int X\,\mathrm{d}Y$ the stochastic integral with respect to a semi-martingale, Theorem 1.2 in [28] yields a large deviations principle on $\mathcal{C}(\mathcal{T},\mathbb{R}^3)$ for $(\varepsilon^b(\rho B_s + \overline{\rho}B^\perp_s), \widetilde{\sigma}(Y^\varepsilon), \varepsilon^b\int_0^\cdot \widetilde{\sigma}(Y^\varepsilon_s)(\rho\mathrm{d}B_s + \overline{\rho}\mathrm{d}B^\perp_s))$, with speed ε^{2b} and good rate function I_6 defined as $I_6(\phi) := I_5(\varphi,\psi)$ if $\phi = \varphi\cdot\psi$ and $\psi\in\mathrm{BV}$ (the space of functions of finite variation), and infinity otherwise. Since, for $t\in\mathcal{T}^*$,

$$\widetilde{X}^\varepsilon_t = -\frac{1}{2}\int_0^t \widetilde{\sigma}(Y^\varepsilon_s)^2\,\mathrm{d}s + \int_0^t \varepsilon^b\widetilde{\sigma}(Y^\varepsilon_s)\,\mathrm{d}W_s =: I(\widetilde{\sigma}(Y^\varepsilon), \widetilde{\sigma}(Y^\varepsilon)\cdot\varepsilon^b(\rho B_s + \overline{\rho}B^\perp_s))(\varepsilon, t),$$

the contraction principle then yields $\widetilde{X}^\varepsilon \sim \mathrm{LDP}(\varepsilon^{2b}, \widetilde{\Lambda})$ with

$$\widetilde{\Lambda}(\phi) := \inf\{I_6(\chi) \mid \phi = I(\varphi, \varphi\cdot\psi),\ \chi = (\varphi,\psi),\ \psi\in\mathrm{BV}\}. \tag{A.2}$$

It remains to prove that X^ε and $\widetilde{X}^\varepsilon$ are exponentially equivalent. Indeed, for $t\in\mathcal{T}^*$,

$$\begin{aligned}\left\|X^\varepsilon - \widetilde{X}^\varepsilon\right\|_\infty \leq \frac{1}{2}\Big\|\int_0^\cdot \Big|\varepsilon^{2b}\sigma^2\Big(\frac{Y^\varepsilon_s}{\varepsilon^b}\Big) - \widetilde{\sigma}(Y^\varepsilon_s)\Big|\mathrm{d}s\Big\|_\infty \\ + \varepsilon^b\Big\|\int_0^\cdot \Big|\varepsilon^b\sigma\Big(\frac{Y^\varepsilon_s}{\varepsilon^b}\Big) - \widetilde{\sigma}(Y^\varepsilon_s)\Big|\mathrm{d}(\rho B_s + \overline{\rho}B^\perp_s)\Big\|_\infty.\end{aligned}$$

Using the linear growth assumption as well as Assumption $\mathcal{A}$, we have, for $s\in[0,t]$,

$$\begin{aligned}\Big|\varepsilon^{2b}\sigma^2\Big(\frac{Y^\varepsilon_s}{\varepsilon^b}\Big) - \widetilde{\sigma}(Y^\varepsilon_s)\Big| &\leq 2C^2\Big\{\varepsilon^{2b}\Big[1 + \Big|\frac{Y^\varepsilon_s}{\varepsilon^b}\Big|^2\Big] + 1 + |Y^\varepsilon_s|^2\Big\} \\ &\leq 2C^2[1 + \varepsilon^{2b} + 2|Y^\varepsilon_s|^2],\end{aligned}$$

and hence

$$\begin{aligned}\Big\|\int_0^\cdot \Big|\varepsilon^{2b}\sigma^2\Big(\frac{Y^\varepsilon_s}{\varepsilon^b}\Big) - \widetilde{\sigma}(Y^\varepsilon_s)\Big|\mathrm{d}s\Big\|_\infty &\leq \Big\|\int_0^\cdot 2C^2(1+\varepsilon^{2b} + 2|Y^\varepsilon_s|^2)\mathrm{d}s\Big\|_\infty \\ &\leq 2C^2(1+\varepsilon^{2b}) + 4C^2\Big\|\int_0^\cdot |Y^\varepsilon_s|^2\,\mathrm{d}s\Big\|_\infty.\end{aligned}$$

Since $Y^\varepsilon_s = \varepsilon^b\mathrm{e}^{\beta s}\Theta + \varepsilon^b\frac{\lambda}{\beta}(\mathrm{e}^{\beta s}-1) + \varepsilon^b\xi[W^H_s + \beta\int_0^s W^H_u\mathrm{e}^{-\beta u}\,\mathrm{d}u]$, we obtain, for $s\in\mathcal{T}$, $|Y^\varepsilon_s| \leq \varepsilon^b\mathrm{e}^{\beta s}|\Theta| + \varepsilon^b\frac{\lambda}{\beta}(\mathrm{e}^{\beta s}-1) + \xi\varepsilon^b|W^H_s|$, because $\beta<0$, and

$$|Y^\varepsilon_s|^2 \leq 2(\varepsilon^{2b}\mathrm{e}^{2\beta s}\Theta^2 + 2\varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2(\mathrm{e}^{2\beta s}+1) + \xi^2\varepsilon^{2b}|W^H_s|^2), \tag{A.3}$$

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

so that

$$\int_0^t |Y_s^\varepsilon|^2 \,\mathrm{d}s \le 2\Big(\varepsilon^{2b} t\Theta^2 + \varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2\Big(2t - \frac{1}{\beta}\Big) + \xi^2\varepsilon^{2b}\int_0^t |W_s^H|^2 \,\mathrm{d}s\Big).$$

Introduce

$$\mathcal{J} := 2C^2\Big\|\int_0^\cdot |Y_s^\varepsilon|^2 \,\mathrm{d}s\Big\|_\infty \qquad \text{and} \qquad \Xi_s^\varepsilon := \varepsilon^b \sigma\Big(\frac{Y_s^\varepsilon}{\varepsilon^b}\Big) - \widetilde{\sigma}(Y_s^\varepsilon).$$

For any $\delta, a_0 > 0$, we can then write

$$\begin{aligned}
\mathrm{P}[\|X^\varepsilon - \widetilde{X}^\varepsilon\|_\infty > \delta] &\le \mathrm{P}\Big[\mathcal{J} + \varepsilon^b\Big\|\int_0^\cdot \Xi_s^\varepsilon \mathrm{d}(\rho B_s + \overline{\rho} B_s^\perp)\Big\|_\infty > \delta_\varepsilon\Big] \\
&\le \mathrm{P}\Big[\mathcal{J} + \varepsilon^b\Big\|\int_0^\cdot \Xi_s^\varepsilon \mathrm{d}(\rho B_s + \overline{\rho} B_s^\perp)\Big\|_\infty > \delta_\varepsilon \,\Big|\, \mathcal{J} < a_0\Big](1 - \mathfrak{P}) \\
&\quad + \mathrm{P}\Big[\mathcal{J} + \varepsilon^b\Big\|\int_0^\cdot \Xi_s^\varepsilon \mathrm{d}(\rho B_s + \overline{\rho} B_s^\perp)\Big\|_\infty > \delta_\varepsilon \,\Big|\, \mathcal{J} \ge a_0\Big]\mathfrak{P} \\
&\le \mathrm{P}\Big[\Big\|\int_0^\cdot \Xi_s^\varepsilon \mathrm{d}(\rho B_s + \overline{\rho} B_s^\perp)\Big\|_\infty > \frac{\delta_\varepsilon - a_0}{\varepsilon^b} \,\Big|\, \mathcal{J} < a_0\Big] + \mathfrak{P},
\end{aligned}$$

with $\mathfrak{P} := \mathrm{P}(\mathcal{J} \ge a_0)$, $\delta_\varepsilon := \delta - C^2(1 + \varepsilon^{2b})$. Let $\mathcal{J}_t := \int_0^t |W_s^H|^2 \,\mathrm{d}s$. By (A.3), we have

$$\begin{aligned}
\mathrm{P}(\mathcal{J} > \bar{\delta}) &\le \mathrm{P}\Big(\Big|\varepsilon^{2b}\Theta^2 + \varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2\Big(2 - \frac{1}{\beta}\Big) + \xi^2\varepsilon^{2b}\mathcal{J}_1\Big| > \frac{\bar{\delta}}{2}\Big) \\
&\le \mathrm{P}\Big(\varepsilon^{2b}\Theta^2 + \varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2\Big|2 - \frac{1}{\beta}\Big| + \xi^2\varepsilon^{2b}\mathcal{J}_1 > \frac{\bar{\delta}}{2}\Big|\xi^2\varepsilon^{2b}\mathcal{J}_1 < a\Big) \\
&\quad \times \mathrm{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 < a) \\
&\quad + \mathrm{P}\Big(\varepsilon^{2b}\Theta^2 + \varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2\Big|2 - \frac{1}{\beta}\Big| + \xi^2\varepsilon^{2b}\mathcal{J}_1 > \frac{\bar{\delta}}{2}\Big|\xi^2\varepsilon^{2b}\mathcal{J}_1 \ge a\Big) \\
&\qquad \times \mathrm{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \ge a) \\
&\le \mathrm{P}\Big(\Theta^2 > \frac{1}{\varepsilon^{2b}}\Big(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2\Big|2 - \frac{1}{\beta}\Big| - a\Big)\Big|\xi^2\varepsilon^{2b}\mathcal{J}_1 < a\Big) \\
&\quad + \mathrm{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \ge a) \\
&\le \mathrm{P}\Big(\Big|\Theta\Big| > \frac{1}{\varepsilon^b}\sqrt{\Big(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2\Big|2 - \frac{1}{\beta}\Big| - a\Big)}\Big|\xi^2\varepsilon^{2b}\mathcal{J}_1 < a\Big) \\
&\quad + \mathrm{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \ge a),
\end{aligned}$$

for any $\bar{\delta} > 0$ and $a > 0$. One can then use Markov's inequality:

$$\mathrm{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \ge a) \le \frac{\mathrm{E}[\xi^2\varepsilon^{2b}\mathcal{J}_1]}{a} = \frac{\xi^2\varepsilon^{2b}}{a}\mathrm{E}\Big(\int_0^1 |W_s^H|^2 \,\mathrm{d}s\Big).$$

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

Since $\mathbb{V}(W_s^H) = s^{2H}$, one obtains $\mathrm{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \geq a) \leq \frac{\xi^2\varepsilon^{2b}}{a}\frac{1}{2H+1}$. Hence there exist $\eta \in [0,\,1)$ and $\bar{\varepsilon} > 0$ such that for all $\varepsilon \leq \bar{\varepsilon}$, $\mathrm{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 < a) \geq 1-\eta$. Assumption $\mathcal{A}_b^\Theta$ then implies that there exist $M,\,\tilde{\varepsilon} > 0$ such that

$$\mathrm{P}\Big(|\Theta| > \frac{1}{\varepsilon^b}\sqrt{\Big(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2\Big|2-\frac{1}{\beta}\Big| - a\Big)}\Big) \leq \exp\Big(-\frac{M}{\varepsilon^{2b}}\Big), \qquad \text{for all } \varepsilon \leq \tilde{\varepsilon}.$$

Thus, for all $\varepsilon \leq \min\{\tilde{\varepsilon},\,\bar{\varepsilon}\}$, Bayes' Theorem yields

$$\begin{aligned}\mathrm{P}\Big(|\Theta| > \frac{1}{\varepsilon^b}\sqrt{\Big(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2\Big|2-\frac{1}{\beta}\Big| - a\Big)}\Big|\xi^2\varepsilon^{2b}\mathcal{J}_1 < a\Big) \\ \leq \frac{\mathrm{P}\Big(|\Theta| > \frac{1}{\varepsilon^b}\sqrt{\Big(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2\Big|2-\frac{1}{\beta}\Big| - a\Big)}\Big)}{\mathrm{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 < a)} \leq \frac{\exp\big(-\frac{M}{\varepsilon^{2b}}\big)}{1-\eta}.\end{aligned}$$

Finally, for all $\varepsilon \leq \min\{\tilde{\varepsilon},\,\bar{\varepsilon}\}$,

$$\begin{aligned}\mathrm{P}(\mathcal{J} > \bar{\delta}) &\leq \mathrm{P}\Big(|\Theta| > \frac{1}{\varepsilon^b}\sqrt{\Big(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\Big(\frac{\lambda}{\beta}\Big)^2\Big|2-\frac{1}{\beta}\Big| - a\Big)}\Big|\xi^2\varepsilon^{2b}\mathcal{J}_1 < a\Big) \\ &\quad + \mathrm{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \geq a) \\ &\leq \frac{\mathrm{e}^{-M/\varepsilon^{2b}}}{1-\eta} + \frac{\xi^2\varepsilon^{2b}}{a}\frac{1}{2H_+}, \end{aligned} \tag{A.4}$$

so that for all $\varepsilon \leq \min\{\tilde{\varepsilon},\,\bar{\varepsilon}\}$, $\mathrm{P}(\mathcal{J} \geq a_0) \leq \frac{\mathrm{e}^{-M/\varepsilon^{2b}}}{1-\eta} + \frac{\xi^2\varepsilon^{2b}}{a}\frac{1}{2H_+}$. Hence, there exist $\breve{\eta} \in [0,\,1)$, $\breve{\varepsilon} < \tilde{\varepsilon} \wedge \bar{\varepsilon}$ such that $\mathrm{P}(\mathcal{J} < a_0) \geq 1-\breve{\eta}$ for all $\varepsilon \leq \breve{\varepsilon}$. By the uniform convergence assumed in $\mathcal{A}$, for all $y \in \mathbb{R}$ there exist $N,\,\varepsilon_0 > 0$ such that $|\varepsilon^b\sigma(\frac{Y_s^\varepsilon}{\varepsilon^b}) - \widetilde{\sigma}(Y_s^\varepsilon)| \leq N$ for all $\varepsilon < \varepsilon_0$. The Borell–TIS inequality [47, Proposition A.1] then yields

$$\begin{aligned}&\mathrm{P}\Big(\Big\|\int_0^\cdot \Big(\varepsilon^b\sigma\Big(\frac{Y_s^\varepsilon}{\varepsilon^b}\Big) - \widetilde{\sigma}(Y_s^\varepsilon)\Big)\mathrm{d}(\rho B_s + \overline{\rho}B_s^\perp)\Big\|_\infty > \frac{1}{\varepsilon^b}(\delta - C^2(1+\varepsilon^{2b}) - a_0)\Big) \\ &\leq 2\exp\Big\{-\frac{(\delta - C^2(1+\varepsilon^{2b}) - a_0)^2}{2N^2\varepsilon^{2b}}\Big\}.\end{aligned}$$

Thus, for all $\varepsilon \leq \min\{\breve{\varepsilon},\,\varepsilon_0\}$, Bayes' Theorem yields

$$\begin{aligned}&\mathrm{P}\Big(\Big\|\int_0^\cdot \Xi_s^\varepsilon\mathrm{d}(\rho B_s + \overline{\rho}B_s^\perp)\Big\|_\infty > \frac{\delta - C^2(1+\varepsilon^{2b}) - a_0}{\varepsilon^b}\Big|\mathcal{J} < a_0\Big) \\ &\leq \frac{\mathrm{P}\Big(\big\|\int_0^\cdot \Xi_s^\varepsilon\mathrm{d}(\rho B_s + \overline{\rho}B_s^\perp)\big\|_\infty > \frac{\delta - C^2(1+\varepsilon^{2b}) - a_0}{\varepsilon^b}\Big)}{\mathrm{P}(\mathcal{J} < a_0)} \\ &\leq \frac{2}{1-\breve{\eta}}\exp\Big\{-\frac{(\delta - C^2(1+\varepsilon^{2b}) - a_0)^2}{2N^2\varepsilon^{2b}}\Big\}.\end{aligned}$$

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

Finally, for all $\varepsilon \le \min\{\breve{\varepsilon}, \varepsilon_0\}$,

$$\mathrm{P}\left(\left\|X^\varepsilon - \widetilde{X}^\varepsilon\right\|_\infty > \delta\right) \le \frac{2\exp\left\{-\frac{(\delta - C^2(1+\varepsilon^{2b}) - a_0)^2}{2N^2\varepsilon^{2b}}\right\}}{1-\breve{\eta}} + \frac{\mathrm{e}^{-M/\varepsilon^{2b}}}{1-\eta} + \frac{\xi^2\varepsilon^{2b}}{2aH_+},$$

so that X^ε and $\widetilde{X}^\varepsilon$ are exponentially equivalent and $X^\varepsilon \sim \mathrm{LDP}(\varepsilon^{2b}, \widetilde{\Lambda})$. □

Remark A.1. For computational purposes, such infinite-dimensional optimisation problems are discretised over a set of basis functions, such as orthonormal polynomials. If we then consider φ continuously differentiable and I_1 finite, we can simplify $I_1(\varphi) = I_1(\varphi, \psi) = \frac{1}{2\rho^2}\int_{\mathcal{T}} [\dot{\varphi}(t)]^2\,\mathrm{d}t$. Further, if $u \in \mathcal{C}^1$, then $I_4(\varphi) = I_3(\varphi)$ and

$$I_3(\varphi) \simeq \frac{1}{2}\inf_{u \in \mathcal{H}\cap\mathcal{C}^1}\left\{\int_0^t \left(\frac{\dot{u}_s}{\rho}\right)^2 \mathrm{d}s + \frac{1}{\overline{\rho}}\int_0^t (\dot{\varphi}_s - \dot{u}_s)^2\,\mathrm{d}s\right\}.$$

A.2. Proof of Theorem 3.4

A.2.1. *Proof of Theorem 3.4(i).* Let us introduce the process $\overline{Y}^\varepsilon$ defined, for $t \in \mathcal{T}^*$, as the unique solution of $\mathrm{d}\overline{Y}^\varepsilon_t = (\varepsilon^{b+2}\lambda + \beta\varepsilon^2\overline{Y}^\varepsilon_t)\mathrm{d}\tau + \varepsilon^{2H+b}\xi\,\mathrm{d}W^H_t$, with $\overline{Y}^\varepsilon_0 = 0$, i.e.

$$\overline{Y}^\varepsilon_t = -\frac{\lambda\varepsilon^b}{\beta}(1 - \mathrm{e}^{\beta\varepsilon^2 t}) + \xi\varepsilon^{2H+b}\int_0^t \mathrm{e}^{\beta\varepsilon^2(t-u)}\,\mathrm{d}W^H_u.$$

By Lemma 3.1, Proposition 3.2, and [19, Theorem 3.4.5], the sequence

$$\left(\varepsilon^{2H+b}\int_0^\cdot G^H_0(\cdot, s)\,\mathrm{d}Z_s\right)_{\varepsilon>0}$$

satisfies an LDP on $\mathcal{C}$, with speed ε^{4H+2b} and good rate function $\Lambda_{G^H_0}$ as in (2.4). Also:

(1) When $H \neq \frac{1}{2}$, $(\varepsilon^{2H+b}\int_0^t G^H_\varepsilon(t, s)\,\mathrm{d}Z_s)_{\varepsilon>0}$ and $(\varepsilon^{2H+b}\int_0^t G^H_0(t, s)\,\mathrm{d}Z_s)_{\varepsilon>0}$ are exponentially equivalent: indeed, the asymptotic expansion of the Gaussian density [1, Formula (26.2.12)] yields, for any $\delta > 0$,

$$\begin{aligned}
&\mathrm{P}\left[\left\|\varepsilon^{2H+b}\int_0^\cdot G^H_\varepsilon(\cdot, s)\,\mathrm{d}Z_s - \varepsilon^{2H+b}\int_0^\cdot G^H_0(\cdot, s)\,\mathrm{d}Z_s\right\|_\infty > \delta\right] \\
&= \mathrm{P}\left[\left\|\int_0^\cdot (G^H_\varepsilon(\cdot, s) - G^H_0(\cdot, s))\mathrm{d}Z_s\right\|_\infty > \frac{\delta}{\varepsilon^{2H+b}}\right] = \mathrm{P}\left[|\mathcal{N}(0,1)| > \frac{\delta\varepsilon^{-2H-b}}{\|V_\varepsilon\|_\infty}\right] \\
&= \sqrt{\frac{2}{\pi}}\frac{\|V_\varepsilon\|_\infty\varepsilon^{2H+b}}{\delta}\exp\left(-\frac{\delta^2}{2\varepsilon^{4H+2b}\|V_\varepsilon\|^2_\infty}\right)(1 + \mathcal{O}(\varepsilon^{4H+2b})),
\end{aligned}$$

with $V^2_\varepsilon := \int_0^t (G^H_\varepsilon(t, s) - G^H_0(t, s))^2\mathrm{d}s$; note that $\lim_{\varepsilon\downarrow 0} V^2_\varepsilon = 0$. Now, for $\varepsilon > 0$ and $0 < s < t$,

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

(i) if $H < \frac{1}{2}$,

$$\begin{aligned}[G_\varepsilon^H(t,s) - G_0^H(t,s)]^2 &\le \frac{2\kappa_H^2\xi^2}{s^{2H_-}}\Big\{{H_-}^2\bigg(\int_s^t \frac{(u-s)^{H_-}}{u^{1-H_-}}(\mathrm{e}^{\beta\varepsilon^2(t-u)}-1)\mathrm{d}u\bigg)^2 \\ &\quad + \beta^2\varepsilon^4\bigg(\int_s^t (u-s)^{H_-}u^{H_-}\mathrm{e}^{\beta\varepsilon^2(t-u)}\,\mathrm{d}u\bigg)^2\Big\} \\ &\le 2\kappa_H^2\xi^2\bigg[\frac{{H_-}^2}{s^{2H_-}}\bigg(\int_s^t \frac{(u-s)^{H_-}}{u^{1-H_-}}\,\mathrm{d}u\bigg)^2 \\ &\quad + \frac{\beta^2\varepsilon^4}{s^{2H_-}}\bigg(\int_s^t (u(u-s))^{H_-}\,\mathrm{d}u\bigg)^2\bigg]\end{aligned}$$

Using [53, Lemma A.3], we further obtain, for $\varepsilon > 0$ and $0 < s < t$, $V_\varepsilon^2 \le 2\kappa_H^2\xi^2[C_H t^{2H} + \beta^2\varepsilon^4\widetilde{C}_H t^{2H+2}]$, with $C_H, \widetilde{C}_H > 0$. Hence,

$$0 < \varepsilon^{4H+2b}V_\varepsilon^2 \le 2\kappa_H^2\xi^2\Big(C_H\varepsilon^{4H+2b}t^{2H} + \beta^2\varepsilon^{4H+2b+4}\widetilde{C}_H t^{2H+2}\Big);$$

(ii) if $H > \frac{1}{2}$,

$$\begin{aligned}[G_\varepsilon^H(t,s) - G_0^H(t,s)]^2 &= \frac{\kappa_H^2\xi^2}{s^{2H_-}}{H_-}^2\bigg(\int_s^t \frac{u^{H_-}}{(u-s)^{1-H_-}}(\mathrm{e}^{\beta\varepsilon^2(t-u)}-1)\,\mathrm{d}u\bigg)^2 \\ &\le \frac{\kappa_H^2\xi^2}{s^{2H_-}}{H_-}^2\bigg(\int_s^t \frac{u^{H_-}\,\mathrm{d}u}{(u-s)^{1-H_-}}\bigg)^2 \\ &\le \frac{\kappa_H^2\xi^2}{s^{2H_-}}{H_-}^2 t^{2H_-}\bigg(\int_s^t (u-s)^{H_--1}\,\mathrm{d}u\bigg)^2 \\ &= \frac{\kappa_H^2\xi^2}{s^{2H_-}}[t(t-s)]^{2H_-}.\end{aligned}$$

Hence, for $H \ne \frac{1}{2}$, as $b > 0$,

$$\lim_{\varepsilon\downarrow 0}\varepsilon^{4H+2b}\|V_\varepsilon\|_\infty^2 = 0 \text{ and } \lim_{\varepsilon\downarrow 0}\varepsilon^{2H+b}\|V_\varepsilon\|_\infty = 0.$$

(2) When $H = \frac{1}{2}$, $(\xi\varepsilon^{1+b}\int_0^t \mathrm{e}^{\beta\varepsilon^2(t-s)}\,\mathrm{d}B_s)_{\varepsilon>0}$ and $(\xi\varepsilon^{1+b}B_t)_{\varepsilon>0}$ are exponentially equivalent: the asymptotic expansion of the Gaussian density near infinity [1, Formula (26.2.12)] yields, for any $\delta > 0$,

$$\begin{aligned}&\mathrm{P}\bigg(\bigg\|\xi\int_0^\cdot \mathrm{e}^{\beta\varepsilon^2(\cdot-s)}\,\mathrm{d}B_s - \xi\int_0^\cdot \mathrm{d}B_s\bigg\|_\infty > \frac{\delta}{\varepsilon^{1+b}}\bigg) \\ &= \mathrm{P}\bigg(\bigg\|\int_0^\cdot \xi(\mathrm{e}^{\beta\varepsilon^2(\cdot-s)}-1)\mathrm{d}B_s\bigg\|_\infty > \frac{\delta}{\varepsilon^{1+b}}\bigg) \\ &= \mathrm{P}\Big(|\mathcal{N}(0,1)| > \frac{\delta}{\|V_\varepsilon\|_\infty\varepsilon^{1+b}\xi}\Big) \\ &= \sqrt{\frac{2}{\pi}}\frac{\|V_\varepsilon\|_\infty\varepsilon^{1+b}\xi}{\delta}\exp\bigg(-\frac{\delta^2}{2\varepsilon^{2+2b}\|V_\varepsilon\|_\infty^2\xi^2}\bigg)(1+\mathcal{O}(\varepsilon^{2+2b})),\end{aligned}$$

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

with

$$V_\varepsilon^2 := \int_0^t [(\mathrm{e}^{\beta\varepsilon^2(t-s)} - 1)]^2 \,\mathrm{d}s = t + \frac{\mathrm{e}^{2\beta\varepsilon^2 t} - 1}{2\beta\varepsilon^2} + \frac{2(1 - \mathrm{e}^{\beta\varepsilon^2 t})}{\beta\varepsilon^2}.$$

Hence $0 < \varepsilon^{2+2b} V_\varepsilon^2 \le \varepsilon^{2+2b} t + \varepsilon^{2b} \frac{\mathrm{e}^{\beta\varepsilon^2 t} - 1}{2\beta}$, and, as $b > 0$, $\lim_{\varepsilon \downarrow 0} \varepsilon^{2+2b} \|V_\varepsilon\|_\infty^2 = 0$. Also, note that $\lim_{\varepsilon \downarrow 0} V_\varepsilon^2 = 0$. The required exponential equivalence then follows, and hence for all $H \in (0, 1)$, $(\varepsilon^{2H+b} \int_0^\cdot G_\varepsilon^H(\cdot, s)\,\mathrm{d}Z_s)_{\varepsilon>0} \sim \mathrm{LDP}(\varepsilon^{4H+2b}, \Lambda_{G_0^H})$.

The final step is exponential equivalence between $(\varepsilon^{2H+b} \int_0^\cdot G_\varepsilon^H(\cdot, s)\,\mathrm{d}Z_s)_{\varepsilon>0}$ and $\overline{Y}^\varepsilon$. For any $\delta, t > 0$,

$$\begin{aligned}
&\mathrm{P}\left(\left\| \overline{Y}^\varepsilon - \varepsilon^{2H+b} \int_0^\cdot G_\varepsilon^H(\cdot, s)\,\mathrm{d}Z_s \right\|_\infty > \delta\right) \\
&= \mathrm{P}\left(\left\| \overline{Y}^\varepsilon - \varepsilon^{2H+b} \int_0^\cdot \mathrm{e}^{\beta\varepsilon^2(\cdot - s)}\,\mathrm{d}W_s^H \right\|_\infty > \delta\right) \\
&= \mathrm{P}\left(\left\| - \frac{\lambda\varepsilon^b}{\beta}(1 - \mathrm{e}^{\beta\varepsilon^2 \cdot}) \right\|_\infty > \delta\right).
\end{aligned}$$

Hence,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon^{4H+2b} \log \mathrm{P}\left(\left\| \overline{Y}^\varepsilon - \varepsilon^{2H+b} \int_0^\cdot G_\varepsilon(\cdot, s)\,\mathrm{d}Z_s \right\|_\infty > \delta\right) = -\infty,$$

yielding exponential equivalence and $\overline{Y}^\varepsilon \sim \mathrm{LDP}(\varepsilon^{4H+2b}, \Lambda_{G_0^H})$. Finally, since

$$\mathrm{P}(\|Y^\varepsilon - \overline{Y}^\varepsilon\|_\infty > \delta) \le \mathrm{P}(\varepsilon^b |\Theta| > \delta),$$

for any $\delta > 0$ as $\beta < 0$, and $\overline{Y}^\varepsilon$ and Y^ε are exponentially equivalent, and $Y^\varepsilon \sim \mathrm{LDP}(\varepsilon^{4H+2b}, \Lambda_{G_0^H})$ by Assumption $\mathcal{A}_b^\Theta$. □

A.2.2. *Proof of Theorem 3.4(ii).* The proof is similar to that of Theorem 3.1(ii).

(1) $H \ne \frac{1}{2}$: we have already established that $\varepsilon^{2H+b}(\overline{\rho}B^\perp + \rho B, W^H)$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, with speed ε^{4H+2b} and good rate function I_3 defined earlier.

(2) $H = \frac{1}{2}$: using the proof of Theorem 3.3, $\varepsilon^{b+1}(\overline{\rho}B^\perp + \rho B, B)'$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, with speed ε^{2+2b} and good rate function Λ^ρ defined in (3.5).

Similar to above, we only prove the case $H \ne \frac{1}{2}$, the other one being analogous. We introduce the process $\widetilde{X}^\varepsilon$ satisfying the SDE $\mathrm{d}\widetilde{X}_t^\varepsilon = -\frac{\varepsilon^{2H+1}}{2}\widetilde{\sigma}^2(Y_t^\varepsilon)\mathrm{d}t + \varepsilon^{2H+b}\widetilde{\sigma}(Y_t^\varepsilon)(\rho\,\mathrm{d}B_t + \overline{\rho}\,\mathrm{d}B_t^\perp)$, and translate this large deviations case into one for the sequence $\widetilde{X}^\varepsilon$. Using that for $\varepsilon > 0$ and $t \in \mathcal{T}^*$, $Y_t^\varepsilon = \varepsilon^b \mathrm{e}^{\beta\varepsilon^2 t}\Theta - \frac{\lambda\varepsilon^b}{\beta}(1 - \mathrm{e}^{\beta\varepsilon^2 t}) + \varepsilon^{2H+b}\xi \int_0^t \mathrm{e}^{\beta\varepsilon^2(t-u)}\,\mathrm{d}W_u^H$, one can define a continuous map $\widetilde{G}((0, 1) \times \mathcal{C} \to \mathcal{C})$ such that $(\varepsilon^{2H+b}(\rho B + \overline{\rho}B^\perp), Y^\varepsilon) = (\varepsilon^{2H+b}(\rho B + \overline{\rho}B^\perp), \widetilde{G}(\varepsilon, \varepsilon^{2H+b}W^H)(t))$. Using $\mathcal{A}_b^\Theta$ and the asymptotic expansion of the Gaussian density near infinity [1, Formula (26.2.12)], $\widetilde{G}(\varepsilon, \varepsilon^{2H+b}W^H)$ is exponentially equivalent to $\widehat{G}_0(\varepsilon^{2H+b}W^H)$, defined by $\widehat{G}_0(\varphi)(t) := \xi \int_0^t \mathrm{d}\varphi_u$.

Hence, the contraction principle yields an LDP on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$ for $(\varepsilon^{2H+b}(\rho B + \overline{\rho}B^\perp), Y^\varepsilon)$, with speed ε^{4H+2b} and good rate function $\widetilde{I}_4 : (\varphi, \psi) \mapsto \inf\{I_3(\varphi, \nu) \mid \psi \in \mathcal{C} : \psi = \widehat{G}_0(\nu)\}$.

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

Under $\mathcal{A}'$, the extended contraction principle [40, Proposition 2.3] yields an LDP on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$ for $(\varepsilon^{2H+b}(\rho B + \overline{\rho} B^{\perp}), \widetilde{\sigma}(Y^{\varepsilon}))$, with speed ε^{4H+2b} and good rate function $\widetilde{I}_5 : (\varphi, \psi) \mapsto \inf\{\widetilde{I}_4(\varphi, \nu) \mid \psi = \widetilde{\sigma}(\nu)\} = \inf\{I_3(\varphi, \nu) \mid \psi = \widetilde{\sigma}(\widehat{G}_0(\nu))\}$. Finally, setting $b \geq \frac{1}{2} - 2H$, the sequence of semi-martingales $(\varepsilon^{2H+b} W)$ is UET and the sequence $(\widetilde{\sigma}(Y^{\varepsilon}))$ is càdlàg (Assumption $\mathcal{A}$), and adapted to the filtration $\mathcal{F}$. Theorem 1.2 in [28] thus gives an LDP on $\mathcal{C}$ for the sequence of stochastic integrals $\varepsilon^{2H+b} \int_0^{\cdot} \widetilde{\sigma}(Y_s^{\varepsilon})(\rho \mathrm{d}B_s + \overline{\rho} \mathrm{d}B_s^{\perp})$, with speed ε^{4H+2b} and good rate function

$$\mathtt{I}(\chi) := \inf\left\{\widetilde{I}_5(\varphi, \psi) : \varphi \cdot \psi = \chi, \ \psi \in \mathrm{BV}\right\}. \tag{A.5}$$

We now prove that $\widetilde{X}^{\varepsilon}$ and $\varepsilon^{2H+b} \int_0^{\cdot} \widetilde{\sigma}(Y_s^{\varepsilon})(\rho \mathrm{d}B_s + \overline{\rho} \mathrm{d}B_s^{\perp})$ are exponentially equivalent.

$$\begin{aligned}
\mathrm{P}\left(\left\| \widetilde{X}^{\varepsilon} - \varepsilon^{2H+b} \int_0^{\cdot} \widetilde{\sigma}(Y_s^{\varepsilon}) \, \mathrm{d}W_s \right\|_{\infty} > \delta\right) &= \mathrm{P}\left(\left\| -\frac{1}{2} \int_0^{\cdot} (\varepsilon^{H+\frac{1}{2}} \widetilde{\sigma}(Y_s^{\varepsilon}))^2 \, \mathrm{d}s \right\|_{\infty} > \delta\right) \\
&= \mathrm{P}\left(\left\| \int_0^{\cdot} (\widetilde{\sigma}(Y_s^{\varepsilon}))^2 \, \mathrm{d}s \right\|_{\infty} > \frac{2\delta}{\varepsilon^{2H+1}}\right) \\
&\leq \mathrm{P}\left(\left\| \int_0^{\cdot} |Y_s^{\varepsilon}|^2 \, \mathrm{d}s \right\|_{\infty} > \frac{\delta}{C^2 \varepsilon^{2H+1}} - 1\right),
\end{aligned}$$

using the growth condition assumption $|\widetilde{\sigma}(x)| \leq C(1 + |x|)$ for $x \in \mathbb{R}$. Since $Y_s^{\varepsilon} = \varepsilon^b \mathrm{e}^{\beta \varepsilon^2 s} \Theta + \varepsilon^b \frac{\lambda}{\beta}(\mathrm{e}^{\beta \varepsilon^2 s} - 1) + \varepsilon^{2H+b} \xi [W_s^H + \beta \varepsilon^2 \int_0^s W_u^H \mathrm{e}^{-\beta \varepsilon^2 u} \, \mathrm{d}u]$ and $\beta < 0$, we obtain, for $s \in \mathcal{T}$,

$$|Y_s^{\varepsilon}| \leq \varepsilon^b \mathrm{e}^{\beta \varepsilon^2 s} |\Theta| + \varepsilon^b \frac{\lambda}{\beta}(\mathrm{e}^{\beta \varepsilon^2 s} - 1) + \xi \varepsilon^{2H+b} |W_s^H|,$$

so that

$$|Y_s^{\varepsilon}|^2 \leq 2\left(\varepsilon^{2b} \mathrm{e}^{2\beta \varepsilon^2 s} \Theta^2 + 2\varepsilon^{2b} \left(\frac{\lambda}{\beta}\right)^2 \left(\mathrm{e}^{2\beta \varepsilon^2 s} + 1\right) + \xi^2 \varepsilon^{4H+2b} |W_s^H|^2\right),$$

and hence

$$\int_0^t |Y_s^{\varepsilon}|^2 \, \mathrm{d}s \leq \varepsilon^{2b} t \Theta^2 + \varepsilon^{2b} \left(\frac{\lambda}{\beta}\right)^2 \left(2t - \frac{1}{\beta \varepsilon^2}\right) + \xi^2 \varepsilon^{4H+2b} \int_0^t |W_s^H|^2 \, \mathrm{d}s.$$

Using a similar argument to (A.4), used in the proof of Theorem 3.1, we have, for all $\varepsilon \leq \min\{\tilde{\varepsilon}, \bar{\varepsilon}\}$,

$$\mathrm{P}\left(\left\| \int_0^{\cdot} |Y_s^{\varepsilon}|^2 \, \mathrm{d}s \right\|_{\infty} > \frac{\delta}{C^2 \varepsilon^{2H_+}} - 1\right) \leq \frac{\mathrm{e}^{-M/\varepsilon^{4H+2b}}}{1 - \eta} + \frac{\xi^2 \varepsilon^{4H+2b}}{a} \frac{1}{2H_+},$$

so that $\widetilde{X}^{\varepsilon}$ and $\varepsilon^{2H+b} \int_0^{\cdot} \widetilde{\sigma}(Y_s^{\varepsilon})(\rho \mathrm{d}B_s + \overline{\rho} \mathrm{d}B_s^{\perp})$ are exponentially equivalent.

Finally, one can use the same arguments as in Appendix A.1.3, replacing β by $\beta \varepsilon^2$ and the speed ε^{2b} by ε^{4H+2b}, to prove that X^{ε} and $\widetilde{X}^{\varepsilon}$ are exponentially equivalent. Hence $X^{\varepsilon} \sim \mathrm{LDP}(\varepsilon^{4H+2b}, \mathtt{I})$. □

A.3. Proof of Theorem 3.6

The proof is similar to that of [47, Theorem 2.9], with some minor modifications. One key ingredient in the proof of the theorem is an exponential equivalence between $(\mathrm{X}^{\varepsilon, \mathrm{x}})_{\varepsilon > 0}$ and its limit system (as ε tends to zero). In this proof, we denote X^{ε} by $\mathrm{X}^{\varepsilon, \mathrm{x}}$ to stress the dependence on the starting point.

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

Lemma A.1. *Under Assumption 3.3, for $R, \delta, \beta > 0$, there exist $\gamma, \rho, \varepsilon_0 > 0$ such that*

$$\varepsilon^2 \log \mathrm{P}(\|X^{\varepsilon,x} - \mathcal{S}_0(\varphi, u)\| > \delta, \|\varepsilon W - \varphi\| \le \gamma) \le -R$$

holds when $\varepsilon \le \varepsilon_0$ for all $u \in supp(X_0)$, $\varphi \in \mathcal{H}$ satisfying $\Lambda^\rho(\varphi) \le \beta$, $x \in \mathcal{B}_\rho((0,0))$.

Proof. Define the Radon–Nikodym derivative

$$\mathrm{D}^\varepsilon(\varphi) := \exp\left\{\frac{1}{\varepsilon}\int_{\mathcal{T}} \dot{\varphi}_s \,\mathrm{dW}_s - \frac{1}{2\varepsilon^2}\int_{\mathcal{T}} \|\dot{\varphi}_s\|^2 \,\mathrm{d}s\right\}.$$

Following [47], consider the family $(\overline{\mathrm{Y}}^\varepsilon)_{\varepsilon>0}$ of (unique strong) solutions of $\mathrm{d}\overline{\mathrm{Y}}^\varepsilon_t = \mathfrak{c}(\varepsilon, t, \overline{\mathrm{Y}}^\varepsilon_t, \mathrm{X}_0)\mathrm{d}t + \varepsilon\mathfrak{a}(\overline{\mathrm{Y}}^\varepsilon_t, \mathrm{X}_0)\mathrm{dB}^\varepsilon_t$ for $t > 0$ with initial condition $\overline{\mathrm{Y}}^\varepsilon_0 = (0,0)' \in \mathbb{R}^2$. We denote $\mathfrak{c}(\varepsilon, t, \mathrm{x}, \mathrm{y}) := \mathfrak{b}(\varepsilon, \mathrm{x}, \mathrm{y}) + \mathfrak{a}(\mathrm{x}, \mathrm{y})\dot{f}_t$ for $\mathrm{x}, \mathrm{y} \in \mathbb{R}^2$, $t \in \mathcal{T}$, and $\varepsilon > 0$, and $\mathrm{B}^\varepsilon_t := W_t - \varepsilon^{-1}\dot{f}_t$ for $t \in \mathcal{T}$. Peithmann's assumptions are updated here as:

(i) $\mathfrak{c} : (0,\infty) \times \mathcal{T} \times \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ converges to $\mathfrak{c}_0 : \mathcal{T} \times \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ in the sense that

$$\lim_{\varepsilon\downarrow 0}\int_{\mathcal{T}} \sup_{\mathrm{x}\in\mathbb{R}^2} \|\mathfrak{c}(\varepsilon, t, \mathrm{x}, \mathrm{y}) - \mathfrak{c}_0(t, \mathrm{x}, \mathrm{y})\| \,\mathrm{d}t = 0, \qquad \text{for all } \mathrm{y} \in \mathbb{R}^2.$$

In particular, note that $\mathfrak{c}_0(t, \mathrm{x}, \mathrm{y}) = \mathfrak{a}(\mathrm{x}, \mathrm{y})\dot{f}_t$.

(ii) There exists $\varrho \in \mathrm{L}^2(\mathcal{T})$ such that for all $\mathrm{x}, \mathrm{y} \in \mathbb{R}^2$, $\|\mathfrak{c}(\varepsilon, t, \mathrm{x}, \mathrm{y})\| + \|\mathfrak{c}_0(t, \mathrm{x}, \mathrm{y})\| \le \varrho(t)$, for $t \in \mathcal{T}$.

(iii) There exists $\kappa \in \mathrm{L}^1(\mathcal{T})$ such that for all $\mathrm{y}, \mathrm{x}_1, \mathrm{x}_2 \in \mathbb{R}^2$, $\|\mathfrak{c}_0(t, \mathrm{x}_1, \mathrm{y}) - \mathfrak{c}_0(t, \mathrm{x}_2, \mathrm{y})\| \le \kappa(t)\|\mathrm{x}_1 - \mathrm{x}_2\|$ on $\mathcal{T}$.

Partitioning $\mathcal{T}$ into $\{t_k := \frac{kT}{n}\}_{k=0,\dots,n}$, set $\overline{\mathrm{Y}}^{\varepsilon,n}_t := \overline{\mathrm{Y}}^\varepsilon_{t_k}$ for $t_k \le t < t_{k+1}$; Peithmann's arguments yield

(i) For any $\delta > 0$, $\lim_{n\uparrow\infty} \limsup_{\varepsilon\downarrow 0} \varepsilon^2 \log \mathrm{P}(\|\overline{Y}^\varepsilon - \overline{\mathrm{Y}}^{\varepsilon,n}\| > \delta) = -\infty$ uniformly with respect to $\mathrm{y}_0 \in \mathbb{R}^2$.

(ii) Given $M^\varepsilon_t := \varepsilon\int_{\mathcal{T}} \mathfrak{a}(\overline{\mathrm{Y}}^\varepsilon_s, \mathrm{X}_0)\mathrm{d}\,\mathrm{W}_s$ and $M^{\varepsilon,n}_t := \varepsilon\int_{\mathcal{T}} \mathfrak{a}(\overline{\mathrm{Y}}^{\varepsilon,n}_s, \mathrm{X}_0)\mathrm{d}\,\mathrm{W}_s$, for all $\delta > 0$,

$$\lim_{\beta\downarrow 0}\limsup_{\varepsilon\downarrow 0} \varepsilon^2 \log \mathrm{P}(\|M^\varepsilon - M^{\varepsilon,n}\| > \delta, \|\mathrm{Y}^\varepsilon - Y^{\varepsilon,n}\| \le \beta) = -\infty,$$

uniformly with respect to $n \in \mathbb{N}$, $y_0 \in \mathbb{R}^2$.

(iii) For all $\delta > 0$, $\lim_{\gamma\downarrow 0} \limsup_{\varepsilon\downarrow 0} \varepsilon^2 \log \mathrm{P}(\|M^\varepsilon\| > \delta, \|\varepsilon\mathrm{W}\| \le \gamma) = -\infty$.

(iv) Let ζ denote the solution of the ODE $\dot{\zeta}_t = \mathfrak{c}_0(t, \zeta_t, Y_t)$ starting from $\zeta_0 = \mathrm{y}$. For all $R > 0$, $\delta > 0$, there exist $\gamma, \rho, \varepsilon_0 > 0$ such that $\mathrm{P}(\{\|\mathrm{Y}^\varepsilon, \zeta\| > \delta\} \cap \{\|\varepsilon\mathrm{W}\| \le \gamma\}) \le \exp(-R/\varepsilon^2)$ for all $\mathrm{y}_0 \in \mathbb{R}^2$, $\mathrm{y} \in \mathcal{B}_\rho(\mathrm{y}_0)$, and $\varepsilon \le \varepsilon_0$.

The theorem follows from Girsanov's Theorem with the Radon–Nikodym derivative $\mathrm{D}^\varepsilon(f)$. □

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

We start with the lower bound. For any open subset G of $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, let $\eta > 0$ and choose $\psi \in G$ such that $I^\alpha(\psi) \le \inf_{\psi \in G} I^\alpha(\psi) + \eta$. Then, let $u \in \mathrm{supp}(\mathrm{X}_0)$ and $\varphi \in \mathcal{H}$ such that $\mathcal{S}_0(\varphi, u) = \psi$ and $\Lambda^\rho(\varphi) = I^\alpha(\psi)$. Let $\delta > 0$ such that $\mathcal{B}_\delta(\psi) \subset G$. Then, for each $\gamma > 0$, $\mathrm{x} \in \mathbb{R}^2$,

$$\begin{aligned}\mathrm{P}(\mathrm{X}^{\varepsilon,\mathrm{x}} \in G) &\ge \mathrm{P}(\|\mathrm{X}^{\varepsilon,\mathrm{x}} - \psi\| \le \delta) \\ &\ge \mathrm{P}(\|\varepsilon \mathrm{W} - \varphi\| \le \gamma) - \mathrm{P}(\|\mathrm{X}^{\varepsilon,\mathrm{x}} - \psi\| > \delta,\, \|\varepsilon \mathrm{W} - \varphi\| \le \gamma).\end{aligned}$$

Schilder's Theorem [16, Theorem 5.2.3] then yields the lower bound

$$\liminf_{\varepsilon \downarrow 0} \varepsilon^2 \mathrm{P}(\|\varepsilon \mathrm{W} - \varphi\| \le \gamma) \ge -\Lambda^\rho(\varphi) = -I^\alpha(\psi) \ge -\inf_{\psi \in G} I^\alpha(\psi) - \eta.$$

Then, we bound the second probability from above using Lemma A.1: fix $\beta \ge \Lambda^\rho(\varphi)$ and $R > \inf_{\psi \in G} I^\alpha(\psi) + \eta$, and find γ, ρ, $\varepsilon_0 > 0$ such that, for $\mathrm{x} \in \mathcal{B}_\rho((0,0)')$, $\varepsilon \le \varepsilon_0$, the bound $\varepsilon^2 \log \mathrm{P}(\|\mathrm{X}^{\varepsilon,\mathrm{x}} - \psi\| > \delta,\, \|\varepsilon \mathrm{W} - \varphi\| \gamma) \le -R$ holds. These two bounds then imply the required lower bound:

$$\liminf_{\varepsilon \downarrow 0} \liminf_{\rho \downarrow 0} \varepsilon^2 \log \inf_{\mathrm{x} \in \mathcal{B}_\rho((0,0)')} \mathrm{P}(\mathrm{X}^{\varepsilon,\mathrm{x}} \in G) \ge \min\left\{ -R, -\inf_{\psi \in G} I^\alpha(\psi) - \eta \right\} = -\inf_{\psi \in G} I^\alpha(\psi) - \eta.$$

We now prove the upper bound. For any closed set F of $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, take $\beta \in (0, \inf_{\psi \in G} I^\alpha(\psi))$ and $R > \beta$. Let $u \in \mathrm{supp}(\mathrm{X}_0)$ and $\psi \in \mathcal{H}$ with $I^\alpha(\psi) \le \beta$. We find $\delta > 0$ such that $\mathcal{B}_\delta(\psi) \cap F = \varnothing$ and $\varphi \in \{\Lambda^\rho \le \beta\}$ such that $\mathcal{S}_0(\varphi, u) = \psi$. Using Lemma A.1, there exist γ, ρ, $\varepsilon_0 > 0$ such that for $\mathrm{x} \in \mathcal{B}_\rho((0,0)')$, $\varepsilon \le \varepsilon_0$,

$$\varepsilon^2 \log \mathrm{P}(\|\mathrm{X}^{\varepsilon,\mathrm{x}} - \psi\| > \delta,\, \|\varepsilon \mathrm{W} - \varphi\| \le \gamma) \le -R.$$

The set $\{\mathcal{B}_\gamma(\varphi) : \psi \in \mathcal{H}, I^\alpha(\psi) \le \beta\}$ forms a cover of the compact set $\{\Lambda^\rho(\varphi) \le \beta\}$, so that we can extract a finite sub-cover $\{\mathcal{B}_{\gamma_i}(\varphi_i)\}_{i=1,\dots,k}$ and set $A := \cup_{i=1}^k \mathcal{B}_{\gamma_i}(f_i)$ and $\psi_i := \mathcal{S}_0(\varphi_i, r)$. For any $i = 1, \dots, k$, there exist δ_i, ρ_i, $\varepsilon_i > 0$ such that, for any $\mathrm{x} \in \mathcal{B}_{\rho_i}((0,0)')$ and $\varepsilon \le \varepsilon_i$,

$$\varepsilon^2 \log \mathrm{P}(\|\mathrm{X}^{\varepsilon,\mathrm{x}} - \psi_i\| > \delta,\, \|\varepsilon \mathrm{W} - \varphi_i\| \le \gamma) \le -R.$$

Set $\varepsilon_0 := \min\{\varepsilon_1, \dots, \varepsilon_k\}$, $\rho_0 := \min\{\rho_1, \dots, \rho_k\}$, take $\varepsilon \le \varepsilon_0$ and $\mathrm{x} \in \mathcal{B}_{\rho_0}((0,0)')$. Since $F \cap \mathcal{B}_{\delta_i}(\psi_i) = \varnothing$ for every $i = 1, \dots, k$, we obtain

$$\begin{aligned}\mathrm{P}(\mathrm{X}^{\varepsilon,\mathrm{x}} \in F) &\le \mathrm{P}(\mathrm{X}^{\varepsilon,\mathrm{x}} \in F,\, \varepsilon \mathrm{W} \in A) + \mathrm{P}(\varepsilon \mathrm{W} \in A^c) \\ &\le \sum_{i=1}^k \mathrm{P}(\|\mathrm{X}^{\varepsilon,\mathrm{x}} - g_i\| > \delta_i,\, \|\varepsilon \mathrm{W} - \varphi_i\| \le \gamma_i) + \mathrm{P}(\varepsilon \mathrm{W} \in A^c) \\ &\le k \exp\left(-\frac{R}{\varepsilon^2}\right) + \exp\left(-\frac{\beta}{\varepsilon^2}\right),\end{aligned}$$

since $I^\alpha(\psi_i) \le \beta$ by definition of β. Finally, since $R > \beta$, the theorem follows from the upper bound

$$\limsup_{\varepsilon \downarrow 0} \varepsilon^2 \log \sup_{\rho \downarrow 0} \sup_{\mathrm{x} \in \mathcal{B}_\rho(\mathrm{X}_0^\varepsilon)} \mathrm{P}(\mathrm{X}^{\varepsilon,\mathrm{x}} \in F) \le -\inf_{\psi \in F} I^\alpha(\psi). \qquad \square$$

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

A.4. Proof of Corollary 4.1

This is a straightforward application of [27, Corollary 7.1]. Taking ε^{-2b} to be k we have, from Theorem 3.1, that $\varepsilon^{2b}X_t \sim \mathrm{LDP}(\varepsilon^{2b}, \widetilde{\Lambda})$ as ε goes to zero. Then,

$$\lim_{k\uparrow\infty} \frac{1}{k} \log \mathrm{P}(X_t \geq k) = -\inf_{y\geq 1} \widetilde{\Lambda}(\phi)|_{\phi_t=y}.$$

Similarly, in Black–Scholes, $\lim_{k\uparrow\infty} \dfrac{\Sigma_t^2(k)}{k^2} \log \mathrm{P}(X_t \geq k) = -\dfrac{1}{2t}$, and the proof follows from [27, Corollary 7.1]. □

A.5. Proof of Corollary 4.2

This is a straightforward application of [27, Corollary 7.1]. Taking ε^2 to be t and $b_H := H + b - 1/2$, from Theorem 3.4, $t^{b_H}X \sim \mathrm{LDP}(t^{2H+b}, \mathtt{I})$; as t goes to zero,

$$\lim_{t\downarrow 0} t^{2H+b} \log \mathrm{P}(t^{b_H}X \geq k) = -\inf_{y\geq k} \mathtt{I}(\phi)|_{\phi_1=y}.$$

Similarly, in the Black–Scholes model we obtain

$$\lim_{t\downarrow 0} \frac{t\Sigma_t^2(t^{-b_H}k)}{t^{-2b_H}} \log \mathrm{P}(t^{b_H}X \geq k) = \lim_{t\downarrow 0} \frac{t\Sigma_t^2(t^{-b_H}k)}{t^{-2b_H}} \log \mathrm{P}(X \geq t^{-b_H}k) = -\frac{k^2}{2},$$

and the result follows from [27, Corollary 7.1]. □

A.6. Proof of Proposition 4.1

For any $\varepsilon > 0$, the pathwise rescaled process $\mathrm{X}^{\varepsilon,\zeta}$ satisfies (4.1). The proof of the proposition relies on the (more general) theorem proved by Millet, Nualart, and Sanz [43] that was recalled in Section 3.1.2, and whose validity is guaranteed by Assumption 3.1. Note first that, from standard large deviations considerations (and in particular contraction mappings), the process $\mathrm{X}^{\varepsilon,0}$ satisfies a large deviations principle with good rate function I given in Theorem 3.2. Recall that by construction $\mathfrak{b}(\varepsilon, \cdot) : \mathbb{R}^2 \to \mathbb{R}^2$ satisfies $\lim_{\varepsilon\downarrow 0} \varepsilon\widetilde{\mathfrak{b}}(\varepsilon, \mathrm{x}/\varepsilon) = \mathfrak{b}(\mathrm{x})$ uniformly as ε tends to zero. Therefore Theorem 3.2 yields a large deviations principle for the sequence $(\mathrm{X}^{\varepsilon,\zeta})_{\varepsilon\geq 0}$ as ε tends to zero, with good rate function I and speed ε^2. In particular, for $A := \{\psi \in \mathcal{C}(\mathcal{T}, \mathbb{R}^n), \forall \mathrm{x} \in \mathbb{R}^n, \psi(1, \mathrm{x}) \geq 1\}$,

$$-\inf_{\psi\in A} I(\psi) \leq \liminf_{\varepsilon\downarrow 0} \varepsilon^2 \log \mathrm{P}(\mathrm{X}_t^{\varepsilon,\zeta} \geq 1) \leq \limsup_{\varepsilon\downarrow 0} \varepsilon^2 \log \mathrm{P}(\mathrm{X}_t^{\varepsilon,\zeta} \geq 1) \leq -\inf_{\psi\in\bar{A}} I(\psi),$$

$$-\inf_{\psi\in A} I(\psi) \leq \liminf_{\varepsilon\downarrow 0} \varepsilon^2 \log \mathrm{P}(\mathrm{X}_t^{\varepsilon,\mathrm{x}_0} \geq 1) \leq \limsup_{\varepsilon\downarrow 0} \varepsilon^2 \log \mathrm{P}(\mathrm{X}_t^{\varepsilon,\mathrm{x}_0} \geq 1) \leq -\inf_{\psi\in\bar{A}} I(\psi).$$

Since Λ is continuous on $\mathcal{H}$, it is upper semi-continuous on A, and so is the good rate function I by [2, Lemma 2.41]. As a good rate function, it is also lower semi-continuous, and hence continuous, on A. Translating the two sets of inequalities above in terms of $\varepsilon\mathrm{X}_t^{\zeta}$ and $\varepsilon\mathrm{X}_t^{\mathrm{x}_0}$, and using the continuity of I, proves the proposition. □

Acknowledgements

The authors are grateful to the anonymous referees for their useful comments. BH acknowledges financial support from the SNSF Early Postdoc. Mobility grant 165248.

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

References

[1] ABRAMOWITZ, M. AND STEGUN, I.A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.

[2] ALIPRANTIS, C. D. AND BORDER, K. C. (2005). *Infinite Dimensional Analysis, a Hitchhiker's Guide*, 3rd edn. Springer, New York.

[3] ALÒS, E., LEÓN, J. AND VIVES, J. (2007). On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance Stoch.* **11,** 571–589.

[4] AZENCOTT, R. (1980). Grandes déviations et applications. *Lecture Notes Math.* **774,** 1–176.

[5] BAUDOIN, F. AND OUYANG, C. (2011). Small-time kernel expansion for solutions of stochastic differential equations driven by fractional Brownian motions. *Stoch. Process. Appl.* **121,** 759–792.

[6] BAUDOIN, F. AND OUYANG, C. (2015). On small-time asymptotics for rough differential equations driven by fractional Brownian motions. In *Large Deviations and Asymptotic Methods in Finance*, Springer Proc. Math. Stats., Vol. **110**.

[7] BAYER, C., FRIZ, P. K. AND GATHERAL, J. (2016). Pricing under rough volatility. *Quant. Finance* **16,** 887–904.

[8] BAYER, C. *et al.* Short-time near-the-money skew in rough fractional volatility models. To appear in *Quant. Finance*.

[9] BEN AROUS, G. (1988). Methods de Laplace et de la phase stationnaire sur l'espace de Wiener. *Stochastics* **25,** 125–153.

[10] BEZUIDENHOUT, C. (1987). A large deviations principle for small perturbations of random evolution equations. *Ann. Prob.* **15,** 646–658.

[11] BISMUT, J. M. (1984). *Large Deviations and the Malliavin Calculus*. Birkhauser, Basel.

[12] CARAVENNA, F. AND CORBETTA, J. (2018). The asymptotic smile of a multiscaling stochastic volatility model. *Stoch. Process. Appl.* **128,** 1034–1071.

[13] CARMONA, R. AND TEHRANCHI, M. R. (2006). *Interest Rate Models: An Infinite-Dimensional Stochastic Analysis Perspective*. Springer, New York.

[14] COMTE, F. AND RENAULT, E. (1998). Long memory in continuous-time stochastic volatility models. *Math. Finance* **8,** 291–323.

[15] COMTE, F., COUTIN, L. AND RENAULT, E. (2012). Affine fractional stochastic volatility models. *Ann. Finance* **8,** 337–378.

[16] DEMBO, A. AND ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, Vol. 2. Springer, New York.

[17] DEUSCHEL, J. D., FRIZ, P. K., JACQUIER, A. AND VIOLANTE, S. (2014). Marginal density expansions for diffusions and stochastic volatility, part I: Theoretical foundations. *Commun. Pure Appl. Math.* **67,** 40–82.

[18] DEUSCHEL, J. D., FRIZ, P. K., JACQUIER, A. AND VIOLANTE, S. (2014). Marginal density expansions for diffusions and stochastic volatility, part II: Applications. *Commun. Pure Appl. Math.* **67,** 321–350.

[19] DEUSCHEL, J. D. AND STROOCK, D. (2001). *Large Deviations*. Vol. 342, American Mathematical Society, Providence, RI.

[20] EL EUCH, O., FUKASAWA, M. AND ROSENBAUM, M. (2018). The microstructural foundations of leverage effect and rough volatility. *Finance Stoch.* **22,** 241–280.

[21] FORDE, M. AND JACQUIER, A. (2009). Small-time asymptotics for implied volatility under the Heston model. *IJTAF* **12,** 861–876.

[22] FORDE, M., JACQUIER, A. AND LEE, R. (2012). The small-time smile and term structure of implied volatility under the Heston model. *SIAM J. Financial Math.* **3,** 690–708.

[23] FORDE, M. AND ZHANG, H. (2017). Asymptotics for rough stochastic volatility models. *SIAM J. Financial Math.* **8,** 114–145.

[24] FRIEDMAN, A. (1975). *Stochastic Differential Equations and Applications*, Vol. 1. Academic Press, New York.

[25] FREIDLIN, M. AND WENTZELL, A. (1984). *Random Perturbations of Dynamical Systems*. Vol. 260, Springer, New York.

[26] FUKASAWA, M. (2017). Short-time at-the-money skew and rough fractional volatility. *Quant. Finance* **17,** 189–198.

[27] GAO, K. AND LEE, R. (2014). Asymptotics of implied volatility to arbitrary order. *Finance Stoch.* **18,** 349–392.

[28] GARCIA, J. (2008). A large deviation principle for stochastic integrals. *J. Theoret. Prob.* **21,** 476–501.

[29] GATHERAL, J. *The Volatility Surface: A Practitioner's Guide*. John Wiley, New York.

[30] GATHERAL, J., JAISSON, T. AND ROSENBAUM, M. (2018). Volatility is rough. *Quant. Finance* **18,** 933–949.

[31] GENIN, A. AND TANKOV, P. (2019). Optimal importance sampling for Lévy processes. To appear in *Stoch. Process. Appl.*

[32] GUASONI, P. AND ROBERTSON, S. (2008). Optimal importance sampling with explicit formulas in continuous time. *Finance Stoch.* **12,** 1–19.

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press

[33] GULISASHVILI, A., VIENS, F. AND ZHANG, X. (2018). Small-time asymptotics for Gaussian self-similar stochastic volatility models. *Appl. Math. Optimization*, 1–41.
[34] JACQUIER, A., KELLER-RESSEL, M. AND MIJATOVIĆ, A. (2013). Large deviations and stochastic volatility with jumps: Asymptotic implied volatility for affine models. *Stochastics* **85,** 321–345.
[35] JACQUIER, A., MARTINI, C. AND MUGURUZA, A. (2018). On VIX futures in the rough Bergomi model. *Quant. Finance* **18,** 45–61.
[36] JACQUIER, A., PAKKANEN, M. AND STONE, H. (2018). Pathwise large deviations for the rough Bergomi model. *J. Appl. Prob.* **55,** 1078–1092.
[37] JACQUIER, A. AND ROOME, P. (2013). Asymptotics of forward implied volatility. *SIAM J. Financial Math.* **6,** 307–351.
[38] JACQUIER, A. AND ROOME, P. (2018). Black–Scholes in a CEV random environment: A new approach to smile modelling. *Math. Financial Econom.* **12,** 445–474.
[39] JACQUIER, A. AND SHI, F. (2019). The randomised Heston model. *SIAM J. Financial Math.* **10,** 89–129.
[40] MAYER-WOLF, E., NUALART, D. AND PÉREZ-ABREU, V. (1992). Large deviations for multiple Wiener–Itô integral processes. *Séminaire de Probabilités* **26,** 11–31.
[41] MECHKOV, S. (2016). 'Hot-start' initialization of the Heston model. *Risk*, November.
[42] MELLOUK, M. (2000). A large-deviation principle for random evolution equations. *Bernoulli* **6,** 977–999.
[43] MILLET, A., NUALART, D. AND SANZ, M. (1992). Large deviations for a class of anticipating stochastic differential equations. *Ann. Prob.* **20,** 1902–1931.
[44] MISHURA, Y. S. (2008). *Stochastic Calculus for Fractional Brownian Motion and Related Processes*. Springer, New York.
[45] NORROS, I., VALKEIKA, E. AND VIRTAMO, J. (1999). An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. *Bernoulli* **5,** 571–587.
[46] PHAM, H. (2010). Large deviations in finance. In *Proc. 3rd SMAI European Summer School in Financial Mathematics*.
[47] PEITHMANN, D. (2007). *Large Deviations and Exit Time Asymptotics for Diffusions and Stochastic Resonance*. PhD thesis, Humboldt University, Berlin.
[48] ROBERTSON, S. (2010). Sample path large deviations and optimal importance sampling for stochastic volatility models. *Stoch. Process. Appl.* **120,** 66–83.
[49] SCHILDER, M. (1966). Asymptotic formulas for Wiener integrals. *Trans. Amer. Math. Soc.* **125,** 63–85.
[50] SCHÖBEL, R. AND ZHU, J. (1999). Stochastic volatility with an Ornstein–Uhlenbeck process: An extension. *Rev. Finance* **3,** 23–46.
[51] STEIN, E. AND STEIN, J. (1991). Stock-price distributions with stochastic volatility – an analytic approach. *Rev. Financial Studies* **4,** 727–752.
[52] VARADHAN, S. R. S. (1967). On the behaviour of the fundamental solution of the heat equation with variable coefficients. *Commun. Pure Appl. Math.* **20** 431–455.
[53] YAN, L., LU, Y. AND XU, Z. (2008). Some properties of the fractional Ornstein–Uhlenbeck process. *J. Phys. A: Math. and Theor.* **41,** 1–17.

https://doi.org/10.1017/jpr.2019.27 Published online by Cambridge University Press