

# PATHWISE MODERATE DEVIATIONS FOR OPTION PRICING

ANTOINE JACQUIER AND KONSTANTINOS SPILIOPOULOS

**ABSTRACT.** We provide a unifying treatment of pathwise moderate deviations for models commonly used in financial applications, and for related integrated functionals. Suitable scaling allows us to transfer these results into small-time, large-time and tail asymptotics for diffusions, as well as for option prices and realised variances. In passing, we highlight some intuitive relationships between moderate deviations rate functions and their large deviations counterparts; these turn out to be useful for numerical purposes, as large deviations rate functions are often difficult to compute.

## 1. INTRODUCTION

We develop a unifying framework for pathwise moderate deviations of (multiscale) Itô diffusions, with applications to small-time, large-time and tail asymptotics for the diffusions and related integrated functionals. The original motivation is a pathwise extension of the moderate deviations proved in [15] in the context of small-time option pricing only. More specifically, we consider a generic two-dimensional stochastic volatility model  $\mathbb{X} = (X, Y)$  of the form

$$(1) \quad \begin{aligned} dX_t &= -\frac{1}{2}\sigma^2(Y_t)dt + \sigma(Y_t)dW_t, \\ dY_t &= f(Y_t)dt + g(Y_t)dZ_t, \\ d\langle W, Z \rangle_t &= \rho dt, \end{aligned}$$

with starting point  $(X_0, Y_0) = (x_0, y_0)$ . For  $0 < \varepsilon, \delta \leq 1$ , the rescaled process  $\mathbb{X}^\varepsilon$  defined by

$$(2) \quad \mathbb{X}_t^\varepsilon := (X_t^\varepsilon, Y_t^\varepsilon) := \left( \delta X_{\frac{\varepsilon}{\delta^2}t}, Y_{\frac{\varepsilon}{\delta^2}t} \right),$$

is a solution of the multiscale system

$$(3) \quad \begin{aligned} dX_t^\varepsilon &= -\frac{1}{2}\frac{\varepsilon}{\delta}\sigma^2(Y_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(Y_t^\varepsilon)dW_t, \\ dY_t^\varepsilon &= \frac{\varepsilon}{\delta^2}f(Y_t^\varepsilon)dt + \frac{\sqrt{\varepsilon}}{\delta}g(Y_t^\varepsilon)dZ_t, \end{aligned}$$

starting from  $\mathbb{X}_0^\varepsilon = (\delta x_0, y_0)$ , again with  $d\langle W, Z \rangle_t = \rho dt$ . One reasonably expects that if  $\delta = 1$ , then small  $\varepsilon$  asymptotics for (3) correspond to small-time asymptotics for (1). Equivalently, if we allow both  $\varepsilon$  and  $\delta$  to tend to zero such that  $\lim_{\varepsilon \downarrow 0} \frac{\varepsilon}{\delta} = \gamma \in (0, \infty)$ , then small  $(\varepsilon, \delta)$  asymptotics for (3) correspond to scaled large-time asymptotics for (1). Hence, one can describe small- and large-time asymptotics for (1) via small  $\varepsilon$  asymptotics for (3) by appropriately choosing  $\delta$ . Such asymptotic results can then be translated to corresponding estimates for option pricing, a quantity of interest in mathematical finance. In addition, one more quantity of interest in financial applications is related to asymptotics of integrated functionals of the form  $\int_0^t H(X_s^\varepsilon, Y_s^\varepsilon)ds$  for appropriate functions  $H$ . In particular,  $H(x, y) = y$  corresponds to the so-called realised variance. The focus of this paper is to quantify such approximations via the lens of moderate deviations (MD), and we shall analyse three situations.

First, considering  $\bar{\mathbb{X}}_t := \lim_{\varepsilon \downarrow 0} \mathbb{X}_t^\varepsilon$  (in some appropriate sense) and a positive function  $h(\cdot)$  such that  $\lim_{\varepsilon \downarrow 0} h(\varepsilon) = \infty$  with  $\lim_{\varepsilon \downarrow 0} \sqrt{\varepsilon}h(\varepsilon) = 0$ , we are interested in the large deviations properties of

$$\eta_t^\varepsilon := \frac{\mathbb{X}_t^\varepsilon - \bar{\mathbb{X}}_t}{\sqrt{\varepsilon}h(\varepsilon)},$$

or equivalently moderate deviations properties for  $\mathbb{X}^\varepsilon$ , and, as a corollary, for  $X^\varepsilon$ , and their consequences on estimates for option prices on  $e^X$ . Moderate deviations for  $X^\varepsilon$  have been developed in the literature

---

*Date:* March 12, 2018.

K.S. was partially supported by NSF DMS 1550918.

both in the  $\delta = 1$  case [1, 13], and in the  $\varepsilon, \delta \downarrow 0$  case [1, 13, 18, 26]; we make here precise the connection of those results to small- and large-time moderate deviations results respectively for (1), and provide exact consequences for option pricing. The mathematical finance literature has seen an explosion of applications of large deviations over the past decade [7, 8, 11, 12, 14, 22, 29], but moderate deviations have only recently been introduced by [15]. However, [15] only considers small-time behaviours, which are not pathwise and are based on assuming that  $X$  admits a continuous density expansion at all times with a specific behaviour for small times. Our methodology here allows us to derive pathwise results without the need for such assumptions.

Second, we obtain moderate deviations principle for some integrated functionals, more precisely large deviations for quantities of the form

$$R^\varepsilon := \frac{1}{\sqrt{\varepsilon h(\varepsilon)}} \int_0^\cdot (H(X_s^\varepsilon, Y_s^\varepsilon) - \overline{H}(X_s^\varepsilon)) ds,$$

for some function  $H$ , where  $\overline{H}$  centers  $H$  around its mean behaviour with respect to the invariant distribution of the process  $Y$ . We also allow the coefficients of the SDE (1) to depend on both  $(x, y)$  and allow certain growth with respect to  $y$  (namely we do not impose global boundedness assumptions). This constitutes the main theoretical result of this paper, presented in Section 4 and proven in Sections 6-7. For example, when  $H(x, y) = y$ , this corresponds to asymptotics of the realised variance and yields estimates for options thereon. Related moderate deviations results have also been established in [20], but with  $H$  uniformly bounded, which is insufficient for many applications. Our proof is based on the weak convergence approach of [9] through stochastic control representations, which then allow us to consider growth in the coefficients of both the underlying SDE models and on  $H$ .

Finally, we show how to apply the pathwise moderate deviations to models used in quantitative finance, and derive asymptotics for option prices. In passing, we obtain interesting connections between the large deviations and the moderate deviations rate functions. The latter in fact characterises the local curvature of the large deviations rate function around its minimum. Even though this is to be intuitively expected, it shows that one can use moderate deviations to get an approximation to large deviations asymptotics, and is useful in practice as the moderate deviations rate function is usually available in closed form as a quadratic, whereas the large deviations rate function is often not explicit. Lastly, we mention that even though the results in this paper are mostly presented for  $\mathbb{X} \in \mathbb{R}^2$ , the actual theorems hold in any finite dimension, as in [1, 5, 13, 26]. We restrict ourselves here to this setting, solely because of the financial motivation.

The rest of the paper is organised as follows. In Section 2 we review the small-noise pathwise moderate deviation results from the literature, and connect them to small-time moderate deviations asymptotics for financial models. In Section 3 we review the small-noise moderate deviations for slow-fast systems (3), with  $\lim_{\varepsilon \downarrow 0} \frac{\varepsilon}{\delta} = \gamma \in (0, \infty)$ , and connect them to large-time and tail moderate deviations for (1). In Section 4 we present our main new theoretical result, on moderate deviations for integrated functionals. Section 5 contains the financial implications of our results on option price asymptotics, Sections 6-7 gather the proof of the moderate deviations for integrated functionals, while Appendices A and B state some results of independent interest, used throughout the paper.

## 2. PATHWISE MODERATE DEVIATIONS FOR SMALL-NOISE DIFFUSIONS

We first consider pathwise moderate deviations for small-noise Itô diffusions, following [1, 13, 26]. By suitable rescaling, this allows us to unify and generalise several recent results in the mathematical finance literature on small-time moderate deviations, in particular those of [15]. Consider the  $\mathbb{R}^d$ -valued SDE

$$(4) \quad d\mathbb{X}_t^\varepsilon = b_\varepsilon(t, \mathbb{X}_t^\varepsilon)dt + \sqrt{\varepsilon}a_\varepsilon(t, \mathbb{X}_t^\varepsilon)dW_t, \quad \mathbb{X}_0^\varepsilon = x_0 \in \mathbb{R}^d,$$

where for every  $\varepsilon \in (0, 1)$ ,  $b_\varepsilon : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $a_\varepsilon : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathcal{M}^{d \times m}$  (the space of real-valued  $d \times m$  matrices) and  $W$  is a standard multidimensional Brownian motion taking values in  $\mathbb{R}^m$  defined on an appropriately filtered probability space. We shall also fix some arbitrary time horizon  $T > 0$ . We will be working with diffusions of the form (4) that have unique strong solutions, and that satisfy the following assumption:

### Assumption 2.1.

- (i) The coefficients  $b_\varepsilon$ ,  $a_\varepsilon$  as well as  $Db_\varepsilon$  converge as  $\varepsilon$  tends to zero to some functions  $b$ ,  $a$  and  $Db$  respectively, uniformly on bounded subsets of  $[0, T] \times \mathcal{C}([0, T]; \mathbb{R}^d)$ .
- (ii) The coefficients  $b_\varepsilon$ ,  $b \in \mathcal{C}^{0,1}$ ,  $a_\varepsilon$ ,  $a$  are locally Lipschitz in  $x$  uniformly in  $\varepsilon \in (0, 1)$  and  $t \in [0, T]$ .

(iii) There exists  $C > 0$  such that for all  $t \in [0, T]$  and  $\phi \in \mathcal{C}([0, T], \mathbb{R}^d)$ ,

$$\max \left\{ \text{Tr} \left[ a_\varepsilon a_\varepsilon^\top(t, \phi) \right], \langle \phi, b_\varepsilon(t, \phi) \rangle \right\}, \max \left\{ \text{Tr} \left[ aa^\top(t, \phi) \right], \langle \phi, b(t, \phi) \rangle \right\} \leq C \left( 1 + \sup_{s \in [0, t]} |\phi_s|^2 \right),$$

Under Assumption 2.1, the SDE (4) has a unique non-explosive solution and  $\sup_{t \in [0, T]} |\mathbb{X}_t^\varepsilon - \bar{\mathbb{X}}_t|$  converges to zero in probability as  $\varepsilon$  tends to zero, where  $\bar{\mathbb{X}}$  satisfies  $\bar{\mathbb{X}}_t = x_0 + \int_0^t b(s, \bar{\mathbb{X}}_s) ds$ . Let now  $h(\cdot)$  be a continuous function on  $(0, 1]$ , diverging to infinity at the origin, such that  $\lim_{\varepsilon \downarrow 0} \sqrt{\varepsilon} h(\varepsilon) = 0$ . Define the moderate deviations process  $\eta^\varepsilon$  pathwise as

$$(5) \quad \eta_t^\varepsilon := \frac{\mathbb{X}_t^\varepsilon - \bar{\mathbb{X}}_t}{\sqrt{\varepsilon h(\varepsilon)}}, \quad \text{for all } t \geq 0.$$

We are interested in the moderate deviations principle (MDP) for  $\{\mathbb{X}^\varepsilon, \varepsilon > 0\}$  in  $\mathcal{C}([0, T]; \mathbb{R}^d)$ , which amounts to establishing the large deviations principle (LDP) for  $\{\eta^\varepsilon, \varepsilon > 0\}$  in  $\mathcal{C}([0, T]; \mathbb{R}^d)$ . The following theorem is in the spirit of [1, 5, 13, 26], and follows by arguments similar to those in [5, 26].

**Theorem 2.2.** *Under Assumption 2.1, the family  $\{\eta^\varepsilon, \varepsilon > 0\}$  satisfies the LDP (or  $\{\mathbb{X}^\varepsilon, \varepsilon > 0\}$  satisfies the MDP) in  $\mathcal{C}([0, T]; \mathbb{R}^d)$  with speed  $h^2(\varepsilon)$  and rate function*

$$S(\xi) = \inf \left\{ \frac{1}{2} \int_0^T |u_s|^2 ds : u \in L^2([0, T]; \mathbb{R}^d), \xi = x_0 + \int_0^\cdot [Db(s, \bar{\mathbb{X}}_s)\xi_s + a(s, \bar{\mathbb{X}}_s)u_s] ds \right\}.$$

Notice that if  $aa^\top(t, x)$  is uniformly non degenerate on the path of  $\{\bar{\mathbb{X}}_s, s \in [0, T]\}$ , then we can write

$$S(\xi) = \begin{cases} \frac{1}{2} \int_0^T \left( \dot{\xi}_s - Db(s, \bar{\mathbb{X}}_s)\xi_s \right)^\top (aa^\top(s, \bar{\mathbb{X}}_s))^{-1} \left( \dot{\xi}_s - Db(s, \bar{\mathbb{X}}_s)\xi_s \right) ds, & \text{if } \xi \in \mathcal{AC}([0, T], \mathbb{R}^d), \xi_0 = x_0, \\ +\infty, & \text{otherwise.} \end{cases}$$

where  $\mathcal{AC}([0, T], \mathbb{R}^d)$  denotes the class of absolutely continuous functions from  $[0, T]$  to  $\mathbb{R}^d$ .

**2.1. Small-time moderate deviations.** We now show how to use Theorem 2.2 in order to prove small-time moderate deviations for a class of two-dimensional systems  $\mathbb{X}$  satisfying the differential equations

$$(6) \quad \begin{aligned} dX_t &= -\frac{1}{2}\sigma^2(X_t, Y_t)dt + \sigma(X_t, Y_t)dW_t, \\ dY_t &= f(t, X_t, Y_t)dt + g(X_t, Y_t)dZ_t, \\ d\langle W, Z \rangle_t &= \rho dt, \end{aligned}$$

with starting point  $\mathbb{X}_0 = (x_0, y_0)$ . Since we are interested in small-time asymptotics, we rescale (6) according to (2) with  $\delta = 1$ , namely we consider  $\mathbb{X}_t^\varepsilon := (X_t^\varepsilon, Y_t^\varepsilon) := (X_{\varepsilon t}, Y_{\varepsilon t})$ . This yields the system

$$\begin{aligned} dX_t^\varepsilon &= -\frac{\varepsilon}{2}\sigma^2(X_t^\varepsilon, Y_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon, Y_t^\varepsilon)dW_t, \\ dY_t^\varepsilon &= \varepsilon f(X_t^\varepsilon, Y_t^\varepsilon)dt + \sqrt{\varepsilon}g(X_t^\varepsilon, Y_t^\varepsilon)dZ_t, \end{aligned}$$

with starting point  $\mathbb{X}_0^\varepsilon = (x_0, y_0)$ . The limiting process  $\bar{\mathbb{X}} := \lim_{\varepsilon \downarrow 0} \mathbb{X}^\varepsilon$  is constant, equal to  $(x_0, y_0)$  almost surely. Define now the process  $\eta^\varepsilon$  as in (5), where  $\sqrt{\varepsilon}h(\varepsilon)$  tends to zero and  $h(\varepsilon)$  tends to infinity as  $\varepsilon$  tends to zero. In the notations of (4), we have

$$(7) \quad b_\varepsilon(t, x, y) = \varepsilon \begin{pmatrix} -\frac{1}{2}\sigma^2(x, y) \\ f(t, x, y) \end{pmatrix} \quad \text{and} \quad a_\varepsilon(t, x, y) = \begin{pmatrix} \sigma(x, y) & 0 \\ \rho g(x, y) & \bar{\rho}g(x, y) \end{pmatrix},$$

where for convenience, we denote  $\bar{\rho} := \sqrt{1 - \rho^2}$ . Then, using Theorem 2.2, we have the following result:

**Proposition 2.3.** *If  $b_\varepsilon$  and  $a_\varepsilon$  in (7) satisfy Assumption 2.1,  $\sigma(\cdot), g(\cdot)$  are non-zero on  $(x_0, y_0)$ , then  $(\mathbb{X}^\varepsilon)_{\varepsilon > 0}$  satisfies the pathwise MDP with speed  $h^2(\varepsilon)$  and rate function*

$$S(\phi, \psi) = \begin{cases} \frac{1}{2(1 - \rho^2)} \int_0^T \left( \left| \frac{\dot{\phi}_s}{\sigma(x_0, y_0)} \right|^2 - \frac{2\rho\dot{\phi}_s\dot{\psi}_s}{\sigma(x_0, y_0)g(x_0, y_0)} + \left| \frac{\dot{\psi}_s}{g(x_0, y_0)} \right|^2 \right) ds, & \text{if } \phi, \psi \in \mathcal{AC}([0, T], \mathbb{R}), \\ +\infty, & \text{and } (\phi_0, \psi_0) = (x_0, y_0), \\ & \text{otherwise.} \end{cases}$$

*Proof.* The matrix  $aa^\top(x_0, y_0)$  is invertible, since  $(x_0, y_0) \neq (0, 0)$ , so that Theorem 2.2 applies.  $\square$

We are primarily interested in the MDP for the  $X^\varepsilon$  component, which we obtain via contraction principle:

**Corollary 2.4.** *Under the assumptions of Proposition 2.3,  $\{X^\varepsilon, \varepsilon > 0\}$  satisfies the pathwise MDP with speed  $h^2(\varepsilon)$  and rate function*

$$I(\phi) = \begin{cases} \frac{1}{2\sigma(x_0, y_0)^2} \int_0^T |\dot{\phi}_s|^2 ds, & \text{if } \phi \in \mathcal{AC}([0, T], \mathbb{R}) \text{ and } \phi_0 = x_0, \\ +\infty, & \text{otherwise.} \end{cases}$$

*Proof.* By Proposition 2.3 and the contraction principle [6, Section 4.2.1], the MDP rate function for the first component of  $(X_t^\varepsilon, Y_t^\varepsilon)$  is

$$I(\phi) = \inf\{S(\phi, \psi) : \psi \in \mathcal{C}([0, T], \mathbb{R}), \psi_0 = y_0\},$$

which can be solved as a variational problem. For a fixed absolutely continuous trajectory  $\phi$ , the Euler-Lagrange equation for the stationary path  $\psi$  is

$$\ddot{\psi}_t = \rho \frac{g(x_0, y_0)}{\sigma(x_0, y_0)} \ddot{\phi}_t,$$

or  $\dot{\psi}_t = \rho \frac{g(x_0, y_0)}{\sigma(x_0, y_0)} \dot{\phi}_t + C$  for some constant  $C$ . Plugging it in the expression for  $S(\cdot)$  in Proposition 2.3, and minimising over  $C$  we obtain that the minimum is at  $C = 0$ , which concludes the proof.  $\square$

**Remark 2.5.** Setting  $T = 1$ , and understanding  $\varepsilon$  as time, Corollary 2.4 immediately provides us with small-time asymptotics of probabilities. For  $x_1 > x_0$ , the minimisation problem trivially yields  $\phi_t = x_0 + (x_1 - x_0)t$  as optimal path on  $[0, 1]$ , so that

$$\lim_{\varepsilon \downarrow 0} \frac{1}{h(\varepsilon)^2} \log \mathbb{P} \left( \frac{X_\varepsilon}{\sqrt{\varepsilon h(\varepsilon)}} \geq x_1 \right) = -\frac{(x_1 - x_0)^2}{2\sigma(x_0, y_0)^2}.$$

## 2.2. Examples.

2.2.1. *Local-stochastic volatility model.* Consider a local stochastic volatility model of the form (6), with

$$\sigma(t, X_t, Y_t) = \sigma_L(t, X_t)\xi(Y_t),$$

and assume that the local volatility component  $\sigma_L(\varepsilon t, \cdot)$  converges uniformly to  $\bar{\sigma}_L(\cdot)$  as  $\varepsilon$  tends to zero. This is consistent with diffusion-type models, but not for jump models [16]. Applying the time rescaling transformation  $t \mapsto \varepsilon t$  yields

$$\begin{aligned} dX_t^\varepsilon &= -\frac{\varepsilon}{2} \sigma_L^2(\varepsilon t, X_t^\varepsilon) \xi(Y_t^\varepsilon)^2 dt + \sqrt{\varepsilon} \sigma_L(\varepsilon t, X_t^\varepsilon) \xi(Y_t^\varepsilon) dW_t, \\ dY_t^\varepsilon &= \varepsilon f(X_t^\varepsilon, Y_t^\varepsilon) dt + \sqrt{\varepsilon} g(X_t^\varepsilon, Y_t^\varepsilon) dZ_t, \end{aligned}$$

with starting point  $(X_0^\varepsilon, Y_0^\varepsilon) = (x_0, y_0)$ . Clearly,  $\mathbb{X}^\varepsilon = (X^\varepsilon, Y^\varepsilon)$  converges to the constant process equal to  $x_0 = (x_0, y_0)$  almost surely as  $\varepsilon$  tends to zero. Choosing  $h(\varepsilon) \equiv \varepsilon^{-\beta}$ , with  $\beta \in (0, 1/2)$ , we then obtain that under Assumption 2.1 and the assumptions of Proposition 2.3, in particular assuming that  $\bar{\sigma}_L$  is not zero outside the origin,  $\eta^\varepsilon := \varepsilon^{\beta-1/2}(\mathbb{X}^\varepsilon - x_0)$  satisfies a LDP as  $\varepsilon$  tends to zero by Theorem 2.2, and hence  $X^\varepsilon$  satisfies a MDP from Corollary 2.4 with rate function

$$I(\phi) = \begin{cases} \frac{1}{2\bar{\sigma}_L(x_0)^2 \xi(y_0)^2} \int_0^T \dot{\phi}_s^2 ds, & \text{if } \phi \in \mathcal{AC}([0, T], \mathbb{R}) \text{ and } \phi_0 = x_0, \\ +\infty, & \text{otherwise.} \end{cases}$$

2.2.2. *Stein-Stein.* Consider the Stein-Stein [31] stochastic volatility model:

$$\begin{aligned} (8) \quad dX_t &= -\frac{1}{2} Y_t^2 dt + Y_t dW_t, & X_0 &= x_0, \\ dY_t &= (a + bY_t) dt + c dZ_t, & Y_0 &= y_0, \\ d\langle W, Z \rangle_t &= \rho dt. \end{aligned}$$

With the time-scaling  $t \mapsto \varepsilon t$ , the system (8) reads

$$\begin{aligned} dX_t^\varepsilon &= -\frac{\varepsilon}{2}(Y_t^\varepsilon)^2 dt + \sqrt{\varepsilon} Y_t^\varepsilon dW_t, & X_0^\varepsilon &= x_0, \\ dY_t^\varepsilon &= (a + bY_t^\varepsilon)\varepsilon dt + \sqrt{\varepsilon} c dZ_t, & Y_0^\varepsilon &= Y_0, \\ d\langle W, Z \rangle_t &= \rho dt. \end{aligned}$$

Clearly,  $\mathbb{X}^\varepsilon = (X^\varepsilon, Y^\varepsilon)$  converges to the constant process equal to  $\mathbf{x}_0 = (x_0, Y_0)$  almost surely as  $\varepsilon$  tends to zero. Choosing  $h(\varepsilon) \equiv \varepsilon^{-\beta}$ , with  $\beta \in (0, 1/2)$ , we then obtain that  $\eta^\varepsilon := \varepsilon^{\beta-1/2}(\mathbb{X}^\varepsilon - \mathbf{x}_0)$  satisfies a LDP as  $\varepsilon$  tends to zero by Theorem 2.2, and hence  $X^\varepsilon$  satisfies a MDP from Corollary 2.4 with rate function

$$I(\phi) = \begin{cases} \frac{1}{2y_0^2} \int_0^T \dot{\phi}_s^2 ds, & \text{if } \phi \in \mathcal{AC}([0, T], \mathbb{R}) \text{ and } \phi_0 = x_0, \\ +\infty, & \text{otherwise.} \end{cases}$$

2.2.3. *Heston.* Consider the Heston stochastic volatility model:

$$(9) \quad \begin{aligned} dX_t &= -\frac{1}{2}Y_t dt + \sqrt{Y_t} dW_t, & X_0 &= x_0, \\ dY_t &= \kappa(\theta - Y_t)dt + \xi\sqrt{Y_t} dZ_t, & Y_0 &= y_0, \\ d\langle W, Z \rangle_t &= \rho dt, \end{aligned}$$

Applying the time rescaling  $t \mapsto \varepsilon t$ , the system (9) turns into

$$\begin{aligned} dX_t^\varepsilon &= -\frac{\varepsilon}{2}Y_t^\varepsilon dt + \sqrt{\varepsilon}\sqrt{Y_t^\varepsilon} dW_t, & X_0^\varepsilon &= x_0, \\ dY_t^\varepsilon &= \kappa(\theta - Y_t^\varepsilon)\varepsilon dt + \sqrt{\varepsilon}\xi\sqrt{Y_t^\varepsilon} dZ_t, & Y_0^\varepsilon &= y_0, \\ d\langle W, Z \rangle_t &= \rho dt. \end{aligned}$$

Clearly,  $\mathbb{X}^\varepsilon = (X^\varepsilon, Y^\varepsilon)$  converges to the constant process  $\mathbf{x}_0 = (x_0, y_0)$  almost surely as  $\varepsilon$  tends to zero. Choosing again  $h(\varepsilon) \equiv \varepsilon^{-\beta}$ , with  $\beta \in (0, 1/2)$ , we then obtain that  $\eta^\varepsilon := \varepsilon^{\beta-1/2}(\mathbb{X}^\varepsilon - \mathbf{x}_0)$  satisfies a LDP as  $\varepsilon$  tends to zero, and that  $X^\varepsilon$  satisfies a MDP from Corollary 2.4 with rate function

$$I(\phi) = \begin{cases} \frac{1}{2y_0} \int_0^T \dot{\phi}_s^2 ds, & \text{if } \phi \in \mathcal{AC}([0, T], \mathbb{R}) \text{ and } \phi_0 = x_0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Together with Remark 2.5, this corresponds to the Moderate Deviations regime for the Heston model in [15, Theorem 6.2], with  $\hat{\beta} = 1/2 - \beta$ .

### 3. LARGE-TIME MODERATE DEVIATIONS

The former section considered small-noise perturbations of stochastic Itô diffusions, which lent himself perfectly to small-time analysis of the solution of the system by scaling. We investigate here proper multiscaled diffusions, in the form of (3), and show how their behaviour yields large-time asymptotics of the solution. Consider a general perturbed two-dimensional diffusion  $\mathbb{X} = (X, Y)$  of the form

$$(10) \quad \begin{aligned} dX_t &= -\frac{1}{2}\sigma^2(Y_t)dt + \sigma(Y_t)dW_t, \\ dY_t &= f(Y_t)dt + g(Y_t)dZ_t, \\ d\langle W, Z \rangle_t &= \rho dt, \end{aligned}$$

with starting point  $\mathbb{X}_0 = (x_0, y_0)$ . For  $0 < \varepsilon, \delta < 1$ , the general rescaling (2) yields

$$(11) \quad \begin{aligned} dX_t^\varepsilon &= -\frac{1}{2}\frac{\varepsilon}{\delta}\sigma^2(Y_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(Y_t^\varepsilon)dW_t, \\ dY_t^\varepsilon &= \frac{\varepsilon}{\delta^2}f(Y_t^\varepsilon)dt + \frac{\sqrt{\varepsilon}}{\delta}g(Y_t^\varepsilon)dZ_t, \\ d\langle W, Z \rangle_t &= \rho dt, \end{aligned}$$

If  $\delta = \varepsilon$ , then  $(X_t^\varepsilon, Y_t^\varepsilon) = (\varepsilon X_{t/\varepsilon}, Y_{t/\varepsilon})$ , but the more general space-time transformation (2) allows us to consider also the case  $\frac{\varepsilon}{\delta} \rightarrow \gamma \in [0, \infty)$ . For  $0 < \varepsilon, \delta \ll 1$  this is in the setup of [18] and more relevant, if we allow the possibility of  $\frac{\varepsilon}{\delta} \rightarrow \gamma \in (0, \infty)$  and correlation  $\rho \neq 0$ , the setup of [26]. To this end let us make the following standing assumption:

**Assumption 3.1.**

(i) The function  $\sigma(\cdot) \in \mathcal{C}^\alpha(\mathbb{R})$  for some  $\alpha > 0$  and there exist constants  $0 < K < \infty$ ,  $0 \leq q_\sigma < 1$  such that

$$|\sigma(y)| \leq K(1 + |y|^{q_\sigma}).$$

(ii) The function  $f(\cdot)$  is locally bounded and we can write  $f(y) = -\kappa y + \tau(y)$  where  $\tau(\cdot)$  is a globally Lipschitz function with Lipschitz constant  $L_\tau < \kappa$ . In addition,  $\lim_{|y| \uparrow \infty} \frac{f(y)y}{|y|^2} = -\kappa$ .

(iii) The function  $g$  is either uniformly continuous and bounded from above and away from zero or alternatively it takes the form  $g(y) = \xi y^{q_g}$  for  $q_g \in [1/2, 1)$  with  $\xi$  a non-zero constant. In the latter case, we also assume that  $f(y) = \kappa(\theta - y)$ .

**Assumption 3.2.** The SDE (10) has a unique strong solution.

Let  $\mathcal{L}_Y$  denote the infinitesimal generator of the  $Y$  process before time/space rescalings, i.e.

$$(12) \quad \mathcal{L}_Y h = fh' + \frac{1}{2}g^2h'',$$

and let  $\mu(d\cdot)$  be the invariant measure corresponding to  $\mathcal{L}_Y$ , which under Assumption 3.1 exists and is unique. With

$$\lim_{\varepsilon \downarrow 0} \frac{\varepsilon}{\delta} =: \gamma \in (0, \infty),$$

let us set

$$\lambda(y) := -\frac{1}{2}\gamma\sigma^2(y) \quad \text{and} \quad \bar{\lambda} := \int_{\mathbb{R}} \lambda(y)\mu(dy).$$

Then, it is a classical result [27] that  $X^\varepsilon$  converges to  $\bar{X}$  in probability as  $\varepsilon$  tends to zero, where

$$\bar{X}_t := x_0 + \bar{\lambda}t.$$

Lastly, we introduce the function  $\Phi$ , solving the Poisson equation

$$(13) \quad \mathcal{L}_Y \Phi(y) = -\frac{1}{\gamma}(\lambda(y) - \bar{\lambda}) = \frac{1}{2}(\sigma^2(y) - \bar{\sigma}^2), \quad \text{with} \quad \int_{\mathcal{Y}} \Phi(y)\mu(dy) = 0,$$

where  $\bar{\sigma}^2 := \int_{\mathbb{R}} \sigma^2(y)\mu(dy)$ . In the case of bounded and non-degenerate  $g$ , Assumption 3.1 implies that Theorems 1 and 2 in [27] hold, providing existence and appropriate smoothness of  $\Phi$  and also polynomial growth (in  $|y|$ ). In the case where  $g$  is only Hölder continuous with exponent  $q_g \in [1/2, 1)$  and can become zero only at zero, the same conclusions hold by Lemma A.2. In order to state Theorem 3.3, we need to know the relative rates at which  $\delta$ ,  $\varepsilon$ , and  $1/h(\varepsilon)$  vanish. In particular,

$$(14) \quad \zeta := \lim_{\varepsilon \downarrow 0} \frac{\varepsilon/\delta - \gamma}{\sqrt{\varepsilon}h(\varepsilon)}$$

specifies the relative rate at which  $\varepsilon/\delta$  converges and  $h(\varepsilon)$  goes to infinity. In order for a MDP to hold, we require that  $\zeta$  be finite. In the special case  $\frac{\varepsilon}{\delta} = \gamma$ , then  $\zeta = 0$ . The following holds:

**Theorem 3.3** (Theorem 2.1 of [26]). *Under Assumptions 3.1 and 3.2, the sequence  $\{X^\varepsilon, \varepsilon > 0\}$  from (11) satisfies the MDP with speed  $h^2(\varepsilon)$  and rate function*

$$I(\phi) = \inf \left\{ \frac{1}{2} \int_0^T |u_s|^2 ds : u \in L^2([0, T]; \mathbb{R}), \phi = \int_0^\cdot (\alpha + q^{1/2}u_s) ds \right\},$$

where

$$\alpha = -\frac{\zeta}{2} \int_{\mathbb{R}} \sigma^2(y)\mu(dy) \quad \text{and} \quad q = \int_{\mathbb{R}} \left[ \sigma^2(y) + |\Phi'(y)g(y)|^2 + 2\rho\sigma(y)g(y)\Phi'(y) \right] \mu(dy),$$

where the finite constant  $\zeta$  is defined in (14), and  $\Phi$  is the solution to the Poisson equation (13).

**Example 3.4** (Heston model). For the Heston model (9), the rescalings (2) yield the system

$$(15) \quad \begin{aligned} dX_t^\varepsilon &= -\frac{\varepsilon}{2\delta} Y_t^\varepsilon dt + \sqrt{\varepsilon} \sqrt{Y_t^\varepsilon} dW_t, & X_0^\varepsilon &= \delta x_0, \\ dY_t^\varepsilon &= \kappa(\theta - Y_t^\varepsilon) \frac{\varepsilon}{\delta^2} dt + \frac{\sqrt{\varepsilon}}{\delta} \xi \sqrt{Y_t^\varepsilon} dZ_t, & Y_0^\varepsilon &= y_0, \\ d\langle W, Z \rangle_t &= \rho dt. \end{aligned}$$

In this case, the invariant measure  $\mu$  has the Gamma density [17, Section 33.4]

$$(16) \quad m(y) = \frac{(2\kappa/\xi^2)^{2\kappa\theta/\xi^2}}{\Gamma(2\kappa\theta/\xi^2)} y^{2\kappa\theta/\xi^2-1} e^{-2\kappa y/\xi^2}, \quad \text{for } y > 0.$$

Furthermore, it is easy to check that Assumption 3.1 is satisfied, and so is Assumption 3.2 by [23, Chapter 5, Theorem 2.9], so that Theorem 3.3 holds with

$$\alpha = -\frac{\zeta\theta}{2} \quad \text{and} \quad q = \int_{\mathbb{R}} y \left( 1 + |\xi\Phi'(y)|^2 + 2\rho\xi\Phi'(y) \right) \mu(dy).$$

The Poisson equation (13) can be solved explicitly as  $\Phi'(y) = -\frac{1}{2\kappa}$ , which yields

$$(17) \quad q = \theta \left( 1 + \frac{\xi^2}{4\kappa^2} - \frac{\rho\xi}{\kappa} \right) > 0.$$

**Remark 3.5** (Connections to large deviations). For the Heston model in Example 3.4, consider the regime  $\varepsilon = \delta$ , so that  $\gamma = 1$  and  $\zeta = \alpha = 0$ . Taking  $h(\varepsilon) = \varepsilon^{-\beta}$  with  $\beta \in (0, 1/2)$  implies an LDP for the sequence of processes  $(\eta^\varepsilon = \varepsilon^{\beta-1/2} X^\varepsilon)_{\varepsilon>0}$ , or, in terms of the original process, an LDP for  $\varepsilon^{\beta+1/2} X_{\cdot/\varepsilon}$ . Setting  $t = 1$  and mapping  $\varepsilon \mapsto t^{-1}$  yields an LDP for  $X_t/t^{\beta+1/2}$  as  $t$  tends to infinity. The MDP from Theorem 3.3 therefore implies that, for any  $x > 0$ :

$$\lim_{t \uparrow \infty} t^{-2\beta} \log \mathbb{P} \left( \frac{X_t}{t^{\beta+1/2}} \geq x \right) = -\inf \{ I(\phi) : \phi_0 = x_0, \phi_1 \geq x \}.$$

The Euler-Lagrange equation for this minimisation problem is simply  $-q\ddot{\phi}_t = 0$ , which yields, with the boundary conditions, the optimal path  $\phi_t = (x - x_0)t + x_0$ . Therefore, we obtain

$$\lim_{t \uparrow \infty} t^{-2\beta} \log \mathbb{P} \left( \frac{X_t}{t^{\beta+1/2}} \geq x \right) = -\frac{(x - x_0)^2}{2q},$$

with  $q$  given in (17). Now, the large-time large deviations regime was proved in [12, 22], with

$$\lim_{t \uparrow \infty} t^{-1} \log \mathbb{P} (t^{-1} X_t \geq x) = -\Lambda^*(x),$$

and rate function  $\Lambda^*$  available in closed form as

$$\Lambda^*(x) := u^*(x)x - \Lambda(u^*(x)), \quad \text{for all } x \in \mathbb{R},$$

with

$$\begin{aligned} \Lambda(u) &:= \frac{\kappa\theta}{\xi^2} (\kappa - \rho\xi u - d(u)) \quad \text{and} \quad d(u) := \sqrt{(\kappa - \rho\xi u)^2 + \xi^2 u (1 - u^2)}, \quad \text{for all } u \in (u_-, u_+), \\ u^*(x) &:= \frac{\xi - 2\kappa\rho + (\kappa\theta\rho + x\xi)\widehat{\zeta} (x^2\xi^2 + 2x\kappa\theta\rho\xi + \kappa^2\theta^2)^{-1/2}}{2\xi(1 - \rho^2)}, \quad u_{\pm} := \frac{\xi - 2\kappa\rho \pm \widehat{\zeta}}{2\xi(1 - \rho^2)}, \end{aligned}$$

and  $\widehat{\zeta} := \sqrt{4\kappa^2 + \xi^2 - 4\kappa\rho\xi}$ . Here again, the moderate rate function turns out to be the second-order Taylor expansion of the large deviations rate function at its minimum. More precisely, it is easy to show that  $\Lambda^*$  attains its minimum at  $-\theta/2$ , which corresponds to the limiting behaviour (at  $t = 1$ ) of  $X_1^\varepsilon$  as  $\varepsilon$  tends to zero. Indeed, due to ergodicity, we have that  $d\bar{X}_t = -\frac{1}{2}\theta dt$ . We can also show that  $\partial_{xx}^2 \Lambda^*(-\theta/2) = q^{-1}$ , where the constant  $q$  is given by the moderate deviations regime (17). Hence, the moderate deviations rate function characterizes the local curvature of the large deviations rate function around its minimum.

#### 4. LIMITING BEHAVIOUR FOR INTEGRATED FUNCTIONALS

We now consider an extension of the previous results to functionals of diffusion processes, and prove moderate deviations thereof. This problem has already been approached in [20], but, using weak convergence techniques, we are able to relax some of the their assumptions, making the results amenable to applications in mathematical finance, see also Remark 4.6. We consider the model

$$(18) \quad \begin{aligned} dX_t^\varepsilon &= -\frac{1}{2} \frac{\varepsilon}{\delta} \sigma^2(X_t^\varepsilon, Y_t^\varepsilon) dt + \sqrt{\varepsilon} \sigma(X_t^\varepsilon, Y_t^\varepsilon) dW_t, \\ dY_t^\varepsilon &= \frac{\varepsilon}{\delta^2} f(X_t^\varepsilon, Y_t^\varepsilon) dt + \frac{\sqrt{\varepsilon}}{\delta} g(X_t^\varepsilon, Y_t^\varepsilon) dZ_t, \\ d\langle W, Z \rangle_t &= \rho dt, \end{aligned}$$

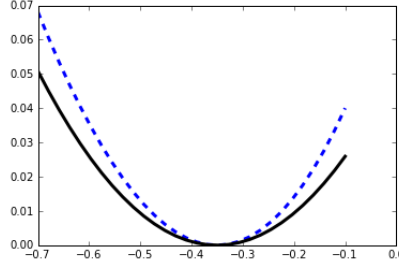


FIGURE 1. Comparison of the MDP rate function (black, solid) and the LDP rate function (blue, dash) in the Heston model. The MDP rate function has been shifted so that the minima match. The parameters are  $\kappa, \theta, \sigma, \rho = 2.7609, 0.7, -0.6, 0.8$ .

which corresponds to the original system (1) using the general rescaling (2) (as in the large-time framework above), whenever the coefficients do not depend on  $X^\varepsilon$ . For given sets  $A, B$ , for  $i, j \in \mathbb{N}$  and  $\alpha > 0$  we denote by  $\mathcal{C}_b^{i,j+\alpha}(A \times B)$ , the space of functions with  $i$  bounded derivatives in  $x$  and  $j$  derivatives in  $y$ , with all partial derivatives being  $\alpha$ -Hölder continuous with respect to  $y$ , uniformly in  $x$ . For any  $x \in \mathbb{R}$ , let  $\mathcal{L}_x^Y$  denote the infinitesimal generator of the  $Y$  process when  $\varepsilon = \delta = 1$  (i.e. before the space/time rescalings):

$$(19) \quad \mathcal{L}_x^Y h(y) = f(x, y)h'(y) + \frac{1}{2}g^2(x, y)h''(y),$$

and let  $\mu_x(dy)$  be the invariant measure corresponding to the infinitesimal generator  $\mathcal{L}_x^Y$ , which is guaranteed to exist by Assumption 4.1 below. With

$$(20) \quad \lim_{\varepsilon \downarrow 0} \frac{\varepsilon}{\delta} = \gamma \in (0, \infty),$$

let us set

$$\lambda(x, y) := -\frac{1}{2}\gamma\sigma^2(x, y) \quad \text{and} \quad \bar{\lambda}(x) := \int_{\mathbb{R}} \lambda(x, y)\mu_x(dy).$$

It is a classical result [27] that  $X^\varepsilon$  converges in probability to  $\bar{X}$  as  $\varepsilon$  tends to zero, where

$$(21) \quad \bar{X}_t := x_0 + \int_0^t \bar{\lambda}(\bar{X}_s)ds.$$

For an appropriate function  $H(x, y)$ , let  $\bar{H}(x) := \int_{\mathbb{R}} H(x, y)\mu(dy)$ ; we are interested in the LDP for

$$(22) \quad R^\varepsilon := \frac{1}{\sqrt{\varepsilon}h(\varepsilon)} \int_0^\cdot (H(X_s^\varepsilon, Y_s^\varepsilon) - \bar{H}(X_s^\varepsilon)) ds.$$

The following assumption is a counterpart to Assumption 3.1 in the more general configuration with full dependence on  $x$  and  $y$  in the coefficients.

**Assumption 4.1.**

- (i) Either  $f, g \in \mathcal{C}^{1,\alpha}(\mathbb{R})$  and  $H \in \mathcal{C}^{2,1+\alpha}(\mathbb{R})$  for some  $\alpha > 0$ , or  $f, g \in \mathcal{C}^{1,2+\alpha}(\mathbb{R})$  and  $H \in \mathcal{C}^{2,\alpha}(\mathbb{R})$  for some  $\alpha > 0$ .
- (ii) There exist constants  $0 < K < \infty$  and  $0 \leq q_\sigma < 1$  such that

$$|\sigma(x, y)| \leq K(1 + |y|^{q_\sigma}).$$

In addition,  $\sigma$  is Lipschitz continuous in  $x$  locally uniformly with respect to  $y$ .

- (iii) We can write  $f(x, y) = -\kappa y + \tau(x, y)$  with  $\kappa > 0$ ,  $\tau(x, y)$  is a globally Lipschitz function in  $y$ , uniformly with respect to  $x$ , with Lipschitz constant  $L_\tau < \kappa$ , and  $\sup_{x \in \mathbb{R}} f(x, y)y \leq -\kappa|y|^2$  for  $|y|$  large enough.
- (iv) The function  $g$  is either uniformly continuous and bounded from above and away from zero or alternatively it takes the form  $g(y) = \xi y^{q_g}$  for  $q_g \in [1/2, 1)$  and  $\xi$  a non-zero constant. In the latter case we also assume that  $f(x, y) = \kappa(\theta - y)$ .
- (v) There exist constants  $0 < K < \infty$  and  $q_H \geq 0$  such that

$$|H(x, y)| + |\partial_x H(x, y)| + |\partial_x^2 H(x, y)| \leq K(1 + |y|^{q_H}).$$

**Assumption 4.2.** The SDE (18) has a unique strong solution.



Let us further consider  $u(x, \cdot)$  to be the solution to the Poisson equation

$$(23) \quad \mathcal{L}_x^Y u(x, y) = H(x, y) - \bar{H}(x).$$

By Theorem A.1 and Lemma A.2,  $u$  is a well-defined classical solution that can grow at most polynomially in  $|y|$ . Next, due to compactness issues, that we will see in the proof of tightness in Section 7.1, we need to restrict the values of the constants  $q_\sigma, q_g, q_H$  that dictate the growth in  $|y|$  of the coefficients  $\sigma, g$  and  $H$  respectively. We collect these constraints in the following assumption:

**Assumption 4.3.**

- (i) If the operator  $\mathcal{L}_x^Y$  given by (19) depends generally on  $(x, y)$ , we assume that

$$\max\{q_\sigma + q_H, q_g + q_H\} < 1.$$

- (ii) If the operator in (19) takes the special form

$$(24) \quad \mathcal{L}^Y h(y) = \kappa(\theta - y)h'(y) + \frac{1}{2}\xi^2 y^{2q_g} h''(y),$$

with  $q_g \in [1/2, 1)$ , and  $H(x, y) = H(y)$  is only a function of  $y$ , then we assume that

$$q_\sigma < 1 \quad \text{and} \quad q_g + q_H < 2,$$

where the constants  $q_g, q_H$  and the function  $h(\cdot)$  are such that  $\lim_{\varepsilon \downarrow 0} \sqrt{\varepsilon} h(\varepsilon)^{\frac{q_g + q_H - 1}{1 - q_g}} = 0$ .

Notice that for Assumption 4.3(ii), if  $q_H = 1$ , then assuming  $h(\varepsilon) = \varepsilon^{-\beta}$  with  $\beta \in (0, 1/2)$  yields  $\sqrt{\varepsilon} h(\varepsilon)^{q_g/(1-q_g)} = \varepsilon^{1/2 - \beta q_g/(1-q_g)}$ , so that Assumption 4.3 is satisfied if and only if  $q_g < 1/(2\beta + 1) \in (1/2, 1)$ . In order to state the main result, we shall need one additional assumption, merely for technical reasons. Given the solution  $u(\cdot, \cdot)$  to the Poisson equation (23) and the constant  $\gamma$  defined in (20), introduce the following:

$$(25) \quad \bar{Q}(x) := \int_{\mathbb{R}} Q(x, y) \mu(dy) \quad \text{and} \quad Q(x, y) := \frac{1}{\gamma^2} |u_y(x, y) g(y)|^2.$$

**Assumption 4.4.** The Lebesgue measure of the set  $\{s \in [0, T] : \bar{Q}(\bar{X}_s) = 0\}$  is equal to zero.

At this point, we mention that the reason we single out the operator (24) is because in this case we can have more detailed information on the behaviour of the solutions to the corresponding Poisson equation (23). This is discussed in detail in Lemma A.2. To be specific, in the special case of (24) with  $H(x, y) \equiv H(y)$ , the solution to the PDE (23) grows like  $|y|^{q_H}$  whereas the derivative grows like  $|y|^{q_H-1}$  for  $q_H \geq 1$ . On the other hand, in the general case, one can only guarantee that both solution and derivatives to the solution of PDE (23) grow like  $|y|^{q_H}$ . Then, we have the following result, proved in Sections 6-7:

**Theorem 4.5.** *Let Assumptions 4.1, 4.2, 4.3 and 4.4 be satisfied. Then, the sequence  $\{R^\varepsilon, \varepsilon > 0\}$  from (22) satisfies the LDP (equivalently MDP for  $\int_0^\cdot (H(X_s^\varepsilon, Y_s^\varepsilon) - \bar{H}(X_s^\varepsilon)) ds$ ), with speed  $h^2(\varepsilon)$  and rate function*

$$I(\phi) = \inf \left\{ \frac{1}{2} \int_0^T |v_s|^2 ds : v \in L^2([0, T]; \mathbb{R}), \phi = \int_0^\cdot \bar{Q}^{1/2}(\bar{X}_s) v_s ds \right\}.$$

**Remark 4.6.** We point out that a result similar to Theorem 4.5 has also been established in [20] using different methods. However, the results of [20] are not sufficient for our purposes, as they assumed that both functions  $g$  and  $H$  are bounded ( $H$  is even assumed to be bounded by  $K(1 + |y|^{-(2+\eta)})$  for some  $\eta > 0$ ). In our work we allow the coefficients to grow according to Assumptions 4.1 and 4.3. In turn this allows us to consider a considerably wider class of models. We see some examples below.

## 5. FINANCIAL APPLICATIONS: ASYMPTOTIC BEHAVIOUR OF OPTION PRICES

We can use Corollary 2.4 for the short-time asymptotics and Theorem 3.3 for the long-time asymptotics to get statements for Call option prices analogous to Theorem 6.2 and 6.3 of [15]. Theorem 4.5 is used to obtain estimates on asymptotics for realised variance.

**5.1. Tail estimates.** We show here how the general result in Theorem 2.2, as well as its lighter versions in Proposition 2.3 and Corollary 2.4 yield, for some suitable scaling, tail estimates for probabilities.

**Proposition 5.1.** *Consider the model (6) under Assumption 2.1, where the coefficients only depend on  $Y$ . Assume further that  $f$  is affine, and that there exist  $\nu_\sigma \in [0, 1]$  and  $\nu_g \in [0, 1 - \nu_\sigma]$  such that  $g(y/\varepsilon) \sim y\varepsilon^{-\nu_g}$  and  $\sigma(y/\varepsilon) \sim y\varepsilon^{-\nu_\sigma}$  as  $\varepsilon$  tends to zero. Then, with  $\zeta := \frac{1-\nu_g+\nu_\sigma}{2(1-\nu_g)}$ , the sequence  $(\varepsilon^\zeta X)$  satisfies a moderate deviations principle on  $\mathcal{C}([0, T], \mathbb{R})$  with speed  $h(\varepsilon)^2$  and rate function*

$$I(\phi) = \begin{cases} \frac{1}{2\sigma(y_0)^2} \int_0^T |\dot{\phi}_s|^2 ds, & \text{if } \phi \in \mathcal{AC}([0, T], \mathbb{R}) \text{ and } \phi_0 = x_0, \\ +\infty, & \text{otherwise.} \end{cases}$$

*Proof.* Let  $\alpha > 0$  to be chosen. The rescaling  $\mathbb{X}_t^\delta = (X_t^\delta, Y_t^\delta) := (\delta X_t, \delta^\alpha Y_t)$  for all  $t \geq 0$  yields the system

$$\begin{aligned} dX_t^\delta &= -\frac{\delta^{1-2\alpha\nu_\sigma}}{2} (Y_t^\delta)^{2\nu_\sigma} dt + \delta^{1-\alpha\nu_\sigma} (Y_t^\delta)^{\nu_\sigma} dW_t, \\ dY_t^\delta &= (a\delta^\alpha + bY_t^\delta) dt + \delta^{\alpha(1-\nu_g)} (Y_t^\delta)^{\nu_g} dZ_t, \\ d\langle W, Z \rangle_t &= \rho dt, \end{aligned}$$

with starting point  $\mathbb{X}_0^\delta = (X_0^\delta, Y_0^\delta) = (\delta x_0, \delta^\alpha y_0)$ . Equating the  $\delta$  powers in the diffusion coefficients yields  $1 - \alpha\nu_\sigma = \alpha(1 - \nu_g)$ , or  $\alpha = (1 - \nu_g + \nu_\sigma)^{-1}$  so that

$$\begin{aligned} dX_t^\delta &= -\frac{\delta^{1-2\alpha\nu_\sigma}}{2} (Y_t^\delta)^{2\nu_\sigma} dt + \delta^\gamma (Y_t^\delta)^{\nu_\sigma} dW_t, \\ dY_t^\delta &= (a\delta^\alpha + bY_t^\delta) dt + \delta^\gamma (Y_t^\delta)^{\nu_g} dZ_t, \\ d\langle W, Z \rangle_t &= \rho dt, \end{aligned}$$

with  $\gamma := 1 - \alpha\nu_\sigma$ . Setting  $\sqrt{\varepsilon} := \delta^\gamma$  finally gives the system

$$\begin{aligned} dX_t^\varepsilon &= -\frac{\varepsilon^{(1-2\alpha\nu_\sigma)/(2\gamma)}}{2} (Y_t^\varepsilon)^{2\nu_\sigma} dt + \sqrt{\varepsilon} (Y_t^\varepsilon)^{\nu_\sigma} dW_t, \\ dY_t^\varepsilon &= \left( a\varepsilon^{\frac{1}{2(1-\nu_g)}} + bY_t^\varepsilon \right) dt + \sqrt{\varepsilon} (Y_t^\varepsilon)^{\nu_g} dZ_t, \\ d\langle W, Z \rangle_t &= \rho dt, \end{aligned}$$

with starting point  $\mathbb{X}_0^\varepsilon = (\varepsilon^{1/(2\gamma)} x_0, \varepsilon^{\alpha/(2\gamma)} y_0)$ . Theorem 2.2 then applies whenever  $\gamma > 0$  and  $1 - 2\nu_\sigma\alpha \geq 0$ , which is equivalent to  $\nu_g \leq 1 - \nu_\sigma$ , and the proposition follows by setting  $\zeta := 1/(2\gamma)$ .  $\square$

**Example 5.2.**

- In the Heston case (9), we have  $\nu_\sigma = 1/2 = \nu_g$ , so that Proposition 5.1 applies with  $\zeta = 1$ ;
- In the Stein-Stein case (8), we have  $\nu_\sigma = 1$  and  $\nu_g = 0$ , so that Proposition 5.1 applies with  $\zeta = 1$ .

In both cases, for any  $x > x_0$ , we can therefore write

$$\lim_{\varepsilon \downarrow 0} \frac{1}{h^2(\varepsilon)} \log \mathbb{P} \left( \frac{X_t^\varepsilon}{\sqrt{\varepsilon} h(\varepsilon)} \geq x \right) = \lim_{\varepsilon \downarrow 0} \frac{1}{h^2(\varepsilon)} \log \mathbb{P} \left( X_t \geq \frac{xh(\varepsilon)}{\sqrt{\varepsilon}} \right) = -\frac{(x - x_0)^2}{2\sigma(y_0)^2 t}.$$

Again, the moderate deviations regime provides a closed-form representation for the rate function, which is to be contrasted with the otherwise more abstract rate function arising from large deviations, as developed in [7, 8, 21].

**5.2. Small-time behaviour.** The following theorem is a reformulation of [15, Theorem 6.3] and [11, Corollary 2.1 and Theorem 2.2], based on the asymptotic behaviour (assuming  $x_0 = 0$  for simplicity)

$$(26) \quad \mathbb{P}(X_t \geq k_t) \sim \exp \left\{ -\frac{k_t^2}{2\sigma^2(0, y_0)t} \right\},$$

as  $t$  tends to zero, from Proposition 2.3 and Corollary 2.4, with  $k_t = t^\gamma l(t)$ ,  $\gamma \in (0, 1/2)$  and  $l(\cdot)$  a strictly positive function slowly varying at zero.

**Theorem 5.3.** *Assume for all  $p \geq 1$  the  $p$ -th moment of  $e^X$  explode at some positive (possibly infinite) time. Then, under the assumptions of Proposition 2.3, the Call price satisfies*

$$\mathbb{E} (e^{X_t} - e^{k_t})_+ \sim \exp \left\{ -\frac{k_t^2}{2\sigma^2(0, y_0)t} \right\}.$$

In the model given by (6), we can write down a pathwise moderate deviations principle for the first component using Corollary 2.4, which generalises (being pathwise and with weaker assumptions) the results from [15], in particular (26). More precisely, from the proposed scaling  $(X_t^\varepsilon, Y_t^\varepsilon) := (X_{\varepsilon t}, Y_{\varepsilon t})$ , Corollary 2.4 and Remark 2.5 yields, as  $t$  tends to zero, for  $k > x_0 = 0$ ,

$$\lim_{t \downarrow 0} \frac{1}{h(\varepsilon)^2} \log \mathbb{P} \left( \frac{X_\varepsilon}{\sqrt{\varepsilon} h(\varepsilon)} \geq k \right) = -\frac{k^2}{2\sigma^2(0, y_0)}.$$

Identifying  $t$  to  $\varepsilon$  and setting  $k_t := kh(t)\sqrt{t}$  yields, as  $t$  tends to zero,

$$\mathbb{P}(X_t \geq k_t) \sim \exp \left\{ -\frac{k^2 h(t)^2}{2\sigma^2(0, y_0)} \right\} = \exp \left\{ -\frac{k_t^2}{2\sigma^2(0, y_0)t} \right\},$$

which corresponds to (26), with  $h(t) = t^{-\beta} l(t)$  for some slowly varying function  $l$  and  $\beta = 1/2 - \gamma \in (0, 1/2)$ .

**5.3. Large-time behaviour.** Following the methodology developed in [12, 22], we can translate the moderate deviations results into large-time asymptotic behaviours of option prices. In order to state the results, let  $J$  denote the (convex) contraction of the rate function  $I$  given in Theorem 3.3 onto the last point, i.e.  $J(x) := \inf \{ I(\phi) : \phi(0) = x_0, \phi(1) = x \}$ . Introduce further the Share measure  $\mathbb{Q}(A) := \mathbb{E}^\mathbb{P} (e^{X_t} \mathbf{1}_{\{A\}})$  for any  $A \in \mathcal{F}_t$ . Before proving the main result of this section, let us state and prove a useful lemma.

**Lemma 5.4.** *Under  $\mathbb{Q}$ , the process  $X$  defined in (10) satisfies*

$$(27) \quad \begin{aligned} dX_t &= \frac{1}{2} \sigma^2(Y_t) dt + \sigma(Y_t) dW_t^\mathbb{Q}, & X_0 &= x_0, \\ dY_t &= f^\mathbb{Q}(Y_t) dt + g(Y_t) dZ_t^\mathbb{Q}, & Y_0 &= y_0, \\ d\langle W^\mathbb{Q}, Z^\mathbb{Q} \rangle_t &= \rho dt, \end{aligned}$$

where  $f^\mathbb{Q}(\cdot) := f(\cdot) + \rho g(\cdot) \sigma(\cdot)$ . Let Assumptions 3.1 and 3.2 be satisfied for (27). With the rescaling (2), the sequence  $(X^\varepsilon)_{\varepsilon > 0}$  satisfies a MDP under  $\mathbb{Q}$  with speed  $h^2(\varepsilon)$  and action functional  $I^\mathbb{Q}$  defined as in Theorem 3.3 with  $\bar{\lambda}^\mathbb{Q} = -\bar{\lambda}$ ,  $\lambda^\mathbb{Q}(\cdot) = -\lambda(\cdot)$  and where the invariant measure  $\mu^\mathbb{Q}$  is that of  $Y$  in (27).

*Proof.* From (10), we can write

$$\mathbb{Q}(A) = \mathbb{E}^\mathbb{P} \left[ \exp \left\{ -\frac{1}{2} \int_0^t \sigma(Y_u)^2 du + \rho \int_0^t \sigma(Y_u) dZ_u + \bar{\rho} \int_0^t \sigma(Y_u) dZ_u^\top \right\} \mathbf{1}_{\{A\}} \right],$$

where we decomposed the Brownian motion  $W$  into  $W = \rho Z + \rho Z^\top$ . Girsanov's Theorem implies that the two processes  $Z^\mathbb{Q}$  and  $\bar{Z}^\mathbb{Q}$  defined as

$$Z_t^\mathbb{Q} := Z_t - \rho \int_0^t \sigma(Y_u) du \quad \text{and} \quad \bar{Z}_t^\mathbb{Q} := Z_t^\top - \bar{\rho} \int_0^t \sigma(Y_u) du$$

are two independent standard Brownian motions under  $\mathbb{Q}$ , and we can rewrite (10) as (27) with  $f^\mathbb{Q}(\cdot) := f(\cdot) + \rho g(\cdot) \sigma(\cdot)$  and  $W^\mathbb{Q} := \rho Z^\mathbb{Q} + \bar{\rho} \bar{Z}^\mathbb{Q}$ . The lemma then follows directly from Theorem 3.3 under the assumptions on the coefficients.  $\square$

Note the flipped sign in the drift of  $X$  in (27). We can thus define  $J^\mathbb{Q}$  as the contraction (onto the last point) of the rate function  $I^\mathbb{Q}$  for  $X^\varepsilon$  under  $\mathbb{Q}$ . The minimising functional  $\phi$  in Theorem 3.3 is linear, of the form  $\phi_t = \alpha t + x_0$ , hence

$$(28) \quad J(x) = \frac{(x - x_0)^2 T}{2q} \quad \text{and} \quad J^\mathbb{Q}(x) = \frac{(x - x_0)^2 T}{2q^\mathbb{Q}}, \quad \text{for any } x \in \mathbb{R}.$$

**Proposition 5.5.** *Let  $\beta \in (0, 1/2)$ . As  $t$  tends to infinity, we observe the following asymptotic behaviours:*

$$\begin{aligned} \frac{1}{t^{1/2+\beta}} \log \mathbb{E}^\mathbb{P} \left( e^{x t^{\beta+1/2}} - e^{X_t} \right)_+ &\approx \begin{cases} x - t^{\beta-1/2} J(x), & \text{if } x < x_0, \\ x, & \text{if } x \geq x_0, \end{cases} \\ \lim_{t \uparrow \infty} \frac{1}{t^{2\beta}} \log \mathbb{E}^\mathbb{P} \left( e^{X_t} - e^{x t^{\beta+1/2}} \right)_+ &= \begin{cases} -J^\mathbb{Q}(x), & \text{if } x > x_0, \\ 0, & \text{if } x \leq x_0. \end{cases} \end{aligned}$$

*Proof.* We follow the methodology developed in [22, Theorem 13] with a few changes. For all  $t \geq 1$  and  $\varepsilon > 0$ , the inequalities

$$e^{xt^{\beta+1/2}} (1 - e^{-\varepsilon}) \mathbf{1}_{\{X_t/t^{\beta+1/2} < x - \varepsilon\}} \leq \left( e^{xt^{\beta+1/2}} - e^{X_t} \right)_+ \leq e^{xt^{\beta+1/2}} \mathbf{1}_{\{X_t/t^{\beta+1/2} < x\}}$$

hold. Taking expectations, logarithms and dividing by  $t^{2\beta}$ , we obtain

$$\frac{x}{t^{\beta-1/2}} + \frac{\log(1 - e^{-\varepsilon})}{t^{2\beta}} + \frac{\log \mathbb{P}\left(\frac{X_t}{t^{\beta+1/2}} < x - \varepsilon\right)}{t^{2\beta}} \leq \frac{1}{t^{2\beta}} \log \mathbb{E} \left( e^{xt^{\beta+1/2}} - e^{X_t} \right)_+ \leq \frac{x}{t^{\beta-1/2}} + \frac{\log \mathbb{P}\left(\frac{X_t}{t^{\beta+1/2}} < x\right)}{t^{2\beta}}.$$

From the moderate deviations principle in Theorem 3.3 and using (28), we can write

$$\lim_{t \uparrow \infty} t^{-2\beta} \log \mathbb{P} \left( \frac{X_t}{t^{\beta+1/2}} < x \right) = - \inf_{z < x} J(z) = \begin{cases} -J(x), & \text{if } x < x_0, \\ 0, & \text{if } x \geq x_0, \end{cases}$$

so that, as  $t$  tends to infinity, the asymptotic behaviour for Put option prices in the proposition holds.

For Call option prices on  $\exp(X)$ , we can use the inequalities, for  $t \geq 1$  and  $\varepsilon > 0$ ,

$$e^{X_t} (1 - e^{-\varepsilon}) \mathbf{1}_{\{X_t/t^{\beta+1/2} > x + \varepsilon\}} \leq \left( e^{X_t} - e^{xt^{\beta+1/2}} \right)_+ \leq e^{X_t} \mathbf{1}_{\{X_t/t^{\beta+1/2} > x\}}.$$

Taking expectations under  $\mathbb{P}$ , this becomes

$$(1 - e^{-\varepsilon}) \mathbb{E}^{\mathbb{P}} \left( e^{X_t} \mathbf{1}_{\{X_t/t^{\beta+1/2} > x + \varepsilon\}} \right) \leq \mathbb{E}^{\mathbb{P}} \left( e^{X_t} - e^{xt^{\beta+1/2}} \right)_+ \leq \mathbb{E}^{\mathbb{P}} \left( e^{X_t} \mathbf{1}_{\{X_t/t^{\beta+1/2} > x\}} \right),$$

and, using the Share measure  $\mathbb{Q}$ , this translates into

$$(1 - e^{-\varepsilon}) \mathbb{Q} \left( \frac{X_t}{t^{\beta+1/2}} > x + \varepsilon \right) \leq \mathbb{E}^{\mathbb{P}} \left( e^{X_t} - e^{xt^{\beta+1/2}} \right)_+ \leq \mathbb{Q} \left( \frac{X_t}{t^{\beta+1/2}} > x \right).$$

Using Lemma 5.4 and (28), the proposition then follows from the computation

$$\lim_{t \uparrow \infty} t^{-2\beta} \mathbb{Q} \left( \frac{X_t}{t^{\beta+1/2}} > x \right) = - \inf_{z > x} J^{\mathbb{Q}}(z) = \begin{cases} -J^{\mathbb{Q}}(x), & \text{if } x > x_0, \\ 0, & \text{if } x \leq x_0. \end{cases}$$

□

**5.4. Asymptotics of the realised variance.** We now show how Theorem 4.5 applies to the realised variance in the Heston model (9). The rescaling (2) yields the same perturbed system (15) as in the large-time case. Likewise, the invariant measure  $\mu(\cdot)$  has Gamma density given in (16). With  $H(x, y) \equiv y$ , we have  $q_\sigma = q_g = 1/2$  and  $q_H = 1$ , so that Assumption 4.3 holds and Theorem 4.5 gives the pathwise large deviations for the process

$$(29) \quad R^\varepsilon = \frac{1}{\sqrt{\varepsilon}h(\varepsilon)} \left( \int_0^\cdot Y_s^\varepsilon ds - \theta \right).$$

In this case, the Poisson equation (23) satisfies  $u_y(y) = -\frac{1}{\kappa}$ , and hence  $\overline{Q}(x) \equiv \frac{\xi^2 \theta}{\kappa^2}$ , yielding the rate function

$$(30) \quad I(\phi) = \begin{cases} \frac{1}{2} \frac{\gamma^2 \kappa^2}{\xi^2 \theta} \int_0^T |\dot{\phi}_s|^2 ds, & \text{if } \phi \in \mathcal{AC}([0, T], \mathbb{R}) \text{ and } \phi_0 = 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Take for simplicity  $\gamma = 1$ , i.e.  $\varepsilon = \delta$ , then we can write, for any  $t \geq 0$ ,

$$R_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}h(\varepsilon)} \left( \int_0^t Y_s^\varepsilon ds - \theta \right) = \frac{1}{\sqrt{\varepsilon}h(\varepsilon)} \int_0^t Y_s^\varepsilon ds - \theta_\varepsilon = \frac{\sqrt{\varepsilon}}{h(\varepsilon)} \int_0^{t/\varepsilon} Y_u ds - \theta_\varepsilon = \frac{\sqrt{\varepsilon}}{h(\varepsilon)} \mathcal{V}_{t/\varepsilon} - \theta_\varepsilon$$

where we denote  $\theta_\varepsilon := \frac{\theta}{\sqrt{\varepsilon}h(\varepsilon)}$ , and  $\mathcal{V}_t := \int_0^t Y_u du$  represents the realised variance over the period  $[0, t]$ . The large deviations for the sequence  $R^\varepsilon$  therefore implies that, taking  $t = 1$ , and applying the contraction principle, for any fixed  $x > 0$ ,

$$\lim_{\varepsilon \downarrow 0} \frac{1}{h^2(\varepsilon)} \log \mathbb{P} \left( \frac{\sqrt{\varepsilon}}{h(\varepsilon)} \mathcal{V}_{1/\varepsilon} - \theta_\varepsilon \geq x \right) = - \inf \left\{ I(\phi) : \phi_1 \geq x \right\},$$

or, taking for example  $h(\varepsilon) = \varepsilon^{-\beta}$  for  $\beta \in (0, 1/2)$ , and renaming  $\varepsilon \mapsto t^{-1}$ ,

$$\lim_{t \uparrow \infty} t^{-2\beta} \log \mathbb{P} \left( \mathcal{V}_t \geq xt^{\beta+1/2} + \theta t \right) = - \inf \left\{ I(\phi) : \phi_1 \geq x \right\}.$$

Here again, this minimum can be computed in closed form from (30). The corresponding Euler-Lagrange equation yields the linear optimal path  $\phi_t = xt$ , for any  $x > 0$ , so that

$$J_{\mathcal{V}}(x) := \inf \left\{ I(\phi) : \phi_1 \geq x \right\} = \frac{\kappa^2 x^2}{2\xi^2 \theta} = \frac{x^2}{2\bar{Q}}.$$

Note that, since the drift and diffusion of  $X$  in (9) do not depend on  $X$  itself, then  $\bar{Q}$  is constant here. In the Heston case (9), we can compare this large-time MDP to the corresponding large-time LDP for the realised variance. Indeed, from [4, Proposition 6.3.4.1], the moment generating function for the realised variance reads, for all  $u$  such that the right-hand side is finite,

$$\Lambda(u, t) := \log \mathbb{E} (e^{u\mathcal{V}_t}) = \frac{2\kappa\theta}{\xi^2} \log \left( \frac{2\gamma(u) \exp \left\{ (\kappa + \gamma(u)) \frac{t}{2} \right\}}{\gamma(u) (e^{\gamma(u)t} + 1) + \kappa (e^{\gamma(u)t} - 1)} \right) + \frac{2uy_0 (e^{\gamma(u)t} - 1)}{\gamma(u) (e^{\gamma(u)t} + 1) + \kappa (e^{\gamma(u)t} - 1)},$$

where  $\gamma(u) := \sqrt{\kappa^2 - 2\xi^2 u}$ . Straightforward computations yield that

$$\Lambda_{\infty}(u) := \lim_{t \uparrow \infty} \frac{\Lambda(u, t)}{t} = \frac{\kappa\theta}{\xi^2} (\kappa - \gamma(u)), \quad \text{for all } u < \frac{\kappa^2}{2\xi^2}.$$

We can therefore use a partial version of the Gärtner-Ellis theorem [6, Section 2.3] to show that the sequence  $(t^{-1}\mathcal{V}_t)_{t>0}$  satisfies a large deviations principle as  $t$  tends to infinity, with rate function  $\Lambda^*$  defined as the Fenchel-Legendre transform of  $\Lambda_{\infty}$ . More precisely, for any  $x > 0$ ,

$$\Lambda^*(x) = \sup_{u < \frac{\kappa^2}{2\xi^2}} \left\{ ux - \Lambda_{\infty}(u) \right\} = \frac{\kappa^2(x - \theta)^2}{2\xi^2 x} \quad \text{and} \quad \partial_{xx} \Lambda^*(x)|_{x=\theta} = \frac{\kappa^2}{\xi^2 \theta} = \bar{Q}^{-1}.$$

Again, we observe that the moderate deviations rate function characterises the local curvature of the large deviations rate function around its minimum. We can translate this behaviour into asymptotics of options on the realised variance, both in the large deviations case and in the moderate deviations one:

**Proposition 5.6.** *Let  $\beta \in (0, 1/2)$ . As  $t$  tends to infinity, we have, for all  $x > 0$ ,*

$$\lim_{t \uparrow \infty} \frac{1}{t} \log \mathbb{E} (e^{xt} - e^{\mathcal{V}_t})_+ = x - \Lambda^*(x) \quad \text{and} \quad \lim_{t \uparrow \infty} \left[ \frac{1}{t^{2\beta}} \log \mathbb{E} \mathbb{P} \left( e^{xt^{\beta+1/2}} - e^{\mathcal{V}_t - \theta t} \right)_+ - \frac{x}{t^{\beta-1/2}} \right] = -J_{\mathcal{V}}(x).$$

**Remark 5.7.** Note that, again, formally setting  $\beta = 1/2$  in the second limit, which arises from moderate deviations, one obtains after simplification,

$$\lim_{t \uparrow \infty} \frac{1}{t} \log \mathbb{E} \left( e^{(x+\theta)t} - e^{\mathcal{V}_t} \right)_+ = x - \theta - J_{\mathcal{V}}(x),$$

which again, as in Remark 3.5, indicates that the moderate deviations rate function represents exactly the curvature of the large deviations one at its minimum.

*Proof.* We start with the first limit, which falls in the scope of large deviations. Mimicking the proof of Proposition 5.5, for all  $t \geq 1$  and  $\varepsilon > 0$ , we can write the inequalities

$$e^{xt} (1 - e^{-\varepsilon}) \mathbf{1}_{\{\mathcal{V}_t/t < x - \varepsilon\}} \leq (e^{xt} - e^{\mathcal{V}_t})_+ \leq e^{xt} \mathbf{1}_{\{\mathcal{V}_t/t < x\}}.$$

Taking expectations, logarithms and dividing by  $t$  and taking limits, we obtain

$$x + \lim_{t \uparrow \infty} \frac{1}{t} \log \mathbb{P} \left( \frac{\mathcal{V}_t}{t} < x - \varepsilon \right) \leq \lim_{t \uparrow \infty} \frac{1}{t} \log \mathbb{E} (e^{xt} - e^{\mathcal{V}_t})_+ \leq x + \lim_{t \uparrow \infty} \frac{1}{t} \log \mathbb{P} \left( \frac{\mathcal{V}_t}{t} < x \right).$$

The large deviations obtained above for the sequence  $(\mathcal{V}_t/t)_{t>0}$  as  $t$  tends to infinity concludes the proof.

We now prove the other limit, which arises from moderate deviations. Now, mimicking the proof of Proposition 5.5, for all  $t \geq 1$  and  $\varepsilon > 0$ , we can write

$$e^{xt^{\beta+1/2}} (1 - e^{-\varepsilon}) \mathbf{1}_{\{\tilde{\mathcal{V}}_t/t^{\beta+1/2} < x - \varepsilon\}} \leq (e^{xt^{\beta+1/2}} - e^{\tilde{\mathcal{V}}_t})_+ \leq e^{xt^{\beta+1/2}} \mathbf{1}_{\{\tilde{\mathcal{V}}_t/t^{\beta+1/2} < x\}},$$

with  $\tilde{\mathcal{V}}_t := \mathcal{V}_t - \theta t$ , so that, taking expectations, logarithms and dividing by  $t^{2\beta}$ , the proposition follows from

$$\frac{x}{t^{\beta-1/2}} + \frac{1}{t^{2\beta}} \log \mathbb{P} \left( \frac{\tilde{\mathcal{V}}_t}{t^{\beta+1/2}} < x - \varepsilon \right) \leq \frac{1}{t^{2\beta}} \log \mathbb{E} (e^{xt^{\beta+1/2}} - e^{\tilde{\mathcal{V}}_t})_+ \leq \frac{x}{t^{\beta-1/2}} + \frac{1}{t^{2\beta}} \log \mathbb{P} \left( \frac{\tilde{\mathcal{V}}_t}{t^{\beta+1/2}} < x \right).$$

□

## 6. CONTROL REPRESENTATIONS FOR THE PROOF OF THEOREM 4.5

In this section we connect Theorem 4.5 to certain stochastic control representation and to the Laplace principle. We start by applying Itô formula to the solution  $u$  to (23). Rearranging terms, we obtain

$$\begin{aligned} R_t^\varepsilon &= \frac{\delta^2}{\varepsilon^{3/2}h(\varepsilon)} (u(X_t^\varepsilon, Y_t^\varepsilon) - u(x_0, y_0)) - \frac{\delta}{\varepsilon h(\varepsilon)} \int_0^t (gu_y)(X_s^\varepsilon, Y_s^\varepsilon) dZ_s - \frac{\delta^2}{\varepsilon h(\varepsilon)} \int_0^t (\sigma u_x)(X_s^\varepsilon, Y_s^\varepsilon) dW_s \\ &\quad - \frac{\rho\delta}{\sqrt{\varepsilon}h(\varepsilon)} \int_0^t (\sigma gu_{xy})(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{\delta}{2\sqrt{\varepsilon}h(\varepsilon)} \int_0^t (\sigma^2 u_x)(X_s^\varepsilon, Y_s^\varepsilon) ds - \frac{\delta^2}{2\sqrt{\varepsilon}h(\varepsilon)} \int_0^t (\sigma^2 u_{xx})(X_s^\varepsilon, Y_s^\varepsilon) ds. \end{aligned}$$

We shall consider  $R_t^\varepsilon$  through this expression together with (18) as the triple  $(R_t^\varepsilon, X_t^\varepsilon, Y_t^\varepsilon)$ . By [9, Section 1.2], the large deviations principle (with speed  $h^2(\varepsilon)$ ) for  $R^\varepsilon$  is equivalent to the Laplace principle, which states that for any bounded continuous function  $G$  mapping  $\mathcal{C}([0, T]; \mathbb{R})$  to  $\mathbb{R}$ ,

$$(31) \quad \lim_{\varepsilon \downarrow 0} \frac{1}{h^2(\varepsilon)} \log \mathbb{E} [\exp \{-h^2(\varepsilon)G(R^\varepsilon)\}] = \inf_{\phi \in \mathcal{C}([0, T]; \mathbb{R})} \{I(\phi) + G(\phi)\},$$

where  $I(\cdot)$  is called the action functional. We prove (31) and then Theorem 4.5 identifies  $I(\phi)$ . The proof of (31) is based on appropriate stochastic control representations developed in [3]. Let  $\mathcal{A}$  be the space of  $\mathcal{F}_t$ -progressively measurable two-dimensional processes  $v = (v_1, v_2)$  such that  $\mathbb{E} \left( \int_0^T |v(s)|^2 ds \right)$  is finite. From [3, Theorem 3.1], we can write the stochastic control representation

$$(32) \quad -\frac{1}{h^2(\varepsilon)} \log \mathbb{E} [\exp \{-h^2(\varepsilon)G(R^\varepsilon)\}] = \inf_{u^\varepsilon \in \mathcal{A}} \mathbb{E} \left[ \frac{1}{2} \int_0^T |u^\varepsilon(s)|^2 ds + G(R^{\varepsilon, u^\varepsilon}) \right]$$

where the controlled process  $(R_t^{\varepsilon, u^\varepsilon}, X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon})$  satisfies the system

$$\begin{aligned} R_t^{\varepsilon, u^\varepsilon} &= \frac{\delta^2}{\varepsilon^{3/2}h(\varepsilon)} (u(X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) - u(x_0, y_0)) - \frac{\delta}{\varepsilon h(\varepsilon)} \int_0^t (gu_y)(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) dZ_s \\ &\quad - \frac{\delta}{\varepsilon} \rho \int_0^t (gu_y)(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) u_1^\varepsilon(s) ds - \frac{\delta}{\varepsilon} \bar{\rho} \int_0^t (gu_y)(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) u_2^\varepsilon(s) ds \\ &\quad - \frac{\delta^2}{\varepsilon} \rho \int_0^t (\sigma u_x)(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) u_1^\varepsilon(s) ds \\ (33) \quad &- \rho \frac{\delta}{\sqrt{\varepsilon}h(\varepsilon)} \int_0^t (\sigma gu_{xy})(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) ds + \frac{1}{2} \frac{\delta}{\sqrt{\varepsilon}h(\varepsilon)} \int_0^t (\sigma^2 u_x)(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) ds \\ &- \frac{1}{2} \frac{\delta^2}{\sqrt{\varepsilon}h(\varepsilon)} \int_0^t (\sigma^2 u_{xx})(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) ds - \frac{\delta^2}{\varepsilon h(\varepsilon)} \int_0^t (\sigma u_x)(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) dW_s \\ dX_t^{\varepsilon, u^\varepsilon} &= -\frac{1}{2} \frac{\varepsilon}{\delta} \sigma^2 (X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) dt + \sqrt{\varepsilon} h(\varepsilon) \sigma (X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) u_1^\varepsilon(t) dt + \sqrt{\varepsilon} \sigma (X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) dW_t, \\ dY_t^{\varepsilon, u^\varepsilon} &= \frac{\varepsilon}{\delta^2} f(X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) dt + \frac{\sqrt{\varepsilon} h(\varepsilon)}{\delta} g(X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) [\rho u_1^\varepsilon(t) + \bar{\rho} u_2^\varepsilon(t)] dt + \frac{\sqrt{\varepsilon}}{\delta} g(X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) dZ_t, \end{aligned}$$

with  $d\langle W, Z \rangle_t = \rho dt$ . With this control representation at hand, we proceed by analysing the limit of the right-hand side of (32) as  $\varepsilon \downarrow 0$ . Similarly to [26], we define the function  $\theta : \mathbb{R}^4 \mapsto \mathbb{R}$  by

$$(34) \quad \theta(x, y, z_1, z_2) := -\gamma^{-1} g(x, y) u_y(x, y) (\rho z_1 + \bar{\rho} z_2).$$

Let  $\mathcal{Z} = \mathbb{R}$  define the control space, and  $\mathcal{Y}$  the state space of  $Y$ . Our assumptions guarantee that  $\theta$  is bounded in  $x$ , affine in  $z_1, z_2$  growing at most polynomially in  $y$ . Next step is to introduce an appropriate family of occupation measures whose role is to single out the correct averaging taking place in the limit. For this reason, let  $0 < \Delta = \Delta(\varepsilon) \downarrow 0$  be a parameter whose role is to exploit a time-scale separation. Let  $A_1, A_2, B, \Gamma$  be Borel sets of  $\mathcal{Z}, \mathcal{Z}, \mathcal{Y}, [0, T]$  respectively. Then, define the occupation measure  $P^{\varepsilon, \Delta}$  by

$$(35) \quad P^{\varepsilon, \Delta}(A_1 \times A_2 \times B \times \Gamma) := \int_\Gamma \left[ \frac{1}{\Delta} \int_t^{t+\Delta} \mathbf{1}_{A_1}(u_1^\varepsilon(s)) \mathbf{1}_{A_2}(u_2^\varepsilon(s)) \mathbf{1}_B(Y_s^{\varepsilon, u^\varepsilon}) ds \right] dt,$$

and assume  $u_i^\varepsilon(s) = 0$  if  $s > T$ . For a Polish space  $S$ , let  $\mathcal{P}(S)$  be the space of probability measures on  $S$ . Next we recall the definition of a viable pair for the moderate deviations case.

**Definition 6.1** (Definition 4.1 in [26]). Let  $\theta(x, y, z_1, z_2): \mathbb{R} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{Z} \rightarrow \mathbb{R}$  be a function that grows at most linearly in  $y$ . For each  $x \in \mathbb{R}$ , let  $\mathcal{L}_x$  be a second-order elliptic partial differential operator and denote by  $\mathcal{D}(\mathcal{L}_x)$  its domain of definition. A pair  $(\psi, P) \in \mathcal{C}([0, T]; \mathbb{R}) \times \mathcal{P}(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T])$  is called a viable pair with respect to  $(\theta, \mathcal{L}_x)$ , and we write  $(\psi, P) \in \mathcal{V}(\theta, \mathcal{L}_x)$ , if

- the function  $\psi$  is absolutely continuous;
- the measure  $P$  is integrable in the sense that

$$\int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T]} (|z_1|^2 + |z_2|^2 + |y|^2) P(dz_1 dz_2 dy ds) < \infty;$$

- for all  $t \in [0, T]$  (with  $\bar{X}$  defined in (21)),

$$(36) \quad \psi_t = \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, t]} \theta(\bar{X}_s, y, z_1, z_2) P(dz_1 dz_2 dy ds);$$

- for all  $t \in [0, T]$  and for every  $F \in \mathcal{D}(\mathcal{L}_x)$ ,

$$(37) \quad \int_0^t \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y}} \mathcal{L}_{\bar{X}_s} F(y) P(dz_1 dz_2 dy ds) = 0;$$

- for all  $t \in [0, T]$ ,

$$(38) \quad P(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, t]) = t.$$

The last item is equivalent to stating that the last marginal of  $P$  is Lebesgue measure, or that  $P$  can be decomposed as  $P(dz_1 dz_2 dy dt) = P_t(dz_1 dz_2 dy) dt$ . Then we shall establish the following result:

**Theorem 6.2.** *Under Assumptions 4.1, 4.2, 4.3 and 4.4, the family of processes  $\{R^\varepsilon, \varepsilon > 0\}$  from (22) satisfies the large deviations principle, with action functional*

$$(39) \quad I(\phi) = \inf_{(\phi, P) \in \mathcal{V}(\theta, \mathcal{L}_x)} \left\{ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T]} (|z_1|^2 + |z_2|^2) P(dz_1 dz_2 dy ds) \right\}$$

with the convention that the infimum over the empty set is infinite.

As will be shown during the proof, Theorem 4.5 follows directly from Theorem 6.2.

## 7. PROOF OF THEOREM 4.5

In this section we offer the proof of Theorem 4.5. As mentioned in Section 6, the proof of Theorem 4.5 is equivalent to that of the Laplace principle in Theorem 6.2. In Subsections 7.1 and 7.2 we prove tightness and convergence of the pair  $(R^{\varepsilon, u^\varepsilon}, P^{\varepsilon, \Delta})$  respectively. In Subsection 7.3, we finally establish the Laplace principle and the representation formula of Theorem 4.5. The proofs of these results make use of the results of [26]. Below we present the main arguments, highlighting the differences and give exact pointers to [26] when appropriate.

**7.1. Tightness of the pair  $\{(R^{\varepsilon, u^\varepsilon}, P^{\varepsilon, \Delta}), \varepsilon > 0\}$ .** In this section we prove the following proposition:

**Proposition 7.1.** *Under Assumptions 4.1, 4.2 and 4.3, for a family  $\{u^\varepsilon, \varepsilon > 0\}$  of controls in  $\mathcal{A}$  satisfying*

$$\sup_{\varepsilon > 0} \int_0^T |u^\varepsilon(t)|^2 dt < N \quad \text{almost surely,}$$

for some  $N < \infty$ , the following hold:

- (i) the family  $\{(R^{\varepsilon, u^\varepsilon}, P^{\varepsilon, \Delta}), \varepsilon > 0\}$  is tight on  $\mathcal{C}([0, T]; \mathbb{R}) \times \mathcal{P}(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T])$ ;
- (ii) With the set

$$\mathcal{B}_M := \left\{ (z_1, z_2, y) \in \mathcal{Z} \times \mathcal{Z} \times \mathbb{R} : |z_1| > M, |z_2| > M, |y| > M \right\},$$

the family  $\{P^{\varepsilon, \Delta}, \varepsilon > 0\}$  is uniformly integrable in the sense that

$$\lim_{M \uparrow \infty} \sup_{\varepsilon > 0} \mathbb{E}_{x_0, y_0} \left[ \int_{(z_1, z_2, y) \in \mathcal{B}_M \times [0, T]} [|z_1| + |z_2| + |y|] P^{\varepsilon, \Delta}(dz_1 dz_2 dy dt) \right] = 0.$$

*Proof of Proposition 7.1.* Tightness of  $\{P^{\varepsilon, \Delta}, \varepsilon, \Delta > 0\}$  on  $\mathcal{P}(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T])$  and the uniform integrability of the family of occupation measures is the subject of [26, Section 5.1.1], and will not be repeated here. It remains to prove tightness of  $\{R^{\varepsilon, u^\varepsilon}, \varepsilon > 0\}$  on  $\mathcal{C}([0, T]; \mathbb{R})$ . We prove it making use of the characterisation of [2, Theorem 8.7], and it follows if we establish that there is  $\varepsilon_0 > 0$  such that for every  $\eta > 0$ ,

(i) there exists  $N < \infty$  such that

$$(40) \quad \mathbb{P} \left( \sup_{0 \leq t \leq T} |R_t^{\varepsilon, u^\varepsilon}| > N \right) \leq \eta, \quad \text{for every } \varepsilon \in (0, \varepsilon_0);$$

(ii) for every  $M < \infty$ ,

$$(41) \quad \lim_{\alpha \downarrow 0} \sup_{\varepsilon \in (0, \varepsilon_0)} \mathbb{P} \left( \sup_{|t_1 - t_2| < \alpha, 0 \leq t_1 < t_2 \leq T} |R_{t_1}^{\varepsilon, u^\varepsilon} - R_{t_2}^{\varepsilon, u^\varepsilon}| \geq \eta, \sup_{t \in [0, T]} |R_t^{\varepsilon, u^\varepsilon}| \leq M \right) = 0.$$

Essentially both statements follow from the control representation (33) together with the results on the growth of the solution to the Poisson equation by Theorem A.1 and Lemma A.2 and using Lemma B.4 to treat each term on the right-hand side of (33). For sake of completeness, we prove the first statement (40). The second statement (41) follows similarly using the general purpose Lemma B.4. Rewrite (33) as

$$(42) \quad R_t^{\varepsilon, u^\varepsilon} = \sum_{i=1}^9 R_t^{i, \varepsilon},$$

where  $R_t^{i, \varepsilon}$  represents the  $i^{\text{th}}$  term on the right-hand side of (33), and notice that

$$\mathbb{P} \left( \sup_{0 \leq t \leq T} |R_t^{\varepsilon, u^\varepsilon}| > N \right) \leq \sum_{i=1}^9 \mathbb{P} \left( \sup_{0 \leq t \leq T} |R_t^{i, \varepsilon}| > \frac{N}{9} \right).$$

Theorem A.1 and an application of Hölder inequality imply that, for  $q_H < 2$ ,

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in [0, T]} |R_t^{1, \varepsilon}| \right) &\leq \frac{2\delta^2}{\varepsilon^{3/2}h(\varepsilon)} \left( 1 + \mathbb{E} \sup_{t \in [0, T]} |Y_t^{\varepsilon, u^\varepsilon}|^{q_H} \right) \leq \frac{2\delta^2}{\varepsilon^{3/2}h(\varepsilon)} \left( 1 + \left( \mathbb{E} \sup_{t \in [0, T]} |Y_t^{\varepsilon, u^\varepsilon}|^2 \right)^{q_H/2} \right) \\ &\leq K \frac{\delta^2}{\varepsilon^{3/2}h(\varepsilon)} \left( 1 + h(\varepsilon)^{\frac{q_H}{1-q_H}} \right) = K \frac{\delta^2}{\varepsilon^2} \sqrt{\varepsilon} h(\varepsilon)^{\frac{q_H + q_H - 1}{1 - q_H}}, \end{aligned}$$

where the second inequality follows from Lemma B.3, which stays bounded by Assumption 4.3 (actually, it goes to zero). Hence we obtain that for some unimportant constant  $K < \infty$ ,

$$\mathbb{P} \left( \sup_{t \in [0, T]} |R_t^{1, \varepsilon}| > \frac{N}{9} \right) \leq \frac{K}{N}.$$

For the stochastic integral term  $R_t^{2, \varepsilon}$ , we have that for a constant  $C < \infty$  that may change from line to line and for  $\nu > 0$  small enough such that  $q_{gu_y}(1 + \nu) < 1$  (with some abuse of notation  $q_{gu_y}$  is the degree of polynomial growth in  $|y|$  of the function  $g(x, \cdot)u_y(x, \cdot)$ ), we have

$$\begin{aligned} \mathbb{P} \left( \sup_{t \in [0, T]} \left| \int_0^t gu_y(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) dZ_s \right| > \frac{\varepsilon h(\varepsilon)N}{9\delta} \right) &\leq \\ &\leq \frac{C}{N^{2(1+\nu)}} \left( \frac{\delta}{\varepsilon h(\varepsilon)} \right)^{2(1+\nu)} \mathbb{E} \left( \sup_{t \in [0, T]} \left| \int_0^t gu_y(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) dW_s \right|^{2(1+\nu)} \right) \\ &\leq \frac{C}{N^{2(1+\nu)}} \left( \frac{\delta}{\varepsilon h(\varepsilon)} \right)^{2(1+\nu)} \mathbb{E} \left( \sup_{t \in [0, T]} \left| \int_0^t |gu_y(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon})|^2 ds \right|^{(1+\nu)} \right), \end{aligned}$$

from which the result follows by Lemma B.4 given that  $q_{gu_y} < 1$ . Similarly, we obtain a corresponding bound for the other stochastic integral term  $R_t^{9, \varepsilon}$ . The Riemann integral terms  $R_t^{3, \varepsilon}, \dots, R_t^{8, \varepsilon}$  are treated very similarly again using Lemma B.4. The conditions on  $q_\sigma, q_g$  and  $q_H$  from Assumption 4.3 guarantee that Lemma B.4 applies in these cases. These considerations yield (40).



The second statement (41) follows by similar arguments using again Lemma B.4 and the growth properties of the involved functions with respect to  $|y|$ . The only term that potentially needs some discussion is the  $R_t^{1,\varepsilon}$  term. In particular, we need to show that for every  $\eta > 0$ , there exists  $\varepsilon_0 > 0$  such that

$$\sup_{\varepsilon \in (0, \varepsilon_0)} \mathbb{P} \left( \sup_{t \in [0, T]} |R_t^{1,\varepsilon}| > \eta \right) \leq \eta.$$

But this follows again by the estimate on  $\mathbb{E} \left( \sup_{t \in [0, T]} |R_t^{1,\varepsilon}| \right)$  above together with Assumption 4.3. This completes the proof of the proposition.  $\square$

**7.2. Convergence of the pair  $\{(R^{\varepsilon, u^\varepsilon}, P^{\varepsilon, \Delta}), \varepsilon > 0\}$ .** In Section 7.1 we proved that the family of processes  $\{(R^{\varepsilon, u^\varepsilon}, P^{\varepsilon, \Delta}), \varepsilon > 0\}$  is tight. It follows that for any subsequence of  $\varepsilon$  converging to 0, there exists a subsubsequence of  $(R^{\varepsilon, u^\varepsilon}, P^{\varepsilon, \Delta})$  which converges in distribution to some limit  $(\bar{R}, \bar{P})$ . The goal of this section is to show that  $(\bar{R}, \bar{P})$  is a viable pair with respect to  $(\theta, \mathcal{L}_x)$ , according to Definition 6.1. To show that the limit point  $(\bar{R}, \bar{P})$  is a viable pair, we must show that it satisfies (36), (37), and (38). The proof of (37) and (38) can be found in [26, Section 5.2], whereas the proof of statement (36) follows via the Skorokhod Representation Theorem [2, Theorem 6.7] and the martingale problem formulation. For completeness, we present the argument below. We will invoke the Skorokhod Representation Theorem, which allows us to assume that there exists a probability space in which the desired convergence occurs with probability one. Let us define

$$\Psi_t^{\varepsilon, u^\varepsilon} := R_t^{\varepsilon, u^\varepsilon} - \frac{\delta^2}{\varepsilon^{3/2} h(\varepsilon)} \left( u(X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) - u(x_0, y_0) \right).$$

As in the proof of tightness, we notice that under Assumption 4.3,

$$\lim_{\varepsilon \downarrow 0} \frac{\delta^2}{\varepsilon^{3/2} h(\varepsilon)} \mathbb{E} \left( \sup_{t \in [0, T]} |u(X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) - u(x_0, y_0)| \right) = 0,$$

so that it is enough to consider the limit in distribution of the family  $\{\Psi^{\varepsilon, u^\varepsilon}, \varepsilon > 0\}$ . The rest of the argument is now classical.

Consider an arbitrary function  $G$  that is real valued, smooth with compact support on  $\mathbb{R}$ . Fix two positive integers  $p_1$  and  $p_2$  and let  $\phi_j, j = 1, \dots, p_1$ , be real valued, smooth functions with compact support on  $\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T]$ . Let  $S_1, S_2$ , and  $t_i, i = 1, \dots, p_2$ , be nonnegative real numbers with  $t_i \leq S_1 < S_1 + S_2 \leq T$ . Let  $\zeta$  be a real-valued, bounded and continuous function with compact support on  $\mathbb{R}^{p_2} \times \mathbb{R}^{p_1 p_2}$ . Given a measure  $P_0 \in \mathcal{P}(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T])$  and  $t \in [0, T]$ , define

$$(P_0, \phi_j)_t := \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, t]} \phi_j(z_1, z_2, y, s) P_0(dz_1 dz_2 dy ds),$$

the empirical measure

$$P_t^{\varepsilon, \Delta}(dz_1 dz_2 dy) := \frac{1}{\Delta} \int_t^{t+\Delta} \mathbf{1}_{dz_1}(u_1^\varepsilon(s)) \mathbf{1}_{dz_2}(u_2^\varepsilon(s)) \mathbf{1}_{dy}(Y_s^{\varepsilon, u^\varepsilon}) ds,$$

as well as the operator  $\widehat{\mathcal{L}}_t^{\varepsilon, \Delta}$  as

$$\widehat{\mathcal{L}}_t^{\varepsilon, \Delta} G(\Psi) := \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y}} G'(\Psi) \theta(\bar{X}_t, y, z_1, z_2) P_t^{\varepsilon, \Delta}(dz_1 dz_2 dy).$$

Based on the martingale problem formulation, proving (36) is equivalent to proving that

$$(43) \quad \lim_{\varepsilon \downarrow 0} \mathbb{E} \left[ \zeta \left( \Psi_{t_i}^{\varepsilon, u^\varepsilon}, ((P^{\varepsilon, \Delta}, \phi_j)_{t_i})_{i \leq p_2, j \leq p_1} \right) \left[ G(\Psi_{S_1+S_2}^{\varepsilon, u^\varepsilon}) - G(\Psi_{S_1}^{\varepsilon, u^\varepsilon}) - \int_{S_1}^{S_1+S_2} \widehat{\mathcal{L}}_t^{\varepsilon, \Delta} G(\Psi_t^{\varepsilon, u^\varepsilon}) dt \right] \right] = 0,$$

and

$$(44) \quad \lim_{\varepsilon \downarrow 0} \left\{ \int_{S_1}^{S_1+S_2} \widehat{\mathcal{L}}_t^{\varepsilon, \Delta} G(\Psi_t^{\varepsilon, u^\varepsilon}) dt - \int_{\mathcal{Z} \times \mathcal{Z} \times \mathbb{R} \times [S_1, S_1+S_2]} G'(\bar{\Psi}_t) \theta(\bar{X}_t, y, z_1, z_2) \bar{P}(dz_1 dz_2 dy dt) \right\} = 0.$$

Note first that for every real-valued, continuous function  $\phi$  with compact support and  $t \in [0, T]$ ,  $(P^{\varepsilon, \Delta}, \phi)_t$  converges to  $(\bar{P}, \phi)_t$  with probability one as  $\varepsilon$  tends to zero. Now (44) follows directly by the first statement

of [26, Lemma 5.1]. In order to prove (43) we first apply the Itô formula to  $G(\Psi)$ . Then it is easy to see that all of the terms apart from

$$\frac{\delta}{\varepsilon} \int_0^t G'(\Psi_s)(gu_y)(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon})(\rho u_1^\varepsilon(s) + \bar{\rho} u_2^\varepsilon(s)) ds$$

in the representation (42) will vanish in the limit. This term surviving in the limit together with the second statement of [26, Lemma 5.1] directly yield (43). This concludes the proof of (36).

**7.3. Laplace principle and compactness of level sets.** We prove here the Laplace principle lower and upper bounds and the compactness of level sets of the action functional  $I(\cdot)$ . We start with the lower bound, i.e. we need to show that for all bounded, continuous functions  $G$  mapping  $\mathcal{C}([0, T]; \mathbb{R})$  into  $\mathbb{R}$ ,

$$\liminf_{\varepsilon \downarrow 0} - \frac{\log \mathbb{E} [\exp \{-h^2(\varepsilon)G(R^\varepsilon)\}]}{h^2(\varepsilon)} \geq \inf_{(\phi, P) \in \mathcal{V}(\theta, \mathcal{L}_x)} \left\{ \frac{1}{2} \int [ |z_1|^2 + |z_2|^2 ] P(dz_1 dz_2 dy ds) + G(\phi) \right\}.$$

It is sufficient to prove it along any subsequence such that  $-\frac{1}{h^2(\varepsilon)} \log \mathbb{E} [e^{-h^2(\varepsilon)G(R^\varepsilon)}]$  converges, which exists due to the uniform bound on the test function  $G$ . In addition, by Lemma B.1, we may assume that

$$\sup_{\varepsilon > 0} \mathbb{E} \int_0^1 |u^\varepsilon(s)|^2 ds \leq N,$$

for some constant  $N$ . For such controls, we construct the family of occupation measures  $P^{\varepsilon, \Delta}$  from (35), and per Proposition 7.1 the family  $\{(R^{\varepsilon, u^\varepsilon}, P^{\varepsilon, \Delta}), \varepsilon > 0\}$  is tight. This indicates that for any subsequence, there is a further subsequence for which  $(R^{\varepsilon, u^\varepsilon}, P^{\varepsilon, \Delta})$  converges to  $(\bar{R}, \bar{P})$  in distribution, with  $(\bar{R}, \bar{P}) \in \mathcal{V}(\theta, \mathcal{L}_x)$  being a viable pair per Definition 6.1. Then using Fatou's lemma, we conclude the proof of the lower bound

$$\begin{aligned} \liminf_{\varepsilon \downarrow 0} - \frac{\log \mathbb{E} [\exp \{-h^2(\varepsilon)G(R^\varepsilon)\}]}{h^2(\varepsilon)} &\geq \liminf_{\varepsilon \downarrow 0} \left\{ \mathbb{E} \left( \frac{1}{2} \int_0^T |u^\varepsilon(s)|^2 ds + G(R^{\varepsilon, u^\varepsilon}) \right) - \varepsilon \right\} \\ &\geq \liminf_{\varepsilon \downarrow 0} \mathbb{E} \left( \frac{1}{2} \int_0^T \frac{1}{\Delta} \int_t^{t+\Delta} |u^\varepsilon(s)|^2 ds dt + G(R^{\varepsilon, u^\varepsilon}) \right) \\ &= \liminf_{\varepsilon \downarrow 0} \mathbb{E} \left( \frac{1}{2} \int_{\mathbb{Z}^2 \times \mathcal{Y} \times [0, T]} [|z_1|^2 + |z_2|^2] P^{\varepsilon, \Delta}(dz_1 dz_2 dy dt) + G(R^{\varepsilon, u^\varepsilon}) \right) \\ &\geq \mathbb{E} \left( \frac{1}{2} \int_{\mathbb{Z}^2 \times \mathcal{Y} \times [0, T]} [|z_1|^2 + |z_2|^2] \bar{P}(dz_1 dz_2 dy dt) + G(\bar{R}) \right) \\ &\geq \inf_{(\xi, P) \in \mathcal{V}(\theta, \mathcal{L}_x)} \left\{ \frac{1}{2} \int_{\mathbb{Z}^2 \times \mathcal{Y} \times [0, T]} [|z_1|^2 + |z_2|^2] P(dz_1 dz_2 dy dt) + G(\phi) \right\}. \end{aligned}$$

Next we want to prove compactness of level sets. Namely, we want to show that for each  $s < \infty$ , the set  $\Phi_s := \{\xi \in \mathcal{C}([0, T]; \mathbb{R}) : I(\phi) \leq s\}$  is a compact subset of  $\mathcal{C}([0, T]; \mathbb{R})$ , which amounts to showing that it is precompact and closed. The proof is analogous to the proof of the lower bound using Fatou's lemma on the form of  $I(\cdot)$  as given by (39) and thus the details are omitted; see also [10] for further details.

It remains to show the upper bound. From [26] we can write  $I(\phi) = \int_0^T L^o(\bar{X}_s, \phi_s, \dot{\phi}_s) ds$ , where

$$\begin{aligned} L^o(x, \eta, \beta) &:= \inf_{(v, \mu) \in \mathcal{A}_{x, \eta, \beta}^o} \frac{1}{2} \int_{\mathcal{Y}} |v(y)|^2 \mu_x(dy), \\ \mathcal{A}_{x, \eta, \beta}^o &:= \left\{ v = (v_1, v_2) : \mathcal{Y} \rightarrow \mathbb{R}^2, \mu \in \mathcal{P}(\mathcal{Y}), \text{ such that } \right. \\ &\quad \left. \begin{aligned} &\int_{\mathcal{Y}} \mathcal{L}_x^Y F(y) \mu_x(dy) = 0, \text{ for all } F \in \mathcal{C}^2(\mathcal{Y}), \\ &\int_{\mathcal{Y}} [|v(y)|^2 + |y|^2] \mu_x(dy) < \infty, \\ &\beta = \int_{\mathcal{Y}} \theta(x, y, v_1(y), v_2(y)) \mu_x(dy) \end{aligned} \right\}. \end{aligned}$$

Then, applying [26, Theorem 5.1] we get that if  $\bar{Q}(x)$  defined in (25) is non-degenerate everywhere then

$$(45) \quad L^o(x, \eta, \beta) = \frac{\beta^2}{2\bar{Q}(x)},$$

and the infimum is attained at

$$u_1(y) = -\frac{\rho\beta g(x,y)u_y(x,y)}{\gamma\bar{Q}(x)} \quad \text{and} \quad u_2(y) = -\frac{\bar{\rho}\beta g(x,y)u_y(x,y)}{\gamma\bar{Q}(x)}.$$

This representation essentially gives the equivalence between Theorems 4.5 and 6.2. Now, we have all the tools needed to prove the Laplace principle upper bound. We need to show that for all bounded, continuous functions  $G$  mapping  $\mathcal{C}([0, T]; \mathbb{R})$  into  $\mathbb{R}$ ,

$$\limsup_{\varepsilon \downarrow 0} -\frac{1}{h^2(\varepsilon)} \log \mathbb{E} [\exp \{-h^2(\varepsilon)G(R^\varepsilon)\}] \leq \inf_{\phi \in \mathcal{C}([0, T]; \mathbb{R})} [I(\phi) + G(\phi)].$$

Let  $\zeta > 0$  be given and consider  $\psi \in \mathcal{C}([0, T]; \mathbb{R})$  with  $\psi_0 = 0$  such that

$$I(\psi) + G(\psi) \leq \inf_{\phi \in \mathcal{C}([0, T]; \mathbb{R})} [I(\phi) + G(\phi)] + \zeta < \infty.$$

Since  $G$  is bounded, this implies that  $I(\psi)$  is finite, and thus  $\psi$  is absolutely continuous. Because  $L^o(x, \eta, \beta)$  is continuous and finite at each  $(x, \eta, \beta) \in \mathbb{R}^3$ , a mollification argument allows us to assume that  $\dot{\psi}$  is piecewise continuous, as in [9, Section 6.5]. Given this  $\psi$  define

$$\bar{u}_1(t, x, y) := -\rho \frac{g(x, y)u_y(x, y)\dot{\psi}_t}{\gamma\bar{Q}(x)} \quad \text{and} \quad \bar{u}_2(t, x, y) := -\bar{\rho} \frac{g(x, y)u_y(x, y)\dot{\psi}_t}{\gamma\bar{Q}(x)}.$$

Define a control in feedback form by

$$\bar{u}^\varepsilon(t) := (\bar{u}_1(t), \bar{u}_2(t)) := (\bar{u}_1(t, \bar{X}_t, Y_t^\varepsilon), \bar{u}_2(t, \bar{X}_t, Y_t^\varepsilon)).$$

Then  $R^{\varepsilon, \bar{u}^\varepsilon}$  converges to  $\bar{R}$  in distribution, where

$$\begin{aligned} \bar{R}_t &= \int_0^t \left[ \int_{\mathbb{R}} -[\gamma^{-1}\rho g(\bar{X}_s, y)u_y(\bar{X}_s, y)\bar{u}_1(s) + \gamma^{-1}\bar{\rho}g(\bar{X}_s, y)u_y(\bar{X}_s, y)\bar{u}_2(s)] \mu_{\bar{X}_s}(dy) \right] ds \\ &= \int_0^t \left[ \int_{\mathbb{R}} [\gamma^{-1}|g(\bar{X}_s, y)u_y(\bar{X}_s, y)|^2] \mu_{\bar{X}_s}(dy) \right] \bar{Q}^{-1}(\bar{X}_s)\dot{\psi}_s ds \\ &= \int_0^t \bar{Q}(\bar{X}_s)\bar{Q}^{-1}(\bar{X}_s)\dot{\psi}_s ds = \int_0^t \dot{\psi}_s ds = \psi_t, \end{aligned}$$

with probability one. If  $\bar{Q}(x)$  becomes zero at a countable number of points, we define  $\bar{Q}^{-1}(x) := 1/\bar{Q}(x)$  if  $\bar{Q}(x) \neq 0$  and 0 otherwise (recall Assumption 4.4). At the same time, the cost satisfies

$$\lim_{\varepsilon \downarrow 0} \mathbb{E} \left( \frac{1}{2} \int_0^T |\bar{u}_s^\varepsilon|^2 ds - \frac{1}{2} \int_0^T \int_{\mathbb{R}} |\bar{u}(s, \bar{X}_s, y)|^2 \mu_{\bar{X}_s}(dy) ds \right)^2 = 0.$$

In addition, the relation (45) implies that

$$\mathbb{E} \left( \frac{1}{2} \int_0^T \int_{\mathbb{R}} |\bar{u}(s, \bar{X}_s, y)|^2 \mu_{\bar{X}_s}(dy) ds \right) = \mathbb{E} (I(\bar{R})) = I(\psi).$$

Then, we can finally write

$$\begin{aligned} \limsup_{\varepsilon \downarrow 0} -\frac{\log \mathbb{E} [\exp \{-h^2(\varepsilon)G(R^\varepsilon)\}]}{h^2(\varepsilon)} &= \limsup_{\varepsilon \downarrow 0} \inf_{u \in \mathcal{A}} \mathbb{E} \left[ \frac{1}{2} \int_0^T |u(t)|^2 dt + G(R^{\varepsilon, u}) \right] \\ &\leq \limsup_{\varepsilon \downarrow 0} \mathbb{E} \left[ \frac{1}{2} \int_0^T |\bar{u}^\varepsilon(t)|^2 dt + G(R^{\varepsilon, \bar{u}^\varepsilon}) \right] \\ &= \mathbb{E} \left[ \frac{1}{2} \int_0^T \int_{\mathbb{R}} |\bar{u}(s, \bar{X}_s, y)|^2 \mu_{\bar{X}_s}(dy) ds + G(\bar{R}) \right] \\ &= [I(\psi) + G(\psi)] \leq \inf_{\phi \in \mathcal{C}([0, 1]; \mathbb{R})} [I(\phi) + G(\phi)] + \zeta. \end{aligned}$$

Since  $\zeta > 0$  is arbitrary, the upper bound is proved. The proof also shows that we have the explicit representation given by Theorem 4.5.

## APPENDIX A. REGULARITY RESULTS ON POISSON EQUATIONS

The following theorem collects results from [27] and [28] that are used in this paper.

**Theorem A.1.** *Under Assumption 4.1, set  $F(x, y) = H(x, y) - \bar{H}(x)$ . If there exist  $K, q_F > 0$  such that*

$$|F(x, y)| + \|\partial_x F(x, y)\| + \|\partial_{xx} F(x, y)\| \leq K(1 + |y|^{q_F}),$$

*then there is a unique solution from the class of functions which grow at most polynomially in  $|y|$  to*

$$\mathcal{L}_Y u(x, y) = -F(x, y), \quad \text{with} \quad \int_Y u(x, y) \mu(dy) = 0.$$

*Moreover, the solution satisfies  $u(\cdot, y) \in \mathcal{C}^2$  for every  $y \in \mathbb{R}$ ,  $\partial_{xx} u \in \mathcal{C}$ , and there exists  $K' > 0$  such that*

$$|u(x, y)| + \|\partial_y u(x, y)\| + \|\partial_x u(x, y)\| + \|\partial_{xx} u(x, y)\| \leq K'(1 + |y|)^{q_F},$$

While Theorem A.1 addresses regularity and growth properties of solutions to general Poisson equations, Lemma A.2 specialises the discussion on Poisson equations whose operators correspond to the so-called CEV model—widely used in mathematical finance—and where the right-hand side  $H(x, y) \equiv H(y)$  is only a function of  $y$ . In this case, we have more precise information on the growth of the solution and its derivatives.

**Lemma A.2.** *Consider the Poisson equation (23) with the operator given by (24) and  $H(x, \cdot) = H(\cdot) \in \mathcal{C}^{2,\alpha}$  such that for some  $K > 0$  and  $q_H \geq 1$ ,  $|H(y)| \leq K(1 + |y|^{q_H})$  holds, then there is a unique solution from the class of functions which grow at most polynomially in  $|y|$  to*

$$\mathcal{L}_Y u(y) = H(y) - \bar{H}, \quad \text{with} \quad \int_Y u(y) \mu(dy) = 0.$$

*Moreover, the solution satisfies  $u \in \mathcal{C}^2$ , and there exists a positive constant  $K'$  such that*

$$|u(y)| \leq K'(1 + |y|^{q_H}), \quad \text{and} \quad |\nabla_y u(y)| \leq K'(1 + |y|^{q_H-1}).$$

*If  $q_H = 0$ , then  $|u(y)| \leq K'(1 + \log(1 + |y|))$ .*

*Proof.* If the generator  $\mathcal{L}_Y$  corresponds to the CIR model, i.e., when  $q_g = 1/2$ , we refer the reader to [17, Lemma 3.2]. We prove it for the more general case  $q_g \in (1/2, 1)$  and  $q_H \geq 1$  using a more direct argument. The invariant measure has density given by the speed measure of the diffusion:

$$m(y) = \frac{1}{\xi^2 y^{2q_g}} \exp \left\{ \int_1^y \frac{2\kappa(\theta - z)}{\xi^2 z^{2q_g}} dz \right\},$$

and a quick computation shows that the derivative to the solution of the Poisson PDE satisfies

$$u'(y) = \frac{1}{\xi^2 y^{2q_g} m(y)} \int_0^y (H(z) - \bar{H}) m(z) dz = -\frac{1}{\xi^2 y^{2q_g} m(y)} \int_y^\infty (H(z) - \bar{H}) m(z) dz,$$

where we used the centering condition to obtain the last equality. Now, without loss of generality we assume that  $\xi = \theta = \kappa = 1$ . For some positive constant  $C$  that may change from line to line, we can write

$$\begin{aligned} |u'(y)| &\leq \frac{1}{y^{2q_g} m(y)} \int_y^\infty z^{q_H} m(z) dz \leq C e^{\frac{y^{2-2q_g}}{1-q_g}} \int_y^\infty z^{q_H-2q_g} e^{-\frac{z^{2-2q_g}}{1-q_g}} dz \\ &\leq C e^{R(y)} y^{1+q_H-2q_g} \int_1^\infty g(x) e^{-y^{2(1-q_g)} R(x)} dx, \end{aligned}$$

with  $g(x) := x^{q_H-2q_g}$  and  $R(x) := \frac{x^{2(1-q_g)}}{1-q_g}$ . Since  $1 - q_g > 0$ , the function  $R(\cdot)$  is smooth and strictly increasing on  $[1, \infty]$ . Furthermore, since  $q_H - 2q_g \in (0, 1)$ , the function  $g(\cdot)$  is smooth on  $[1, \infty]$ . As  $y$  becomes large, this integral is exactly of Laplace type, albeit without a saddlepoint for  $R(\cdot)$ , its minimum being attained at the left boundary of the domain. Using [25, Section 3.3], we can write, for large  $y$ ,

$$\int_1^\infty g(u) e^{-y^{2(1-q_g)} R(u)} du \sim \frac{g(1) e^{-y^{2(1-q_g)} R(1)}}{y^{2(1-q_g)} R'(1)} = \frac{1}{2} y^{2(q_g-1)} e^{-y^{2(1-q_g)} R(1)},$$

and therefore, as  $y$  tends to infinity,  $|u'(y)| = \mathcal{O}(y^{q_H-1})$ , and the lemma follows.  $\square$

## APPENDIX B. PRELIMINARY RESULTS ON THE INVOLVED CONTROLLED PROCESSES

In this section we recall some results from [26] and we prove some additional corollaries related to certain bounds involving the controlled processes (33) appearing via the weak convergence control representation.

**Lemma B.1** (Lemma B.1 of [26]). *Under Assumptions 4.1 and 4.2, let  $(X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon})$  be the strong solution to (33). Then the infimum of the representation in (32) can be taken over all controls such that*

$$\int_0^T |u^\varepsilon(s)|^2 ds < N, \text{ almost surely,}$$

where the constant  $N$  does not depend on  $\varepsilon$  or  $\delta$ .

**Lemma B.2** (Lemma B.2 of [26]). *Under Assumptions 4.1 and 4.2, for  $N \in \mathbb{N}$ , let  $u^\varepsilon \in \mathcal{A}$  such that*

$$\sup_{\varepsilon > 0} \int_0^T |u^\varepsilon(s)|^2 ds < N$$

holds almost surely. Then there exist  $\varepsilon_0 > 0$  small enough such that

$$\sup_{\varepsilon \in (0, \varepsilon_0)} \mathbb{E} \left( \int_0^T |Y_s^{\varepsilon, u^\varepsilon}|^2 ds \right) \leq K(N, T),$$

for some finite constant  $K(N, T)$  that may depend on  $(N, T)$ , but not on  $\varepsilon, \delta$ .

Lemma B.3 follows from Lemma B.2, but since its short proof was not included in [26], we include it here.

**Lemma B.3.** *Let Assumptions 4.1 and 4.2 be satisfied. For  $N \in \mathbb{N}$ , let  $u^\varepsilon \in \mathcal{A}$  such that*

$$\sup_{\varepsilon > 0} \int_0^T |u^\varepsilon(s)|^2 ds < N,$$

almost surely. Then, for  $0 \leq q_g < 1$  and for  $\varepsilon > 0$  small enough we have

$$\mathbb{E} \left( \sup_{t \in [0, T]} |Y_t^{\varepsilon, u^\varepsilon}|^2 \right) \leq K(N, T) \left( 1 + h(\varepsilon)^{\frac{2}{1-q_g}} \right),$$

for some finite constant  $K(N, T)$  that does not depend on  $\varepsilon$  and  $\delta$ .

*Proof of Lemma B.3.* By Assumption 4.1, we can write that  $f(x, y) = -\kappa y + \tau(x, y)$ , and we can assume that  $\tau$  is globally Lipschitz in  $y$  with Lipschitz constant  $L_\tau < \kappa$ , where  $\kappa$  is from Assumption 4.1. From presentation purposes and without loss of generality, we only prove the case  $\rho = 0$ , i.e., when the Brownian motions are independent. It is easy to see that

$$Y_t^{\varepsilon, u^\varepsilon} = e^{-\frac{\varepsilon \kappa}{\delta^2} t} y_0 + \frac{\varepsilon}{\delta^2} \int_0^t e^{-\frac{\varepsilon}{\delta^2} \kappa(t-s)} \tau(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) ds + Q_t^\varepsilon + M_t^\varepsilon,$$

where

$$Q_t^\varepsilon := \frac{\sqrt{\varepsilon} h(\varepsilon)}{\delta} \int_0^t e^{-\frac{\varepsilon}{\delta^2} \kappa(t-s)} g(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) u_s^\varepsilon ds \quad \text{and} \quad M_t^\varepsilon := \frac{\sqrt{\varepsilon}}{\delta} \int_0^t e^{-\frac{\varepsilon}{\delta^2} \kappa(t-s)} g(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) dZ_s.$$

Let us also set  $\Lambda_t^\varepsilon := Y_t^{\varepsilon, u^\varepsilon} - Q_t^\varepsilon - M_t^\varepsilon$ , so that

$$d\Lambda_t^\varepsilon = -\frac{\varepsilon}{\delta^2} \kappa \Lambda_t^\varepsilon dt + \frac{\varepsilon}{\delta^2} \tau(X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) dt.$$

Hence, we have

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} |\Lambda_t^\varepsilon|^2 &= d\Lambda_t^\varepsilon \cdot \Lambda_t^\varepsilon = -\frac{\varepsilon}{\delta^2} \kappa |\Lambda_t^\varepsilon|^2 + \frac{\varepsilon}{\delta^2} \tau(X_t^{\varepsilon, u^\varepsilon}, Y_t^{\varepsilon, u^\varepsilon}) \Lambda_t^\varepsilon \\ &\leq -\frac{\varepsilon}{\delta^2} \kappa |\Lambda_t^\varepsilon|^2 + \frac{\varepsilon}{\delta^2} \left( \tau(X_t^{\varepsilon, u^\varepsilon}, \Lambda_t^\varepsilon + Q_t^\varepsilon + M_t^\varepsilon) - \tau(X_t^{\varepsilon, u^\varepsilon}, Q_t^\varepsilon + M_t^\varepsilon) \right) \Lambda_t^\varepsilon + \frac{\varepsilon}{\delta^2} \tau(X_t^{\varepsilon, u^\varepsilon}, Q_t^\varepsilon + M_t^\varepsilon) \Lambda_t^\varepsilon \\ &\leq -\frac{\varepsilon}{2\delta^2} (\kappa - L_\tau) |\Lambda_t^\varepsilon|^2 + K \frac{\varepsilon}{\delta^2} (1 + |Q_t^\varepsilon|^2 + |M_t^\varepsilon|^2), \end{aligned}$$

for some unimportant positive constant  $K$ . In the last line we used the Lipschitz and growth properties of  $\tau$  together with Young's inequality. Then we obtain that for some finite constant  $K > 0$ ,

$$|\Lambda_t^\varepsilon|^2 \leq e^{-\frac{\varepsilon}{\delta^2}(\kappa-L\tau)t} |y_0|^2 + K \frac{\varepsilon}{\delta^2} \int_0^t e^{-\frac{\varepsilon}{\delta^2}(\kappa-L\tau)(t-s)} (1 + |Q_s^\varepsilon|^2 + |M_s^\varepsilon|^2) ds.$$

We now bound each term separately. We have for some constant  $K$  that may change from line to line

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in [0, T]} |Q_t^\varepsilon|^2 \right) &\leq \frac{\varepsilon h^2(\varepsilon)}{\delta^2} \left( \int_0^T e^{-\frac{\varepsilon}{\delta^2} \kappa(T-s)} g(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) u_s^\varepsilon ds \right)^2 \\ &\leq \frac{\varepsilon h^2(\varepsilon)}{\delta^2} \int_0^T e^{-2\frac{\varepsilon}{\delta^2} \kappa(T-s)} \left| g(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) \right|^2 ds \int_0^T |u_s^\varepsilon|^2 ds \\ &\leq \frac{\varepsilon h^2(\varepsilon)}{\delta^2} N K \int_0^T e^{-2\frac{\varepsilon}{\delta^2} \kappa(T-s)} \left( 1 + \left| Y_s^{\varepsilon, u^\varepsilon} \right|^{2q_g} \right) ds \\ &\leq \frac{\varepsilon h^2(\varepsilon)}{\delta^2} N K \int_0^T e^{-2\frac{\varepsilon}{\delta^2} \kappa(T-s)} ds \left( 1 + \mathbb{E} \sup_{t \in [0, T]} \left| Y_t^{\varepsilon, u^\varepsilon} \right|^{2q_g} \right) \\ &\leq K(N, T, \kappa) h^2(\varepsilon) \left( 1 + \mathbb{E} \sup_{t \in [0, T]} \left| Y_t^{\varepsilon, u^\varepsilon} \right|^{2q_g} \right). \end{aligned}$$

Using the Burkholder-Davis-Gundy inequality and proceeding as above, we get for the stochastic integral

$$\mathbb{E} \left( \sup_{t \in [0, T]} |M_t^\varepsilon|^2 \right) \leq K \frac{\varepsilon}{\delta^2} \int_0^T e^{-2\frac{\varepsilon}{\delta^2} \kappa(T-s)} \left( 1 + \left| Y_s^{\varepsilon, u^\varepsilon} \right|^{2q_g} \right) ds \leq K(T, \kappa) \left\{ 1 + \mathbb{E} \left( \sup_{t \in [0, T]} \left| Y_t^{\varepsilon, u^\varepsilon} \right|^{2q_g} \right) \right\}.$$

Consequently, we obtain that for some appropriate constant  $K < \infty$ ,

$$\mathbb{E} \left( \sup_{t \in [0, T]} |\Lambda_t^\varepsilon|^2 \right) \leq K \left\{ 1 + \mathbb{E} \left( \sup_{t \in [0, T]} \left| Y_t^{\varepsilon, u^\varepsilon} \right|^{2q_g} \right) + h^2(\varepsilon) \mathbb{E} \left( \sup_{t \in [0, T]} \left| Y_t^{\varepsilon, u^\varepsilon} \right|^{2q_g} \right) \right\}$$

Hence, recalling the original definition of  $\Lambda_t^\varepsilon = Y_t^{\varepsilon, u^\varepsilon} - Q_t^\varepsilon - M_t^\varepsilon$ , we have for  $\varepsilon > 0$  small enough such that  $h(\varepsilon) > 1$  and for some constant  $K < \infty$  that changes from line to line,

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in [0, T]} |Y_t^{\varepsilon, u^\varepsilon}|^2 \right) &\leq K \left( 1 + \mathbb{E} \left( \sup_{t \in [0, T]} \left| Y_t^{\varepsilon, u^\varepsilon} \right|^{2q_g} \right) + h^2(\varepsilon) \mathbb{E} \left( \sup_{t \in [0, T]} \left| Y_t^{\varepsilon, u^\varepsilon} \right|^{2q_g} \right) \right) \\ &\leq K \left( 1 + h^2(\varepsilon) \mathbb{E} \left( \sup_{t \in [0, T]} \left| Y_t^{\varepsilon, u^\varepsilon} \right|^{2q_g} \right) \right) \\ &\leq K \left( 1 + h(\varepsilon)^{\frac{2}{1-q_g}} \right) + \frac{1}{2} \mathbb{E} \left( \sup_{t \in [0, T]} \left| Y_t^{\varepsilon, u^\varepsilon} \right|^2 \right), \end{aligned}$$

where in the last step we used Young's generalised inequality. Hence, by rearranging terms we obtain  $\mathbb{E} \left( \sup_{t \in [0, T]} |Y_t^{\varepsilon, u^\varepsilon}|^2 \right) \leq K \left( 1 + h(\varepsilon)^{\frac{2}{1-q_g}} \right)$ , completing the proof of the lemma.  $\square$

**Lemma B.4** (Lemma B.3 in [26]). *Let Assumption 4.1 hold,  $N \in \mathbb{N}$ , and  $u^\varepsilon = (u_1^\varepsilon, u_2^\varepsilon) \in \mathcal{A}$  such that*

$$\sup_{\varepsilon > 0} \int_0^T |u^\varepsilon(s)|^2 ds < N$$

*holds almost surely. Let  $A(x, y)$  and  $B(x, y)$  be given functions and  $K, \theta \in (0, 1)$  such that*

$$|A(x, y)| \leq K(1 + |y|^\theta) \quad \text{and} \quad |B(x, y)| \leq K(1 + |y|^{2\theta}).$$

*Then for  $\alpha \in \{1, 2\}$ ,*

*(i) for any  $p \in (1, 1/\theta]$ , there exists  $C > 0$  such that*

$$\mathbb{E} \left( \sup_{t \in [0, T]} \left| \int_0^T A(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) u_\alpha^\varepsilon(s) ds \right|^{2p} + \sup_{t \in [0, T]} \left| \int_0^T B(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) ds \right|^p \right) \leq C;$$

(ii) for any  $p \in (1, 1/\theta]$ , there exists  $C > 0$  such that for fixed  $e > 0$  and for all  $0 \leq t_1 < t_1 + e \leq T$ ,

$$\mathbb{E} \left( \sup_{\substack{0 \leq t_1 < t_2 \leq T \\ |t_2 - t_1| < e}} \left| \int_{t_1}^{t_2} A(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) u_\alpha^\varepsilon(s) ds \right|^{2p} + \sup_{\substack{0 \leq t_1 < t_2 \leq T \\ |t_2 - t_1| < e}} \left| \int_{t_1}^{t_2} B(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) ds \right|^p \right) \leq C|e|^{r/\theta-1};$$

(iii) for all  $\zeta > 0$ ,

$$\lim_{e \downarrow 0} \limsup_{\varepsilon \downarrow 0} \mathbb{P} \left[ \sup_{\substack{0 \leq t_1 < t_2 \leq T \\ |t_2 - t_1| < e}} \left| \int_{t_1}^{t_2} A(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) u_\alpha^\varepsilon(s) ds \right| > \zeta \right] = 0$$

and

$$\lim_{e \downarrow 0} \limsup_{\varepsilon \downarrow 0} \mathbb{P} \left[ \sup_{\substack{0 \leq t_1 < t_2 \leq T \\ |t_2 - t_1| < e}} \left| \int_{t_1}^{t_2} B(X_s^{\varepsilon, u^\varepsilon}, Y_s^{\varepsilon, u^\varepsilon}) ds \right| > \zeta \right] = 0.$$

## REFERENCES

- [1] D. Baier and M. I. Freidlin. Theorems on large deviations and stability for random perturbations. *Dokl. Akad. Nauk SSSR* **235**: 253-256, 1977.
- [2] P. Billingsley. Convergence of Probability Measures, Second Edition, Wiley, New York, 1968.
- [3] M. Boué and P. Dupuis. A variational representation for certain functionals of Brownian motion. *Annals of Probability*, **26**(4): 1641-1659, 1998.
- [4] M. Chesney, M. Jeanblanc and M. Yor. Mathematical methods for financial markets. Springer-Verlag London, 2009.
- [5] A. Chiarini and M. Fischer. On large deviations for small noise Itô processes. *Advances App. Prob.*, **46**: 1126-1147, 2014.
- [6] A. Dembo and O. Zeitouni. Large deviations techniques and applications. Springer-Verlag New-York, 1998.
- [7] J.D. Deuschel, P.K. Friz, A. Jacquier and S. Violante. Marginal density expansions for diffusions and stochastic volatility, Part I: Theoretical foundations. *Communications on Pure and Applied Mathematics*, **67**(1): 40-82, 2014.
- [8] J.D. Deuschel, P.K. Friz, A. Jacquier and S. Violante. Marginal density expansions for diffusions and stochastic volatility, Part II: Applications. *Communications on Pure and Applied Mathematics*, **67**(2): 321-350, 2014.
- [9] P. Dupuis and R.S. Ellis. A weak convergence approach to the theory of large deviations. John Wiley & Sons, New York, 1997.
- [10] P. Dupuis and K. Spiliopoulos. Large deviations for multiscale problems via weak convergence methods. *Stochastic Processes and their Applications*, **122**: 1947-1987, 2012.
- [11] J. Feng, J.-P. Fouque and R. Kumar. Small-time asymptotics for fast mean-reverting stochastic volatility models. *Annals of Applied Probability*, **22**(4): 1541-1575, 2012.
- [12] M. Forde and A. Jacquier. The large-maturity smile for the Heston model *Finance and Stochastics*, **15**(4): 755-780, 2011.
- [13] M. Freidlin. The averaging principle and theorems on large deviations. *Russian Math. Surveys*, **33**(5): 117-176, 1978.
- [14] P. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier and J. Teichmann. Large deviations and asymptotic methods in finance. Springer Proceedings in Mathematics and Statistics, **110**, 2015.
- [15] P. Friz, S. Gerhold and A. Pinter. Option pricing in the moderate deviations regime. Forthcoming in *Mathematical Finance*.
- [16] P. Friz, S. Gerhold and M. Yor. How to make Dupire's local volatility work with jumps. *Quant. Fin.*, **14**(8): 1327-1331, 2014.
- [17] J.-P. Fouque, G. Papanicolaou, R. Sircar and K. Solna. Multiscale stochastic volatility for equity, interest rate, and credit derivatives. Cambridge University Press, Cambridge, 2011.
- [18] A. Guillin. Averaging principle of SDE with small diffusion: moderate deviations. *Annals of Probability*, **31**(1): 413-443, 2003.
- [19] A. Guillin and R. Liptser. Examples of moderate deviation principle for diffusion processes. *Discrete and Continuous Dynamical Systems (B)*, **6**(4): 803-828, 2006.
- [20] A. Guillin and R. Liptser. MDP for integral functionals of fast and slow processes with averaging. *Stochastic Processes and their applications*, **115**: 1187-1207, 2005.
- [21] A. Gulisashvili and E. Stein. Asymptotic behavior of the stock price distribution density and implied volatility in stochastic volatility models. *Applied Mathematics and Optimization*, **61**(3): 287-315.
- [22] A. Jacquier, M. Keller-Ressel and A. Mijatović. Large deviations and stochastic volatility with jumps. *Stochastics*, **85**(2): 321-345, 2013.
- [23] I. Karatzas and S. Shreve. Brownian Motion and Stochastic Calculus. Springer, 1991.
- [24] R. Lipster and V. Spokoyny. Moderate deviations type evaluation for integral functional of diffusion processes. *Electronic Journal of Probability*, **4**: 1-25, 1999.
- [25] P.D. Miller. Applied asymptotic analysis. Graduate Studies in Mathematics, **75**, AMS, 2006.
- [26] M. Morse and K. Spiliopoulos. Moderate deviations principle for systems of slow-fast diffusions. *Asymptotic Analysis*, **105**: 97-135, 2017.
- [27] E. Pardoux and A.Yu. Veretennikov. On the Poisson equation and diffusion approximation I. *Annals of Probability*, **29**(3): 1061-1085, 2001.

- [28] E. Pardoux and A.Yu. Veretennikov. On Poisson equation and diffusion approximation II. *Annals of Probability*, 31(3): 1166-1192, 2003.
- [29] S. Robertson. Sample path large deviations and optimal importance sampling for stochastic volatility models. *Stochastic Processes and their Applications*, 120(1), 66-83, 2010.
- [30] K. Spiliopoulos. Large deviations and importance sampling for systems of slow-fast motion. *Applied Mathematics and Optimization*, 67: 123-161, 2013.
- [31] E. Stein and J. Stein. Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies*, 4(4): 727-752, 1991.

DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE LONDON AND BARUCH COLLEGE, CUNY

*E-mail address:* `a.jacquier@imperial.ac.uk`

DEPARTMENT OF MATHEMATICS AND STATISTICS, BOSTON UNIVERSITY

*E-mail address:* `kspiliop@math.bu.edu`