

LARGE AND MODERATE DEVIATIONS FOR IMPORTANCE SAMPLING IN THE HESTON MODEL

MARC GEHA, ANTOINE JACQUIER, AND ŽAN ŽURIČ

ABSTRACT. We provide a detailed importance sampling analysis for variance reduction in stochastic volatility models. The optimal change of measure is obtained using a variety of results from large and moderate deviations: small-time, large-time, small-noise. Specialising the results to the Heston model, we derive many closed-form solutions, making the whole approach easy to implement. We support our theoretical results with a detailed numerical analysis of the variance reduction gains.

CONTENTS

| | |
|--|----|
| 1. Introduction and general overview | 2 |
| 1.1. Introduction | 2 |
| 1.2. Overview of the importance sampling methodology | 2 |
| 1.3. Choosing an approximated random variable G_ε | 3 |
| 1.4. General approach | 3 |
| 2. Importance sampling via large deviations | 5 |
| 2.1. Small-noise LDP | 5 |
| 2.2. Small-time LDP | 9 |
| 2.3. Large-time LDP | 10 |
| 3. Importance sampling via moderate deviations | 10 |
| 3.1. Small-noise moderate deviations | 11 |
| 3.2. Small-time moderate deviations | 14 |
| 3.3. Large-time moderate deviations | 15 |
| 4. Examples | 17 |
| 4.1. Black-Scholes with deterministic volatility | 17 |
| 4.2. Ornstein-Uhlenbeck | 19 |
| 4.3. Rough volatility | 19 |
| 4.4. The Heston model | 19 |
| 5. Numerical results | 21 |
| 5.1. Numerical results for Black-Scholes | 22 |
| 5.2. Numerical results for the Heston model | 23 |
| References | 27 |
| Appendix A. Technical proofs | 28 |
| A.1. Proof of Lemma 3.1.1 | 28 |
| A.2. Proof of Lemma 3.1.2 | 28 |
| A.3. Proof of Lemma 3.1.3 | 29 |
| A.4. Proof of Theorem 3.1.1 | 30 |
| Appendix B. Variations around Varadhan's Lemma and Exponential Equivalence | 32 |

Date: November 17, 2023.

2020 Mathematics Subject Classification. 60F10, 65C05, 91G20.

Key words and phrases. Heston, volatility, importance sampling, large deviations, moderate deviations.

AJ acknowledges financial support from the EPSRC/T032146 grant. ŽŽ is supported by the EP-SRC/S023925 CDT in Mathematics of Random Systems: Analysis, Modelling and Simulation.

Conflict of Interests: The authors declare that there is no conflict of interests regarding the publication of this article.

1. INTRODUCTION AND GENERAL OVERVIEW

1.1. Introduction. Monte Carlo simulation is the standard (if not the only) technique for most numerical problems in stochastic modelling. It has a long history and has been successfully applied in many fields, such as biology [32], statistical Physics [4], Finance [17] among others. The default order of magnitude for the variance of the estimator is $\mathcal{O}(N^{-1/2})$ with N the number of sample paths. It has long been recognised though that several tricks achieve lower variance with equivalent (hopefully zero) bias; among those antithetic variables and importance sampling have become ubiquitous. We focus on the latter, for which large and moderate deviations (LDP and MDP) provide closed-form formulae, making their applications pain-free and without additional computer costs.

The first attempt to reduce the variance of a Monte Carlo estimator based on asymptotics originated, rather heuristically, in [40]. This was then made rigorous by Glasserman and Wang [18], who also highlighted pitfalls of the method and by Dupuis and Wang [12], who provided clear explanations on the trade-off between asymptotic approximations and the restrictions they entail on the induced change of measure. Guasoni and Robertson [20] put this into practice for out-the-money path-dependent options in the Black-Scholes model, and Robertson [36] developed a thorough analysis for the Heston model using sample path large deviations. This is our starting point, and the goal of our current enterprise is to analyse different asymptotic regimes (small-time, large-time, small-noise), both in the large deviations and in the moderate deviations regimes, in the Heston model and to show how these yield closed-form formulae for an optimal change of measure for importance sampling purposes.

We propose, in particular, a specific form of adaptive drift, allowing for fast computation and increase in variance reduction. For geometric Asian Call options in the Heston model, MDP-based estimators with deterministic changes of drift turn out to be no better than those computed with deterministic volatility approximation in the LDP approach. However, MDP-based estimators with adaptive changes of drift perform much better than their LDP counterparts with deterministic volatility approximation, and in fact show a performance very close to the LDP-based estimators in Heston. These adaptive MDP-based estimators therefore provide an efficient alternative in models where LDP is difficult to compute.

Setting and notations Throughout this paper we work on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ with a finite time horizon $T > 0$, where $\Omega = \mathcal{C}([0, T] \rightarrow \mathbb{R}^2)$ is the space of all continuous functions, \mathcal{F} is the Borel- σ -algebra on Ω and $\mathbb{F} := (\mathcal{F}_t)_{t \in [0, T]}$ is the natural filtration of a given two-dimensional standard Brownian motion $\mathbf{W} := (W, W^\perp)$. For a pair of (possibly deterministic) process (X, Y) , with X predictable and Y a semi-martingale, we write the stochastic integral $X \circ Y := \int_0^\cdot X_s dY_s$ and $X \circ_t W := (X \circ W)_t$ for any $t \in [0, T]$. We denote any d -dimensional path by $\mathbf{h} := (h_1, \dots, h_d)$ for $d \in \mathbb{N}$, and for such a path, $\|\mathbf{h}\|_T^2 := \int_0^T (|h_1(t)|^2 + \dots + |h_d(t)|^2) dt$. The Cameron-Martin space \mathbb{H}_T^0 of Brownian motion is isomorphic to the space of absolutely continuous functions $\mathcal{AC}([0, T])$ starting at zero and with square integrable first derivatives. We define a similar space $\mathbb{H}_T^x := \{\varphi \in \mathcal{C}([0, T] \rightarrow \mathbb{R}^d) : \varphi_t = x + \int_0^t \dot{\varphi}_s ds, \dot{\varphi} \in L^2([0, T]; \mathbb{R}^d)\}$ for processes starting at $x \in \mathbb{R}^d$ and a subspace $\mathbb{H}_T^{x,+} \subset \mathbb{H}_T^x$ where functions map to $(\mathbb{R}^+)^d$ instead of \mathbb{R}^d . Whenever a variable has an obvious time-dependence, we drop the explicit reference in the notation. We also write $\mathcal{C}_T := \mathcal{C}([0, T] \rightarrow \mathbb{R})$ to simplify statements. Finally $\{X^\varepsilon\} \sim \text{LDP}(\mathbf{I}^X, \mathcal{C}_T)$ means that the sequence $\{X^\varepsilon\}$ satisfies a large deviations principle as ε tends to zero on \mathcal{C}_T with good rate function \mathbf{I}^X . For a given function f , we denote by $\mathcal{D}(f)$ its effective domain. We finally write $\mathbb{R}^+ := [0, \infty)$ and \mathcal{BV} for the space of paths with bounded variation.

1.2. Overview of the importance sampling methodology. We consider a given risk-neutral probability measure \mathbb{P} , so that the fundamental theorem of asset pricing implies that the price of an option with attainable payoff $G \in L^2(\Omega; \mathbb{R})$ is equal to $\mathbb{E}^\mathbb{P}[G]$. While, strictly speaking, we do not need $L^2(\Omega; \mathbb{R})$ for pricing purposes, we require it to estimate the variance of payoff estimators. Monte-Carlo estimators rely on the (strong) law of large number, whereby for iid samples $\{G_i\}_{1 \leq i \leq n}$ from $\mathbb{P} \circ G^{-1}$, the empirical mean $\hat{G}_n := \frac{1}{n} \sum_{i=1}^n G_i$ converges to the true expectation \mathbb{P} -almost surely:

$$\lim_{n \uparrow \infty} \hat{G}_n = \mathbb{E}^\mathbb{P}[G].$$

Importance sampling is a method to reduce the variance of the estimator \widehat{G}_n , yielding a new law \mathbb{Q} such that $\mathbb{E}^{\mathbb{Q}}[G] = \mathbb{E}^{\mathbb{P}}[G]$ and $\mathbb{V}^{\mathbb{Q}}[G] < \mathbb{V}^{\mathbb{P}}[G]$ (and of course both the equality and inequality remain true with G replaced by \widehat{G}_n). Let for example $Z := \frac{d\mathbb{Q}}{d\mathbb{P}}$ denote the Radon-Nikodym derivative of the change of measure, so that $\mathbb{E}^{\mathbb{P}}[G] = \mathbb{E}^{\mathbb{Q}}[GZ^{-1}]$. The variance of the Monte-Carlo estimator based on iid samples of $\widehat{G}_n Z^{-1}$ under \mathbb{Q} is then

$$\mathbb{V}^{\mathbb{Q}}[\widehat{G}_n Z^{-1}] = \mathbb{E}^{\mathbb{Q}}[\widehat{G}_n^2 Z^{-2}] - \mathbb{E}^{\mathbb{Q}}[\widehat{G}_n Z^{-1}]^2 = \mathbb{E}^{\mathbb{P}}[\widehat{G}_n^2 Z^{-1}] - \mathbb{E}^{\mathbb{P}}[\widehat{G}_n]^2.$$

If Z is chosen such that $\mathbb{E}^{\mathbb{P}}[\widehat{G}_n^2 Z^{-1}] < \mathbb{E}^{\mathbb{P}}[\widehat{G}_n^2]$, the variance is thus reduced. Finding such Z however is usually hard, and we shall instead consider the approximation

$$(1.1) \quad \mathbb{E}^{\mathbb{P}}[\widehat{G}_n^2 Z^{-1}] \approx \varepsilon \log \mathbb{E}^{\mathbb{P}} \left[\exp \left\{ \frac{1}{\varepsilon} \log(G_\varepsilon^2 Z_\varepsilon^{-1}) \right\} \right],$$

for small $\varepsilon > 0$, for two random variables G_ε and Z_ε whose choices will be discussed later. The computation of this expression is then further simplified by the use of the Varadhan's lemma (Theorem B.0.1), which casts the problem into a deterministic optimisation over the appropriate Cameron-Martin space.

1.3. Choosing an approximated random variable G_ε . Consider an attainable payoff $G(X)$, where X is a unique strong solution to the stochastic differential equation

$$(1.2) \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0,$$

where $b, \sigma : \mathbb{R} \rightarrow \mathbb{R}$ are sufficiently well-behaved deterministic functions and W is a standard Brownian motion. The approximation of G is then defined as $G_\varepsilon := G(X^\varepsilon)$, where the following are possible approximations of X :

Definition 1.3.1. *Let X be a unique strong solution to (1.2). The process X^ε is called*

i) Small-noise approximation if

$$(1.3) \quad dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dW_t, \quad X_0^\varepsilon = x_0.$$

ii) Small-time approximation if

$$(1.4) \quad dX_t^\varepsilon = \varepsilon b(X_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dW_t, \quad X_0^\varepsilon = x_0$$

iii) Large-time approximation if

$$(1.5) \quad dX_t^\varepsilon = \frac{1}{\varepsilon}b(X_t^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}}\sigma(X_t^\varepsilon)dW_t, \quad X_0^\varepsilon = x_0.$$

The terminology here is straightforward since (1.4) follows from (1.2) via the mapping $t \mapsto \varepsilon t$ and (1.5) follows from (1.2) via the mapping $t \mapsto t/\varepsilon$. The small-noise (1.3) comes from the early works on random perturbations of deterministic systems by Varadhan [41] and Freidlin-Wentzell [13].

1.4. General approach. We consider an asset price $S := \{S_t\}_{t \in [0, T]}$ and the corresponding log-price process $X := \log(S) := \{X_t\}_{t \in [0, T]}$, with dynamics

$$(1.6) \quad \begin{aligned} dX_t &= -\frac{1}{2}V_t dt + \sqrt{V_t}dB_t, & X_0 &= 0, \\ dV_t &= f(V_t)dt + g(V_t)dW_t, & V_0 &= v_0 > 0, \end{aligned}$$

where $\mathbf{W} = (W, W^\perp)$ is a standard two-dimensional Brownian motion and $B := \rho W + \bar{\rho} W^\perp$ with correlation coefficient $\rho \in (-1, 1)$ and $\bar{\rho} := \sqrt{1 - \rho^2}$. The drift and diffusion coefficients of the volatility process satisfy $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ and $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ and Assumption 1.4.1 if not stated otherwise (e.g. in the case of large-time approximation in Section 3.3 additional assumptions are required for ergodicity purposes).

Assumption 1.4.1.

- (i) *The function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ is globally Lipschitz continuous;*
- (ii) *The function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is increasing, strictly positive outside origin. Furthermore, there exist $K > 0$ and $p \geq \frac{1}{2}$ such that, for all $x, y \in \mathbb{R}^+$,*

$$|g(x) - g(y)| \leq K|x - y|^p.$$

Remark 1.4.1. Condition (ii) in fact implies p -polynomial growth condition for $p \geq \frac{1}{2}$, i.e., $|g(x)| \leq C(1 + |x|^p)$ for all $x \in \mathbb{R}^+$. Indeed, let $y = 0$ and $x \in \mathbb{R}^+$, then $|g(x) - g(0)| \leq K|x|^p$ and by the triangle inequality $|g(x)| \leq |g(0)| + K|x|^p$.

Under this assumption, Yamada-Watanabe's theorem [42, Theorem 1] ensures the existence of a unique strong solution to (1.6). Consider now the continuous map $G \in \mathcal{C}_T$, yielding the option price $\mathbb{E}[G(X)]$. Finding the optimal Radon-Nikodym derivative encoding the change of measure from \mathbb{P} to \mathbb{Q} is hard in general and we instead consider the particular class of change of measure

$$(1.7) \quad Z_T := \frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_T} = \exp \left\{ -\frac{1}{2} \|\dot{\mathbf{h}}\|_T^2 + \dot{\mathbf{h}} \circ_t \mathbf{W}^\top \right\},$$

which is well defined and satisfies $\mathbb{E}^\mathbb{P}[Z_T] = 1$ whenever $\mathbf{h} \in \mathbb{H}_T^0$. Now let $F := \log |G|$, so that $\mathbb{E}^\mathbb{P}[Z_T^{-1}G(X)^2] = \mathbb{E}^\mathbb{P}[e^{2F(X) - \log(Z_T)}]$; the approximation (1.1) then yields for $\mathbf{W}^\varepsilon := \sqrt{\varepsilon} \mathbf{W}$ the estimate

$$(1.8) \quad L(\mathbf{h}) := \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E}^\mathbb{P} \left[\exp \left\{ \frac{1}{\varepsilon} \left\{ 2F(X^\varepsilon) - \left(-\frac{1}{2} \|\dot{\mathbf{h}}\|^2 + \dot{\mathbf{h}} \circ_T \mathbf{W}^{\varepsilon\top} \right) \right\} \right\} \right]$$

to compute, for some proxies X^ε .

Definition 1.4.1. A path $\mathbf{h} \in \mathbb{H}_T^0$ minimising $L(\mathbf{h})$ is called an asymptotically optimal change of drift.

From this point onward, several approaches exist in the literature. In [11, 23, 10] fully adaptive schemes are considered, where \mathbf{h} is function of (t, X_t, V_t) . These schemes effectively reduce variance, but are expensive to compute. For that reason, we look at the case where \mathbf{h} is an absolutely continuous function with derivatives in $L^2([0, T]; \mathbb{R})$. The main advantage is the fast computation in comparison to the fully adaptive schemes. Conserving this advantage, we also look at paths \mathbf{h} of the form $\int_0^\cdot \dot{\mathbf{h}}(t) \sqrt{V_t} dt$ (yielding a stochastic change of measure) for which computations are usually just as fast and variance reduction is higher. In the case where \mathbf{h} is a deterministic function (the approach is similar in other settings), the main methodology we shall develop below then goes through the following steps:

- i) We choose appropriate approximations X^ε as in Section 1.3;
- ii) We prove an LDP (MDP) with good rate function I^X for $\{X^\varepsilon\}_{\varepsilon > 0}$;
- iii) We show that Varadhan's Lemma applies, so that the function $L(\mathbf{h})$ in (1.8) reads

$$(1.9) \quad L(\mathbf{h}) = \sup_{X \in \mathcal{D}(I^X)} \mathfrak{L}(X; \mathbf{h}),$$

where

$$(1.10) \quad \mathfrak{L}(X; \mathbf{h}) := 2F(X) - \left(-\frac{1}{2} \|\dot{\mathbf{h}}\|^2 + \dot{\mathbf{h}} \circ_T \mathbf{W}^{\varepsilon\top} \right) - I^X(X).$$

- iv) We consider the dual problem of (1.9) in the sense of Definition 1.4.2. For more details see the remark below.

Definition 1.4.2. The primal problem is defined as

$$(1.11) \quad \inf_{\mathbf{h} \in \mathbb{H}_0^T} \sup_{X \in \mathcal{D}(I^X)} \mathfrak{L}(X; \mathbf{h}),$$

while the dual consists in

$$(1.12) \quad \sup_{X \in \mathcal{D}(I^X)} \inf_{\mathbf{h} \in \mathbb{H}_0^T} \mathfrak{L}(X; \mathbf{h}).$$

Remark 1.4.2. In many cases the primal optimisation problem (1.11) may be difficult to solve analytically, so we deal with the dual problem which turns out to be much simpler. With further assumptions, it may be possible to prove strong duality, however this is outside the scope of this paper. Note in particular that the functional \mathfrak{L} does not admit any obvious convexity property, making classical duality results unusable (see for example [37, Section 4]). Recent results about duality and optimality conditions in stochastic optimisation [3, 35] could provide some hints though.

Remark 1.4.3. *Small-time approximations may induce important loss of information because the drift term in H^ε (the quadratic part in \dot{h}) can be negligible and can thus lead to a trivial dual problem. In the Black-Scholes setting (Section 4.1), a small-time approximation leads to the following problem: let $F : \mathcal{C}_T \rightarrow \mathbb{R}^+$ be a smooth enough function so that Varadhan's Lemma (Theorem B.0.1) holds, then the small-noise approximation problem reads*

$$\mathfrak{L}(x; h) = 2F(x) - \int_0^T \dot{h}_t \dot{x}_t dt - \frac{1}{2} \int_0^T \dot{x}_t^2 dt + \frac{1}{2} \int_0^T \dot{h}_t^2 dt.$$

However, in the small-time case we actually have

$$\mathfrak{L}(x; h) = 2F(x) - \int_0^T \dot{h}_t \dot{x}_t dt - \frac{1}{2} \int_0^T \dot{x}_t^2 dt.$$

In this small-time setting, the dual problem (1.12) then reads

$$\begin{aligned} \sup_{x \in \mathcal{D}(\mathbf{I}^X)} \inf_{h \in \mathbb{H}_T^0} \mathfrak{L}(X; h) &= \sup_{x \in \mathcal{D}(\mathbf{I}^X)} \inf_{h \in \mathbb{H}_T^0} \left\{ 2F(x) - \int_0^T \dot{h}_t \dot{x}_t dt - \frac{1}{2} \int_0^T \dot{x}_t^2 dt \right\} \\ &= \sup_{x \in \mathcal{D}(\mathbf{I}^X)} \left\{ 2F(x) - \frac{1}{2} \int_0^T \dot{x}_t^2 dt - \sup_{h \in \mathbb{H}_T^0} \int_0^T \dot{h}_t \dot{x}_t dt \right\}. \end{aligned}$$

Clearly the path h can be multiplied by any arbitrarily large positive constant to increase the inner supremum, and therefore the optimisation does not admit a maximiser. In these cases, we thus do not consider the small-time approximation.

Remark 1.4.4. *Another drawback of the outlined method exists. In most cases, solving the dual problem in step iv) involves numerical optimisation, which in turn means that the Novikov conditions of the resulting optimal change of drift h cannot be rigorously verified. However, practically we see notable reduction in variance, implying that the obtained Radon-Nikodym derivative induces a change of measure and is indeed a martingale.*

The paper's structure is as follows. In Section 2 we look at stochastic volatility models satisfying Assumption 1.4.1 and derive explicit solutions for large deviations approximations for *path-dependent* payoffs of the form $F(\int_0^T \vartheta_t dX_t)$ for general deterministic paths $\vartheta \in \mathcal{C}_T$. This includes *state-dependent* payoffs of European type, i.e., functions of X_T (for the choice of $\vartheta = 1$) and of Asian type $\frac{1}{T} \int_0^T X_t dt$ (for $\vartheta_t = \frac{T-t}{T}$). In contrast to [36] we consider a more general approach using Garcia's theorem [14], which includes small-time approximations and studies stochastic changes of measure. Later in Section 3 we study moderate deviations, where we derive small-noise, small-time and large-time MDPs, whose advantages, compared to LDP, are simpler forms of rate functions. In Section 4, we provide examples of models used in mathematical finance, which satisfy our setup, with a particular emphasis on the Heston model. Finally, in Section 5 we highlight numerical results for the Heston model and compare variance reduction results for different approximation types. Some of the technical proofs are relegated to the Appendix A.

2. IMPORTANCE SAMPLING VIA LARGE DEVIATIONS

2.1. Small-noise LDP. We start with the small-noise approximation of (1.6):

$$(2.1) \quad \begin{cases} dX_t^\varepsilon &= -\frac{1}{2} V_t^\varepsilon dt + \sqrt{\varepsilon} \sqrt{V_t^\varepsilon} dB_t, & X_0^\varepsilon = 0, \\ dV_t^\varepsilon &= f(V_t^\varepsilon) dt + \sqrt{\varepsilon} g(V_t^\varepsilon) dW_t, & V_0^\varepsilon = v_0. \end{cases}$$

2.1.1. Large deviations. In the spirit of [13], we provide an LDP in $\mathcal{C}([0, T] \rightarrow \mathbb{R}^2)$ for $\{X^\varepsilon, V^\varepsilon\}$; usual assumptions involve non-degenerate and locally-Lipschitz diffusion though, which clearly fails for square-root type stochastic volatility models. Hence, to derive LDP for the (log-)price process, it is first necessary to derive LDP for the pair $\{V^\varepsilon, Y^\varepsilon\}$ given by

$$\begin{aligned} dY_t^\varepsilon &= \sqrt{V_t^\varepsilon} dW_t^{\varepsilon, \perp}, \\ dV_t^\varepsilon &= f(V_t^\varepsilon) dt + g(V_t^\varepsilon) dW_t^\varepsilon, \end{aligned}$$

assuming an LDP for V . To be more exact we follow [36] and assume the following:

Assumption 2.1.1. $\{V^\varepsilon\}_{\varepsilon>0} \sim \text{LDP}(\mathbf{I}^V, \mathcal{C}_T)$ with a good rate function \mathbf{I}^V with $\mathcal{D}(\mathbf{I}^V) \subset \mathbb{H}_T^{v_0}$.

Thus, to continue, we first need to show that Assumption 2.1.1 holds in our setting and check whether any further assumption on the coefficients are necessary. Many processes arising from volatility models, where the classical Freidlin-Wentzell does not apply, have been studied in the literature. For example Donati-Martin, Rouault, Yor and Zani [9] show that V^ε satisfies an LDP in the case of the Heston model (4.4), then Chiarini and Fischer [5] show existence of an LDP for class of models with uniformly continuous coefficients on compacts, Conforti, De Marco and Deuschel [6] show an LDP for non-Lipschitz diffusion coefficient of CEV type. Most notably, Baldi and Caramellino [1] cover the case of Assumption 1.4.1 with $f(0) > 0$ and sub-linear growth when $f, g \rightarrow \infty$. We now state their main result, which requires a small additional assumption.

Assumption 2.1.2. *Both $f, g : \mathbb{R} \rightarrow \mathbb{R}$ have at most linear growth at infinity and $f(0) > 0$.*

Theorem 2.1.1 (Theorem 1.2 in [1]). *Let V^ε solve (2.1) on $[0, T]$ with $v_0 > 0$. Then under Assumptions 1.4.1 and 2.1.2 the process V^ε satisfies an LDP with good rate function*

$$I^V(\psi) = \begin{cases} \frac{1}{2} \int_0^T \dot{w}_t^2 dt, & \text{if } w \in \mathbb{H}_T^0, \psi_t = f(\psi_t) + g(\psi_t) \dot{w}_t, \psi_0 = v_0, \\ +\infty, & \text{otherwise.} \end{cases}$$

The process $Y^\varepsilon = \int_0^\cdot \sqrt{V_t^\varepsilon} dW_t^{\varepsilon, \perp}$ is simply the Itô integral of a square root diffusion against a Brownian motion, which falls exactly into the setting of Garcia's Theorem [14]. Before stating it (Theorem 2.1.2) though, we require the notion of *uniform exponential tightness*.

Definition 2.1.1 (Definition 1.2 in [14]). *Let \mathcal{U} denote the space of simple, real-valued, adapted processes H such that $\sup_{t \geq 0} |H_t| \leq 1$. A sequence of semi-martingales $\{M^\varepsilon\}_{\varepsilon > 0}$ is uniformly exponentially tight (UET) if, for every $c, t > 0$, there exists $K_{c,t} > 0$ such that*

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \left(\sup_{H \in \mathcal{U}} \mathbb{P} \left[\sup_{s \leq t} \left| \int_0^s H_u dM_u^\varepsilon \right| \geq K_{c,t} \right] \right) \leq -c.$$

Theorem 2.1.2 (Theorem 1.2 in [14]). *Let $\{Z^\varepsilon\}_{\varepsilon > 0}$ be a sequence of adapted, càdlàg stochastic processes, and $\{M^\varepsilon\}_{\varepsilon > 0}$ a sequence of uniformly exponentially tight semi-martingales. If the sequence $\{Z^\varepsilon, M^\varepsilon\}_{\varepsilon > 0}$ satisfies an LDP with rate function \hat{I} , then the sequence of triples $\{Z^\varepsilon, M^\varepsilon, Z^\varepsilon \circ M^\varepsilon\}_{\varepsilon > 0}$ satisfies an LDP with the good rate function*

$$I(z, m, \varphi) = \begin{cases} \hat{I}(z, m), & \varphi = z \circ m, m \in \mathcal{BV}, \\ \infty, & \text{otherwise.} \end{cases}$$

In particular, the sequence of stochastic integrals $\{Z^\varepsilon \circ M^\varepsilon\}_{\varepsilon > 0}$ satisfies an LDP with rate function

$$\hat{I}(\varphi) := \inf \{I(z, m) : \varphi = z \circ m, m \in \mathcal{BV}\}.$$

Theorem 2.1.3. *Under Assumption 2.1.1, $\{V, W^\perp, Y^\varepsilon\} \sim \text{LDP}(I^{V, W, Y}, \mathcal{C}([0, T] \rightarrow \mathbb{R}^3))$ with*

$$I^{V, W, Y} = \begin{cases} I^V(\psi) + I^W(w), & \phi = \sqrt{\psi} \circ w, w \in \mathbb{H}_0^T, \\ \infty, & \text{otherwise.} \end{cases}$$

Moreover, we also have $\{V^\varepsilon, Y^\varepsilon\} \sim \text{LDP}(I^{V, Y}, \mathcal{C}([0, T] \rightarrow \mathbb{R}^2))$, where

$$I^{V, Y}(\psi, \phi) = \inf \left\{ I^V(\psi) + I^W(w) : \phi = \int_0^\cdot \sqrt{\psi_t} \dot{w}_t dt \right\},$$

with $I^W(w) = \frac{1}{2} \int_0^T \dot{w}_t^2 dt$ if $w \in \mathbb{H}_T^0$ and infinite otherwise.

Proof. Since the scaled Brownian motion $W^{\varepsilon, \perp}$ is a uniformly exponentially tight martingale [14, Example 2.1], we only need the LDP for the pair $\{\sqrt{V^\varepsilon}, W^{\varepsilon, \perp}\}$ to apply Theorem 2.1.2 in order to obtain an LDP for the stochastic integral Y^ε . The LDP for $\sqrt{V^\varepsilon}$ is immediate from the Contraction principle [8, Theorem 4.2.1], where the corresponding rate function reads $I^{\sqrt{V}}(\psi) = I^V(\psi^2)$ and the joint $\{\sqrt{V^\varepsilon}, W^{\varepsilon, \perp}\} \sim \text{LDP}(I^{\sqrt{V}, W}, \mathcal{C}([0, T] \rightarrow \mathbb{R}^2))$ is thus a classical result, where the rate function is simply a sum of corresponding rate functions: $I^{\sqrt{V}, W}(\psi, w) = I^V(\psi^2) + I^W(w)$. Here, I^W is nothing else than the usual energy function for the

Brownian motion from Schilder's theorem [38]. This allows us to apply Theorem 2.1.2 to obtain $\{\sqrt{V}, W^\perp, Y^\varepsilon\} \sim \text{LDP} \left(\mathbb{I}^{\sqrt{V}, W, Y}, \mathcal{C}([0, T] \rightarrow \mathbb{R}^3) \right)$ with $\mathbb{I}^{\sqrt{V}, W, Y}(\psi, w, \phi) = \mathbb{I}^{\sqrt{V}, W}(\psi, w)$ if $\phi = \psi \circ w$ for $w \in \mathcal{BV}$ and infinite otherwise. Let us now consider a continuous map $(x, y, z) \mapsto (x^2, y, z)$ through which we recover the LDP for $\{V^\varepsilon, W^{\varepsilon, \perp}, Y^\varepsilon\}$, where the rate function $\mathbb{I}^{V, W, Y}(\psi, w, \phi) = \mathbb{I}^{\sqrt{V}, W, Y}(\sqrt{\psi}, w, \phi)$ is granted by the Contraction principle. Finally, by projection, we have the LDP for the pair $\{V^\varepsilon, Y^\varepsilon\}$ with the rate function

$$\begin{aligned} \mathbb{I}^{V, Y}(\psi, \phi) &= \inf \left\{ \mathbb{I}^{V, W, Y}(\psi, w, \phi) : w \in \mathbb{H}_0^T \right\} = \inf \left\{ \mathbb{I}^{\sqrt{V}, Y}(\sqrt{\psi}, \phi) : \phi = \sqrt{\psi} \circ w, w \in \mathbb{H}_0^T \right\} \\ &= \inf \left\{ \mathbb{I}^V(\psi) + \mathbb{I}^W(w) : \phi = \sqrt{\psi} \circ w, w \in \mathbb{H}_0^T \right\}, \end{aligned}$$

where in the second line, the fact that absolute continuity implies bounded variation was used after application of Theorem 2.1.2. \square

Remark 2.1.1. *The form of the rate function is the same as in [36, Lemma 3.1], but now holds for a general volatility process adhering to aforementioned Assumption 2.1.1.*

We now continue similarly to [36, Section 3]: the LDP for X^ε and its corresponding price process S^ε are obtained from Theorem 2.1.3 using the Contraction principle [8, Theorem 4.2.1], since both maps $X^\varepsilon = X^\varepsilon(V^\varepsilon, Y^\varepsilon)$ and $S^\varepsilon = S^\varepsilon(V^\varepsilon, Y^\varepsilon)$ are continuous.

2.1.2. LDP-based importance sampling. We consider two changes of measure, with a deterministic and a stochastic change of drift, and start with the former.

Deterministic change of drift. The drift is of the form

$$\left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_T} := \exp \left\{ \dot{\mathbf{h}} \circ_T \mathbf{W}^\top - \frac{1}{2} \|\dot{\mathbf{h}}\|_T^2 \right\},$$

with $\mathbf{h} \in \mathbb{H}_T^0$. The limit (1.8) then reads

$$L(\mathbf{h}) = \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{1}{\varepsilon} \left(2F(X^\varepsilon) - \dot{\mathbf{h}} \circ_T \mathbf{W}^{\varepsilon \top} + \frac{1}{2} \|\dot{\mathbf{h}}\|_T^2 \right) \right\} \right].$$

We now follow the same approach as in the case of deterministic volatility in Section 4.1. Let $F : \mathcal{C}_T \rightarrow \mathbb{R}^+$ be continuous and bounded from above and $\dot{\mathbf{h}}$ be of finite variation, then the tail condition of Varadhan's Lemma in Section B.0.1 is immediate from Lemma 4.1.1 and the functional L in (1.9) reads

$$L(\mathbf{h}) = \sup_{\mathbf{x} \in \mathbb{H}_T^0} \mathfrak{L}(\mathbf{x}, \mathbf{h}), \quad \text{with} \quad \mathfrak{L}(\mathbf{x}, \mathbf{h}) := 2F(\varphi_T(\mathbf{x})) - \dot{\mathbf{h}} \circ_T \mathbf{x}^\top + \frac{1}{2} \left(\|\dot{\mathbf{h}}\|_T^2 - \|\dot{\mathbf{x}}\|_T^2 \right),$$

where $\varphi(\mathbf{x})$ is the unique solution on $[0, T]$ to

$$\dot{\varphi}_t(\mathbf{x}) = -\frac{1}{2} \psi_t(\mathbf{x}) + \sqrt{\psi_t(\mathbf{x})} \boldsymbol{\varrho} \dot{\mathbf{x}}(t)^\top, \quad \text{with} \quad \dot{\psi}_t(\mathbf{x}) = f(\psi_t(\mathbf{x})) + g(\psi_t(\mathbf{x})) \dot{x}_1(t),$$

with initial conditions $\varphi_0(\mathbf{x}) = 0$ and $\psi_0 = v_0$, and $\boldsymbol{\varrho} := (\rho, \bar{\rho})$. To solve the dual problem (1.12), the inner optimisation reads

$$\begin{aligned} \inf_{\mathbf{h} \in \mathbb{H}_T^0} \mathfrak{L}(\mathbf{x}, \mathbf{h}) &= \inf_{\mathbf{h} \in \mathbb{H}_T^0} \left\{ 2F(\varphi_T(\mathbf{x})) - \dot{\mathbf{h}} \circ_T \mathbf{x}^\top + \frac{1}{2} \left(\|\dot{\mathbf{h}}\|_T^2 - \|\dot{\mathbf{x}}\|_T^2 \right) \right\} \\ &= 2F(\varphi_T(\mathbf{x})) + \inf_{\mathbf{h} \in \mathbb{H}_T^0} \left\{ \frac{\|\dot{\mathbf{h}}\|_T^2}{2} - \dot{\mathbf{h}} \circ_T \mathbf{x}^\top - \frac{\|\dot{\mathbf{x}}\|_T^2}{2} \right\} \\ &= 2F(\varphi_T(\mathbf{x})) - \|\dot{\mathbf{x}}\|_T^2, \end{aligned}$$

and can then be solved as

$$(2.2) \quad \mathbf{h}^* = \arg \max_{\mathbf{x} \in \mathbb{H}_T^0} \left\{ F(\varphi_T(\mathbf{x})) - \frac{\|\dot{\mathbf{x}}\|_T^2}{2} \right\},$$

which is an asymptotically optimal change of drift in the sense of Definition 1.4.1.

Stochastic change of drift. We now consider the stochastic change of measure

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_T} := \exp \left\{ \left(\dot{\mathbf{h}}\sqrt{V} \right) \circ_T \mathbf{W}^\top - \frac{1}{2} \left\| \dot{\mathbf{h}}\sqrt{V} \right\|_T^2 \right\},$$

with $\dot{\mathbf{h}}$ a deterministic function of finite variation such that $\mathbb{E}[\frac{d\mathbb{Q}}{d\mathbb{P}}] = 1$; this holds for example under the Novikov condition $\mathbb{E} \left[\exp(\frac{1}{2} \int_0^T \|\dot{\mathbf{h}}(t)\|^2 V_t dt) \right] < \infty$. We again consider $G : \mathcal{C}_T \rightarrow \mathbb{R}^+$ and let $F := \log |G|$. The minimisation problem (1.8) now reads

$$\begin{aligned} L(\mathbf{h}) &= \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{1}{\varepsilon} \left(2F(X^\varepsilon) - (\dot{\mathbf{h}}\sqrt{V^\varepsilon}) \circ_T \mathbf{W}^{\varepsilon\top} + \frac{1}{2} \|\dot{\mathbf{h}}\sqrt{V^\varepsilon}\|_T^2 \right) \right\} \right] \\ &= \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{1}{\varepsilon} \left(2F(X^\varepsilon) - \dot{\mathbf{h}} \circ_T \mathbf{Y}^{\varepsilon\top} + \frac{1}{2} \|\dot{\mathbf{h}}\sqrt{V^\varepsilon}\|_T^2 \right) \right\} \right]. \end{aligned}$$

Since F is continuous, the term inside the exponential is a continuous function of $\{V^\varepsilon, \mathbf{Y}^\varepsilon\}$. Varadhan's Lemma then yields $L(\mathbf{h}) = \sup_{\mathbf{x} \in \mathbb{H}_T^0} \mathfrak{L}(\mathbf{x}, \mathbf{h})$, with

$$\mathfrak{L}(\mathbf{x}, \mathbf{h}) = 2F(\varphi_T(\mathbf{x})) - \left(\dot{\mathbf{h}}\sqrt{\psi(\mathbf{x})} \right) \circ_T \mathbf{x}^\top + \frac{\|\dot{\mathbf{h}}\sqrt{\psi(\mathbf{x})}\|_T^2}{2} - \frac{\|\dot{\mathbf{h}}\|_T^2}{2},$$

where $\mathbf{x} = (x_1 \ x_2)^\top$ and $\{\varphi(\mathbf{x}), \psi(\mathbf{x})\}$ are unique solutions on $[0, T]$ to

$$\dot{\varphi}_t(\mathbf{x}) = -\frac{1}{2}\psi_t(\mathbf{x}) + \sqrt{\psi_t(\mathbf{x})}\boldsymbol{\varrho}\dot{\mathbf{x}}(t)^\top, \quad \text{with} \quad \dot{\psi}_t(\mathbf{x}) = f(\psi_t(\mathbf{x})) + g(\psi_t(\mathbf{x}))\dot{x}_1(t),$$

with initial conditions $(\varphi_0, \psi_0) = (0, v_0)$. For the dual problem, we search for a change of measure with \mathbf{h} such that:

$$(2.3) \quad \begin{cases} \mathbf{x}^* &= \arg \max_{\mathbf{x} \in \mathbb{H}_T^0} \left\{ F(\varphi_T(\mathbf{x})) - \frac{\|\dot{\mathbf{x}}\|_T^2}{2} \right\}, \\ \dot{\varphi}_t(\mathbf{x}) &= -\frac{1}{2}\psi_t(\mathbf{x}) + \sqrt{\psi_t(\mathbf{x})}\boldsymbol{\varrho}\dot{\mathbf{x}}(t)^\top, & \varphi_0 = 0, \\ \dot{\psi}_t(\mathbf{x}) &= f(\psi_t(\mathbf{x})) + g(\psi_t(\mathbf{x}))\dot{x}_1(t), & \psi_0 = v_0, \\ \mathbf{h}^* &= \int_0^\cdot \frac{\dot{\mathbf{x}}^*(t)}{\sqrt{\psi_t(\mathbf{x})}} dt. \end{cases}$$

The maximisation problem is very similar to the one with deterministic change of drift (2.2). However, as we will see in Section 5, the stochastic version usually gives better results.

2.1.3. Application to options with path-dependent payoff. Consider a payoff $G(\vartheta \circ_T X)$ with $G : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ differentiable, ϑ a positive function of class \mathcal{C}^1 and $F := \log |G|$. We only look at the deterministic case, namely the optimisation problem (2.2) since the solutions to (2.3) can be easily deduced from it. It reads

$$\begin{cases} \mathbf{h}^* &= \arg \max_{\mathbf{x} \in \mathbb{H}_T^0} \left\{ F \left(\int_0^T \vartheta_t \dot{\varphi}_t(\mathbf{x}) dt \right) - \frac{\|\dot{\mathbf{x}}\|_T^2}{2} \right\}, \\ \dot{\varphi}_t(\mathbf{x}) &= -\frac{1}{2}\psi_t(\mathbf{x}) + \sqrt{\psi_t(\mathbf{x})}\boldsymbol{\varrho}\dot{\mathbf{x}}(t)^\top, & \varphi_0 = 0, \\ \dot{\psi}_t(\mathbf{x}) &= f(\psi_t(\mathbf{x})) + g(\psi_t(\mathbf{x}))\dot{x}_1(t), & \psi_0 = v_0. \end{cases}$$

The following lemma helps transforming the above optimisation problem.

Lemma 2.1.1. *Let $\mathbf{x} \in \mathbb{H}_T^0 \times \mathbb{H}_T^0$. The function $\mathfrak{K} : \mathbb{H}_T^0 \times \mathbb{H}_T^0 \rightarrow \mathbb{H}_T^0 \times \mathbb{H}_T^{v_0,+}$ such that $\mathfrak{K}(\mathbf{x}) = (\varphi, \psi)$ is solution to*

$$\dot{\varphi}_t(\mathbf{x}) = -\frac{1}{2}\psi_t(\mathbf{x}) + \sqrt{\psi_t(\mathbf{x})}\boldsymbol{\varrho}\dot{\mathbf{x}}(t)^\top, \quad \text{with} \quad \dot{\psi}_t(\mathbf{x}) = f(\psi_t(\mathbf{x})) + g(\psi_t(\mathbf{x}))\dot{x}_1(t),$$

with initial conditions $\varphi_0 = 0$ and $\psi_0 = v_0$, is well defined and is a bijection.

Proof. Clearly $\dot{\psi} \in L^2([0, T]; \mathbb{R})$ and the unique solution ψ to the ODE $\dot{\psi}_t = f(\psi_t) + g(\psi_t)\dot{w}_t$ with $\psi_0 = v_0 > 0$ is strictly positive under Assumption 1.4.1 by [1, Proposition 3.11]. Therefore $\psi \in \mathbb{H}_T^{v_0,+}$ and $\mathfrak{K}(\mathbf{x}) = (\varphi, \psi)$ is well defined. Finally, \mathfrak{K} is clearly a bijection and its

inverse can be computed explicitly as

$$\mathfrak{K}^{\leftarrow}(\mathbf{z}) = \left(\int_0^t \frac{\dot{z}_2(t) - f(z_2(t))}{g(z_2(t))} dt, \frac{1}{\bar{\rho}} \int_0^t \left\{ \frac{\dot{z}_1(t) + \frac{1}{2}z_2(t)}{\sqrt{z_2(t)}} - \rho \frac{\dot{z}_2(t) - f(z_2(t))}{g(z_2(t))} \right\} dt \right).$$

□

Using Lemma 2.1.1, we can substantially simplify the optimisation problem by writing it in terms of $\mathfrak{K}(\mathbf{x})$. To be more precise, we will make use of the following transformation:

$$(2.4) \quad U(\psi) = \frac{\dot{\psi} - f(\psi)}{g(\psi)} \quad \text{and} \quad Z(\varphi, \psi) = \frac{\dot{\varphi} + \frac{1}{2}\psi}{\sqrt{\psi}},$$

which stems from the two components of $\mathfrak{K}(\mathbf{x})$. The optimisation problem becomes

$$\begin{cases} \varphi^*, \psi^* &= \arg \max_{\{\varphi, \psi\} \in \mathbb{H}_T^0 \times \mathbb{H}_T^{v_0, +}} \left\{ F \left(\int_0^T \vartheta_t \dot{\varphi}_t dt \right) - \frac{1}{2} \int_0^T \left\{ U(\psi_t)^2 + \left(\frac{Z(\varphi_t, \psi_t) - \rho U(\psi_t)}{\bar{\rho}} \right)^2 \right\} dt \right\}, \\ h_1^* &= U(\psi^*), \\ h_2^* &= \frac{Z(\varphi^*, \psi^*) - \rho U(\psi^*)}{\bar{\rho}}. \end{cases}$$

This allows us to apply Euler-Lagrange to the problem seen as an optimisation over $\{\int_0^T \vartheta_t \dot{\varphi}_t dt, \psi\}$:

$$(2.5) \quad \begin{cases} \frac{d}{dt} \left\{ \frac{Z - \rho U}{\vartheta \bar{\rho}^2 \sqrt{\psi}} \right\} = 0, \\ \frac{d}{dt} \left\{ \frac{U}{g(\psi)} - \frac{\rho}{\bar{\rho}^2} \frac{Z - \rho U}{g(\psi)} \right\} = -\frac{U}{g^2(\psi)} \zeta + \frac{Z - \rho U}{\bar{\rho}^2} \left\{ \frac{1}{2\sqrt{\psi}} - \frac{Z}{2\psi} + \frac{\rho}{g^2(\psi)} \zeta \right\}, \end{cases}$$

where $\zeta := f'(\psi)g(\psi) + (\dot{\psi} - f(\psi))g'(\psi)$. This system of equations is still hard to solve for general f and g , but can be solved for the Heston model, as done in Section 4.4 below.

2.2. Small-time LDP. Applying the mapping $t \mapsto \varepsilon t$ to (1.6) yields

$$(2.6) \quad \begin{aligned} dX_t^\varepsilon &= -\frac{1}{2}\varepsilon V_t^\varepsilon dt + \sqrt{\varepsilon} \sqrt{V_t^\varepsilon} dB_t, & X_0^\varepsilon &= 0, \\ dV_t^\varepsilon &= \varepsilon f(V_t^\varepsilon) dt + \sqrt{\varepsilon} g(V_t^\varepsilon) dW_t, & V_0 &= v_0. \end{aligned}$$

Robertson [36, Proposition 3.2] showed that $\varepsilon \int_0^T V_t^\varepsilon dt$ is in fact exponentially equivalent to zero (Definition B.0.1), so that the drift of V^ε can be ignored at the large deviations level. In the case of a general drift f , the following lemma provides a similar statement:

Lemma 2.2.1. *The process $\varepsilon \int_0^T f(V_t^\varepsilon) dt$ is exponentially equivalent to zero.*

Proof. Markov's inequality implies that, for any $\delta > 0$,

$$\mathbb{P} \left[\int_0^T f(V_s^\varepsilon) ds > \frac{2\delta}{\varepsilon} \right] \leq \exp \left\{ -\frac{2\delta}{\varepsilon^2} \right\} \mathbb{E} \left[\exp \left\{ \frac{1}{\varepsilon} \int_0^T f(V_s^\varepsilon) ds \right\} \right],$$

and we are therefore left to show that $\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{1}{\varepsilon} \int_0^T f(V_s^\varepsilon) ds \right\} \right]$ is finite. To that end we apply the integral Jensen inequality

$$\mathbb{E} \left[\exp \left\{ \frac{1}{\varepsilon} \int_0^T f(V_s^\varepsilon) ds \right\} \right] \leq \frac{1}{T} \int_0^T \mathbb{E} \left[\exp \left\{ \frac{T}{\varepsilon} f(V_s^\varepsilon) \right\} \right] ds,$$

and the linear growth condition from the global Lipschitz condition in Assumption 1.4.1:

$$\frac{1}{T} \int_0^T \mathbb{E} \left[\exp \left\{ \frac{T}{\varepsilon} f(V_s^\varepsilon) \right\} \right] ds \leq \frac{1}{T} \int_0^T \mathbb{E} \left[\exp \left\{ \frac{T}{\varepsilon} (1 + |V_s^\varepsilon|) \right\} \right] ds = \frac{1}{T} e^{\frac{T}{\varepsilon}} \int_0^T \mathbb{E} \left[\exp \left\{ \frac{T}{\varepsilon} |V_s^\varepsilon| \right\} \right] ds.$$

Next, by the properties of the logarithm and the supremum

$$\begin{aligned} \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{1}{\varepsilon} \int_0^T f(V_s^\varepsilon) ds \right\} \right] \\ \leq \limsup_{\varepsilon \downarrow 0} \left\{ \varepsilon \log \frac{1}{T} \right\} + T + \limsup_{\varepsilon \downarrow 0} \varepsilon \log \left\{ \int_0^T \mathbb{E} \left[\exp \left\{ \frac{T}{\varepsilon} |V_s^\varepsilon| \right\} \right] ds \right\}. \end{aligned}$$

We can now apply Gronwall's Lemma to the last term, which yields for some $C > 0$,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{1}{\varepsilon} \int_0^T f(V_s^\varepsilon) ds \right\} \right] \leq T + \limsup_{\varepsilon \downarrow 0} \varepsilon \log \left\{ \exp \left\{ CT + \frac{T}{\varepsilon} v_0 \right\} \right\} = T(1 + v_0),$$

which is finite. \square

Following this lemma, the results from the previous section could simply be adapted so that $\{V^\varepsilon, W^\varepsilon\}$ satisfy the same LDP by simply setting $f = 0$ (or equivalently $\kappa = 0$ in the case of Heston (4.4)). However, this violates the condition $f(0) > 0$ in Baldi and Caramellino [1]. Fortunately, Conforti, Deuschel and De Marco [6] removed the need for strict positivity on the drift at the initial time by imposing more stringent conditions on the diffusion.

Assumption 2.2.1.

- (i) *There exists $\xi > 0$ such that $g(y) = \xi|y|^\gamma$ for $\gamma \in [1/2, 1)$ for all $y \geq 0$;*
- (ii) *The equality $f(y) = \tau(y) + Ky$ holds for all $y \geq 0$, where τ is a Lipschitz continuous and bounded function, and $\tau(y) \geq 0$ in a neighbourhood of the origin.*

Theorem 2.2.1 (Theorem 1.1 in [6]). *Under Assumption 2.2.1, the solution V^ε to (2.1) satisfies $\{V^\varepsilon\} \sim \text{LDP}(\mathbf{I}^V, \mathcal{C}([0, T] \rightarrow \mathbb{R}^+))$ with*

$$\mathbf{I}^V(\varphi) = \frac{1}{2\xi^2} \int_0^T \left(\frac{\dot{\varphi}_t - K\varphi_t}{\varphi_t^\gamma} \right)^2 \mathbf{1}_{\{\varphi_t \neq 0\}} dt.$$

Therefore by setting $K = 0$ we can use the methodology from the previous section since the LDPs are the same. Similarly as before, we only consider the deterministic change of drift, since the stochastic case is very similar. We therefore search for \mathbf{h} such that

$$\mathbf{h}^* = \arg \max_{\mathbf{x} \in \mathbb{H}_T^0} \left\{ F(\varphi_T(\mathbf{x})) - \frac{\|\dot{\mathbf{x}}\|_T^2}{2} \right\},$$

where $\varphi_t(\mathbf{x})$ is the unique solution on $[0, T]$ to

$$\begin{cases} \varphi_t(\mathbf{x}) &= \sqrt{\psi_t(\mathbf{x})} \boldsymbol{\varrho} \dot{\mathbf{x}}(t)^\top, & \varphi_0(\mathbf{x}) = 0, \\ \dot{\psi}_t(\mathbf{x}) &= g(\psi_t(\mathbf{x})) \dot{x}_1(t), & \psi_0 = v_0. \end{cases}$$

2.3. Large-time LDP. Variance reduction for affine stochastic volatility processes via importance sampling through the large-time approximation is extensively covered in [19], so we do not repeat the study and refer the reader to the aforementioned work.

3. IMPORTANCE SAMPLING VIA MODERATE DEVIATIONS

In the previous sections, large deviations provided us with a way of computing the asymptotic change of measure for importance sampling, via an ε -approximation of the log-price X^ε . While the large deviation rate function is a convenient quadratic function in the deterministic volatility setting, it is in general rather cumbersome to compute numerically, unfortunately offsetting any importance sampling gain. Moderate deviations act on a cruder scale, but provide quadratic rate functions, easier to compute. Suppose that the sequence $\{X^\varepsilon\}_{\varepsilon > 0}$ converges in probability to \bar{X} . Moderate deviations for $\{X^\varepsilon\}_{\varepsilon > 0}$ are defined as large deviations for the rescaled sequence

$$\left\{ \frac{X^\varepsilon - \bar{X}}{\sqrt{\varepsilon h(\varepsilon)}} \right\}_{\varepsilon > 0},$$

where $h(\varepsilon)$ tends to infinity and $\sqrt{\varepsilon}h(\varepsilon)$ to zero as ε tends to zero. A typical choice is $h(\varepsilon) = \varepsilon^{-\tilde{\alpha}}$ for $\tilde{\alpha} \in (0, \frac{1}{2})$, equivalently $\frac{1}{\sqrt{\varepsilon}h(\varepsilon)} = \varepsilon^{-\alpha}$ for $\alpha \in (0, \frac{1}{2})$. We shall stick to this choice of h here in order to highlight rates of convergence. We now introduce the approximation

$$(3.1) \quad \tilde{X}^\varepsilon := \bar{X} + \varepsilon^{-\alpha} (X^\varepsilon - \bar{X}).$$

This process is centered around \bar{X} and is a simple candidate. Furthermore, in stochastic volatility models, and particularly in large-time setting, the moderate deviations rate function is simply the second-order Taylor expansion of the large deviations rate function around its minimum \bar{X} [27, Remark 3.5]. We again consider the dynamics (1.6) with Assumption 1.4.1 for the coefficients. We further assume the following conditions:

Assumption 3.0.1.

- (i) For $f \in \mathcal{C}^2(\mathbb{R}^+ \rightarrow \mathbb{R})$, the equation $\bar{\psi}_t = v_0 + \int_0^t f(\bar{\psi}_s) ds$ admits a unique strictly positive solution $\bar{\psi} \in \mathcal{C}^2([0, T] \rightarrow \mathbb{R})$;
- (ii) The small-noise approximation (2.1) of V satisfies an LDP with the good rate function I^V and speed ε such that I^V admits a unique minimum and is null there.

Remark 3.0.1. When a large deviations principle holds, then the infimum of the rate function has to be zero, indeed, since $\mathbb{P}(X^\varepsilon \in B) \sim \exp\{-\frac{1}{\varepsilon} \inf_{x \in B} I(x)\}$, then taking B to be the whole space, the probability on the left-hand side has to be equal to one, and the infimum is thus null. It may not be attained though. However it is in the case of good rate functions (see for example Dembo-Zeitouni's book, bottom of Page 5). As stated at the end of Section 1.1, we are only concerned here with large deviations with good rate functions. The only gap that could appear is the case of several points where zero is attained, but this does not seem restrictive in our setup.

As it will be shown in Lemma 3.1.3, the sequence $\{V^\varepsilon\}_{\varepsilon>0}$ converges in probability to the function $\bar{\psi}$ as a consequence of Assumption 3.0.1. This provides a natural choice for the centered process $\bar{X}_t = -\frac{1}{2} \int_0^t \bar{\psi}_s ds$, so that the approximation (3.1) reads, for any $t \in [0, T]$,

$$(3.2) \quad \tilde{X}_t^\varepsilon = -\frac{1}{2} \int_0^t \bar{\psi}_s ds + \varepsilon^{-\alpha} \left(X_t^\varepsilon + \frac{1}{2} \int_0^t \bar{\psi}_s ds \right).$$

3.1. Small-noise moderate deviations. For the small-noise approximation (2.1), the process \tilde{X}^ε in (3.2) satisfies the SDE

$$\begin{aligned} d\tilde{X}_t^\varepsilon &= \left(-\frac{1}{2} + \frac{1}{2\varepsilon^\alpha} \right) \bar{\psi}_t dt + \varepsilon^{-\alpha} dX_t^\varepsilon \\ &= -\frac{1}{2} (1 - \varepsilon^{-\alpha}) \bar{\psi}_t dt + \varepsilon^{-\alpha} \left(-\frac{1}{2} V_t^\varepsilon dt + \sqrt{\varepsilon} \sqrt{V_t^\varepsilon} dB_t \right) \\ &= -\frac{1}{2} \bar{\psi}_t dt - \frac{1}{2} \varepsilon^{-\alpha} (V_t^\varepsilon - \bar{\psi}_t) dt + \varepsilon^{\frac{1}{2}-\alpha} \sqrt{V_t^\varepsilon} dB_t, \end{aligned}$$

starting at $\tilde{X}_0^\varepsilon = 0$, where the small-noise approximation (2.1) of the variance reads

$$dV_t^\varepsilon = f(V_t^\varepsilon) dt + \sqrt{\varepsilon} g(V_t^\varepsilon) dW_t, \quad V_0^\varepsilon = v_0.$$

This transformation creates a discrepancy between the decreasing speeds of \tilde{X}^ε and that of V^ε (speed of convergence to zero of the diffusion part of the volatility process i.e. $\varepsilon^{\frac{1}{2}-\alpha}$ versus $\varepsilon^{\frac{1}{2}}$). Since \tilde{X}^ε is our reference, we adjust the speed of the LDP via $\varepsilon^{\frac{1-2\alpha}{2}} \mapsto \sqrt{\varepsilon}$. With

$$\beta := \frac{\alpha}{(1-2\alpha)} \quad \text{and} \quad \eta_t^\varepsilon := \varepsilon^{-\beta} (V_t^\varepsilon - \bar{\psi}_t),$$

we obtain the system

$$(3.3) \quad \begin{aligned} d\tilde{X}_t^\varepsilon &= -\frac{1}{2} \bar{\psi}_t dt - \frac{1}{2} \eta_t^\varepsilon dt + \sqrt{\varepsilon} \sqrt{V_t^\varepsilon} dB_t, & \tilde{X}_0^\varepsilon &= 0, \\ d\eta_t^\varepsilon &= \varepsilon^{-\beta} (f(V_t^\varepsilon) - f(\bar{\psi}_t)) dt + \sqrt{\varepsilon} g(V_t^\varepsilon) dW_t, & \eta_0^\varepsilon &= 0, \\ dV_t^{\varepsilon, \beta} &= f(V_t^{\varepsilon, \beta}) dt + \varepsilon^{\frac{1}{2}+\beta} g(V_t^{\varepsilon, \beta}) dW_t, & V_0^{\varepsilon, \beta} &= v_0, \end{aligned}$$

where the notation $V^{\varepsilon, \beta}$ was used to emphasise the change of speed to $\varepsilon^{\frac{1}{2} + \beta}$. In the following, we provide an LDP for $\{\eta^\varepsilon\}$ (equivalently an MDP for $\{V^{\varepsilon, \beta}\}$). We relegate more technical proofs to Appendix A.

3.1.1. Theoretical results. The main moderate deviations result of this section is Theorem 3.1.1, but we first start with the following three technical lemmata, useful for the theorem but also of independent interest, proved in Appendices A.1-A.2-A.3:

Lemma 3.1.1. *Let $\{Z^\varepsilon\}_{\varepsilon > 0}$ be a family of random variables mapping to any metrisable space \mathcal{X} and satisfying an LDP with good rate function $\tilde{\mathbf{I}}$. If there exists a unique x_0 such that $\tilde{\mathbf{I}}(x_0) = 0$, then for all $\beta > 1$, Z^{ε^β} satisfies an LDP with the good rate function*

$$\mathbf{I}(x) = \begin{cases} 0, & \text{for } x = x_0, \\ +\infty, & \text{elsewhere.} \end{cases}$$

If for $\beta > 1$, Z^ε satisfies an LDP with speed ε^β and the good rate function \mathbf{I} , then Z^ε is exponentially equivalent to x_0 with speed ε .

As a consequence of this lemma, the sequence $\{Z^\varepsilon\}$ converges in probability to x_0 .

Lemma 3.1.2. *Let $\{Z^\varepsilon\}_{\varepsilon > 0}$ be a sequence of random variables mapping to any metrisable space \mathcal{X} and satisfying an LDP with good rate function \mathbf{I} such that $\mathbf{I}(x) = 0$ if and only if $x = x_0$ for some $x_0 \in \mathcal{X}$. If Z^ε is uniformly integrable, then $\lim_{\varepsilon \downarrow 0} \mathbb{E}[Z^\varepsilon] = x_0$.*

Lemma 3.1.3. *Let $V^{\varepsilon, \beta}$ be given by (3.3) such that $f, g \in \mathcal{C}(\mathbb{R}^+ \rightarrow \mathbb{R})$ satisfy Assumption 1.4.1. Then $\{V^{\varepsilon, \beta}\}$ converges almost surely to the unique solution of $\dot{\bar{\psi}}_t = f(\bar{\psi}_t)$ on $[0, T]$ with boundary condition $\bar{\psi}(0) = v_0$.*

Theorem 3.1.1. *Let $\beta, v_0 > 0$ and let $f, g \in \mathcal{C}(\mathbb{R}^+ \rightarrow \mathbb{R})$ be such that Assumption 1.4.1 is satisfied and let $W^\varepsilon := \sqrt{\varepsilon}W$ and $V^{\varepsilon, \beta}, \eta^\varepsilon$ as defined in (3.3), then under Assumption 3.0.1, the triple $\{V^{\varepsilon, \beta}, \eta^\varepsilon, W^\varepsilon\}$ satisfies an LDP with speed ε and the good rate function*

$$\mathbf{I}^{V, \eta, W}(v, \eta, w) = \begin{cases} \frac{1}{2} \|\dot{w}\|_T^2, & \text{if } w \in \mathbb{H}_T^0, v = \bar{\psi}, \dot{\eta} = \dot{f}(\bar{\psi})\eta + g(\bar{\psi})\dot{w}, \\ +\infty, & \text{otherwise.} \end{cases}$$

3.1.2. MDP-based importance sampling. Consider the system (3.3), and let $\mathbf{W}^\varepsilon := \sqrt{\varepsilon}W$ and $\mathbf{Y}^\varepsilon := \sqrt{V^{\varepsilon, \beta}} \circ \mathbf{W}^\varepsilon$. First we apply Theorem 3.1.1, which in turn allows us the use of Theorem 2.1.3. Next, the Contraction principle and [8, Exercise 4.2.7] imply that the triple $\{\eta^\varepsilon, \mathbf{W}^\varepsilon, \mathbf{Y}^\varepsilon\}$ satisfies an LDP with speed ε and good rate function

$$\mathbf{I}^{\eta, \mathbf{W}, \mathbf{Y}}(\eta, \mathbf{w}, \mathbf{y}) = \begin{cases} \frac{1}{2} \|\dot{\mathbf{w}}\|_T^2, & \text{if } \mathbf{w} \in \mathbb{H}_T^0, \eta = \int_0^\cdot \left\{ \dot{f}(\bar{\psi}_s)\eta_s + g(\bar{\psi}_s)\dot{w}_1(s) \right\} ds, \mathbf{y} = \sqrt{\bar{\psi}} \circ \mathbf{w}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Remember that the exponential tightness condition necessary for [8, Exercise 4.2.7] is automatically granted, given that the space we work on is Polish [31, Lemma 2.6].

Consider a payoff $G \in \mathcal{C}_T \rightarrow \mathbb{R}^+$ and let $F := \log |G|$. Recall that we are interested in finding a measure change minimising $\mathbb{E} \left[e^{2F} \frac{d\mathbb{P}}{d\mathbb{Q}} \right]$.

Deterministic change of drift. We first consider a deterministic change of drift, via

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_T} = \exp \left\{ -\frac{1}{2} \|\dot{\mathbf{h}}\|_T^2 + \dot{\mathbf{h}} \circ_T \mathbf{W}^\top \right\},$$

for $\mathbf{h} \in \mathbb{H}_T^0$ with $\dot{\mathbf{h}}$ of finite variation. In the spirit of moderate deviations, we use the approximation

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_T} \approx \exp \left\{ -\frac{1}{2} \|\dot{\mathbf{h}}\|_T^2 + \dot{\mathbf{h}} \circ_T \mathbf{W}^{\varepsilon \top} \right\},$$

and thus aim at minimising

$$L(\mathbf{h}) = \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{1}{\varepsilon} \left(2F(\tilde{X}^\varepsilon) + \frac{1}{2} \|\dot{\mathbf{h}}\|_T^2 - \dot{\mathbf{h}} \circ_T \mathbf{W}^{\varepsilon \top} \right) \right\} \right].$$

Under the conditions of Varadhan's lemma B.0.2 (for example if F is bounded), then $L(\mathbf{h}) = \sup_{\mathbf{x} \in \mathbb{H}_T^0} \mathfrak{L}(\mathbf{x}, \mathbf{h})$ where

$$\mathfrak{L}(\mathbf{x}, \mathbf{h}) = 2F \left(-\frac{1}{2} \int_0^\cdot (\bar{\psi}_t + \eta_t(\mathbf{x})) dt + \boldsymbol{\varrho} \mathbf{Y}^\top(\mathbf{x}) \right) - \dot{\mathbf{h}} \circ_T \mathbf{x}^\top + \frac{\|\dot{\mathbf{h}}\|_T^2}{2} - \frac{\|\dot{\mathbf{x}}\|_T^2}{2},$$

with $\boldsymbol{\varrho} := (\rho \ \bar{\rho})^\top$,

$$(3.4) \quad \eta(\mathbf{x}) = \int_0^\cdot \left\{ \dot{f}(\bar{\psi}_s) \eta_s(\mathbf{x}) + g(\bar{\psi}_s) \dot{x}_1(s) \right\} ds \quad \text{and} \quad \mathbf{Y}(\mathbf{x}) = \sqrt{\bar{\psi}} \circ \mathbf{x}.$$

Minimising $L(\mathbf{h})$ is far from trivial, and hence, as before, we define the optimal change of drift \mathbf{h}^* as a solution to the dual problem $\inf_{\mathbf{h} \in \mathbb{H}_T^0} \mathfrak{L}(\mathbf{x}, \mathbf{h})$, so that, with η as in (3.4),

$$(3.5) \quad \left\{ \begin{array}{l} \mathbf{Y}(\mathbf{x}) = \sqrt{\bar{\psi}} \circ \mathbf{x}, \\ \bar{\psi}_t = v_0 + \int_0^t f(\bar{\psi}_s) ds, \\ \eta_t(\mathbf{x}) = \int_0^t \left\{ \dot{f}(\bar{\psi}_s) \eta_s(\mathbf{x}) + g(\bar{\psi}_s) \dot{x}_1(s) \right\} ds, \\ \mathbf{h}^* = \arg \max_{\mathbf{x} \in \mathbb{H}_T^0} \left\{ F \left(-\frac{1}{2} \int_0^\cdot (\bar{\psi}_t + \eta_t(\mathbf{x})) dt + \boldsymbol{\varrho} \mathbf{Y}^\top(\mathbf{x}) \right) - \frac{\|\dot{\mathbf{x}}\|_T^2}{2} \right\}. \end{array} \right.$$

Stochastic change of drift. We now consider instead a stochastic change of drift in the form $\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_T} = \exp \left\{ -\frac{1}{2} \|\dot{\mathbf{h}} \sqrt{V}\|_T^2 + (\dot{\mathbf{h}} \sqrt{V}) \circ_T \mathbf{W}^\top \right\}$, for $\mathbf{h} \in \mathbb{H}_T^0$, $\dot{\mathbf{h}}$ of finite variation and with $\mathbb{E} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \right] = 1$. Again, we use the approximation

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_T} \approx \exp \left\{ -\frac{1}{2} \|\dot{\mathbf{h}} \sqrt{V^{\varepsilon, \beta}}\|_T^2 + (\dot{\mathbf{h}} \sqrt{V^{\varepsilon, \beta}}) \circ_T (\mathbf{W}^\varepsilon)^\top \right\},$$

and aim at minimising

$$L(\mathbf{h}) = \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \frac{1}{\varepsilon} \left\{ 2F(\tilde{X}^\varepsilon) + \frac{1}{2} \|\dot{\mathbf{h}} \sqrt{V^{\varepsilon, \beta}}\|_T^2 - (\dot{\mathbf{h}} \sqrt{V^{\varepsilon, \beta}}) \circ_T \mathbf{W}^{\varepsilon \top} \right\} \right].$$

If Varadhan's lemma conditions hold, then again $L(\mathbf{h}) = \sup_{\mathbf{x} \in \mathbb{H}_T^0} \mathfrak{L}(\mathbf{x}, \mathbf{h})$ where

$$\mathfrak{L}(\mathbf{x}, \mathbf{h}) = 2F \left(-\frac{1}{2} \int_0^\cdot (\bar{\psi}_t + \eta_t(\mathbf{x})) dt + \boldsymbol{\varrho} \mathbf{Y}^\top(\mathbf{x}) \right) - \left(\dot{\mathbf{h}} \sqrt{\bar{\psi}} \right) \circ_T \mathbf{x}^\top + \frac{\|\dot{\mathbf{h}} \sqrt{\bar{\psi}}\|_T^2}{2} - \frac{\|\dot{\mathbf{x}}\|_T^2}{2},$$

with η defined as in (3.4). As minimising $L(\mathbf{h})$ is *a priori* complicated, we define our optimal change of drift \mathbf{h}^* as a solution to the dual problem

$$\left\{ \begin{array}{l} \mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathbb{H}_T^0} \left\{ F \left(-\frac{1}{2} \int_0^\cdot (\bar{\psi}_t + \eta_t(\mathbf{x})) dt - \boldsymbol{\varrho} \mathbf{Y}^\top(\mathbf{x}) \right) - \frac{\|\dot{\mathbf{x}}\|_T^2}{2} \right\}, \\ \bar{\psi}_t = v_0 + \int_0^t f(\bar{\psi}_s) ds, \\ \eta_t(\mathbf{x}) = \int_0^t \left\{ \dot{f}(\bar{\psi}_s) \eta_s(\mathbf{x}) + g(\bar{\psi}_s) \dot{x}_1(s) \right\} ds, \\ \mathbf{h}^* = \int_0^\cdot \frac{\dot{\mathbf{x}}^*(t)}{\sqrt{\bar{\psi}_t}} dt. \end{array} \right.$$

3.1.3. Application to options with path-dependent payoff. Consider an option with payoff $G(\vartheta \circ_T X)$, where $G : \mathbb{R} \rightarrow \mathbb{R}^+$ is a differentiable function and ϑ a positive (almost everywhere) function of class $\mathcal{C}^1([0, T] \rightarrow \mathbb{R}^+)$. The payoff is then a continuous function of the path. Now let $F = \log |G|$ and $\bar{F}(x) = F(x - \frac{1}{2} \int_0^T \bar{\psi}_s \vartheta_s ds)$ and suppose that Assumptions 1.4.1 and 3.0.1 hold.

Deterministic change of drift. We proceed in a similar fashion as in Section 2 and consider the transformations

$$\dot{\varphi} = \sqrt{\bar{\psi}}(\rho\dot{x}_1 + \bar{\rho}\dot{x}_2) - \frac{\eta}{2} \quad \text{and} \quad \dot{\eta} = \dot{f}(\bar{\psi})\eta + g(\bar{\psi})\dot{x}_1,$$

so that the optimisation problem (3.5) for a path-dependent payoff can then be written as (3.6)

$$\begin{cases} \varphi^*, \eta^* = \arg \max_{\varphi, \eta \in \mathbb{H}_T^0} \left\{ \bar{F} \left(\int_0^T \vartheta_t \dot{\varphi}_t dt \right) - \frac{1}{2} \int_0^T \left\{ U(\eta)^2 + \left| \frac{Z(\varphi, \eta) - \rho U(\eta)}{\bar{\rho}} \right|^2 \right\} dt \right\}, \\ \mathbf{h}^* = \left(U(\eta^*), \frac{1}{\bar{\rho}} (Z(\varphi^*, \eta^*) - \rho U(\eta^*)) \right), \end{cases}$$

with

$$U(\eta) = \frac{\dot{\eta} - \dot{f}(\bar{\psi})\eta}{g(\bar{\psi})} \quad \text{and} \quad Z(\varphi, \eta) = \frac{\dot{\varphi} + \frac{1}{2}\eta}{\sqrt{\bar{\psi}}}.$$

Then, by applying Euler-Lagrange to the problem seen as an optimisation over $\{\int_0^T \vartheta_t \dot{\varphi}_t dt, \eta\}$, we obtain the system of ODEs

$$\begin{cases} Z - \rho U &= \frac{1}{2} \lambda \bar{\rho}^2 \vartheta \sqrt{\bar{\psi}}, \\ \frac{d}{dt} \left\{ \frac{2U}{g(\bar{\psi})} - \lambda \rho \vartheta \frac{\sqrt{\bar{\psi}}}{g(\bar{\psi})} \right\} &= \lambda \vartheta \sqrt{\bar{\psi}} \left(\frac{1}{2\sqrt{\bar{\psi}}} + \rho \frac{\dot{f}(\bar{\psi})}{g(\bar{\psi})} \right) - 2U \frac{\dot{f}(\bar{\psi})}{g(\bar{\psi})}, \end{cases}$$

with boundary conditions $\lambda = -2\bar{F}'_T$ and $U_T = -\rho \vartheta_T \sqrt{\bar{\psi}_T} \bar{F}'_T$, where $\bar{F}'_T := \bar{F}' \left(\int_0^T \vartheta_t \dot{\varphi}_t dt \right)$.

Introducing $A := \frac{2U}{g(\bar{\psi})} - \lambda \rho \vartheta \frac{\sqrt{\bar{\psi}}}{g(\bar{\psi})}$ simplifies the problem to the linear ODE: $\dot{A} - \dot{f}A = \frac{1}{2} \lambda \vartheta$ with $A_T = 0$, with solution

$$A_t = \frac{\lambda}{2} e^{-B_t} \left(\int_0^t e^{B_s} \vartheta_s ds - \gamma_T \right), \quad \text{for } t \in [0, T],$$

where $B := \int_0^\cdot f(\bar{\psi}_t) dt$ and $\gamma := \int_0^\cdot e^{B_t} \vartheta_t dt$. We can now solve for U and Z :

$$U = \lambda \left(u + \frac{1}{2} \rho \vartheta \sqrt{\bar{\psi}} \right) \quad \text{and} \quad Z = \lambda \left(\rho u + \frac{1}{2} \vartheta \sqrt{\bar{\psi}} \right),$$

with $u = \frac{1}{4} g(\bar{\psi}) e^{-B} (\gamma - \gamma_T)$. Our optimisation problem was posed over $\{\int_0^T \vartheta_t \dot{\varphi}_t dt, \eta\}$ so we require the solution in terms of this couple, and therefore

$$\begin{aligned} \text{(i)} \quad & \int_0^t \vartheta_s \dot{\varphi}_s ds = \lambda \int_0^t \vartheta_s \left\{ \dot{\varphi}_s^1 - \frac{1}{2} \dot{\varphi}_s^2 \right\} ds + \eta_0 T e^{B_T}, \text{ where} \\ & \dot{\varphi}^1 = \left(u + \frac{1}{2} \vartheta \sqrt{\bar{\psi}} \right) \sqrt{\bar{\psi}} \quad \text{and} \quad \dot{\varphi}^2 = e^B \int_0^\cdot e^{-B_s} \left(u_s + \frac{1}{2} \vartheta_s \sqrt{\bar{\psi}_s} \right) g(\bar{\psi}_s) ds; \\ \text{(ii)} \quad & \eta_t = \lambda e^{B_t} \int_0^t e^{-B_s} \left(u_s + \frac{1}{2} \rho \vartheta_s \sqrt{\bar{\psi}_s} \right) g(\bar{\psi}_s) ds + \eta_0 e^{B_t}. \end{aligned}$$

Here, $\lambda, \eta_0 \in \mathbb{R}$ are parameters over which we perform our optimisation. Thus the original optimisation objective (3.6) becomes

$$\lambda^*, \eta_0^* = \arg \max_{\lambda, \eta_0 \in \mathbb{R}} \left\{ \bar{F} \left(\lambda \int_0^T \vartheta_s \left\{ \dot{\varphi}_s^1 - \frac{1}{2} \dot{\varphi}_s^2 \right\} ds + \eta_0 T e^{B_T} \right) - \frac{\lambda^2}{2} \int_0^T \left\{ |\dot{\varphi}_t^2|^2 + \frac{\bar{\rho}^2}{4} \vartheta_t^2 \bar{\psi}_t \right\} dt \right\}.$$

Stochastic change of drift. The stochastic change of drift objective is equivalent to the one with deterministic drift, the difference being how \mathbf{h} is calculated.

3.2. Small-time moderate deviations. We now mimic the results of the previous section, but in the context of small-time moderate deviations. We again consider the log-price dynamics (1.6) under Assumption 1.4.1 and Assumption 3.0.1. Let $\alpha \in (0, \frac{1}{2})$ and define $\tilde{X}^\varepsilon := \varepsilon^{-\alpha} X^\varepsilon$, so that

$$\begin{aligned} d\tilde{X}_t^\varepsilon &= -\frac{1}{2} \varepsilon^{1-\alpha} V_t^\varepsilon dt + \varepsilon^{\frac{1}{2}-\alpha} \sqrt{V_t^\varepsilon} dB_t, \quad \tilde{X}_0^\varepsilon = 0, \\ dV_t^\varepsilon &= \varepsilon f(V_t^\varepsilon) dt + \sqrt{\varepsilon} g(V_t^\varepsilon) dW_t, \quad V_0^\varepsilon = v_0. \end{aligned}$$

As a consequence of Theorem 3.2.1 the results remain the same as in the case of small-noise moderate deviations of the previous section with $f = 0$ and $\bar{\psi} = v_0$. This being the case, we do not repeat them here. Let us nevertheless note that in the case of change of measure with deterministic drift, the problem is similar to the one, where V is constant equal to v_0 .

3.2.1. *Theoretical results.* Analogously to Lemma 3.1.3 and Theorem 3.1.1, we have

Lemma 3.2.1. *Let V^ε be given by (3.3) such that $f, g \in \mathcal{C}(\mathbb{R}^+ \rightarrow \mathbb{R})$ satisfy Assumption 1.4.1. Then $\{V^\varepsilon\}$ converges in probability to v_0 .*

Theorem 3.2.1. *Let $v_0 > 0$ and let $f, g \in \mathcal{C}(\mathbb{R}^+ \rightarrow \mathbb{R})$ be such that Assumption 1.4.1 is satisfied and let $W^\varepsilon := \sqrt{\varepsilon}W$ and $V^{\varepsilon, \beta}, \eta^\varepsilon$ defined as in (3.3), then, under Assumption 3.0.1, the triple $\{V^{\varepsilon, \beta}, \eta^\varepsilon, W^\varepsilon\}$ satisfies an LDP with speed ε and the good rate function*

$$I^{V, \eta, W}(v, \eta, w) = \begin{cases} \frac{1}{2} \|\dot{w}\|_T^2, & \text{if } w \in \mathbb{H}_T^0, v = \bar{\psi}, \dot{\eta} = g(\bar{\psi})\dot{w}, \\ +\infty, & \text{otherwise.} \end{cases}$$

3.3. **Large-time moderate deviations.** We now consider a rescaling of (1.6) defined as $V_t^\varepsilon = V_{t/\varepsilon}$ and $X_t^\varepsilon = \varepsilon X_{t/\varepsilon}$, so that under Assumption 1.4.1 and Assumption 3.3.1,

$$(3.7) \quad \begin{cases} dX_t^\varepsilon = -\frac{1}{2}V_t^\varepsilon dt + \sqrt{\varepsilon}\sqrt{V_t^\varepsilon}dB_t, & X_0^\varepsilon = 0, \\ dV_t^\varepsilon = \frac{1}{\varepsilon}f(V_t^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}}g(V_t^\varepsilon)dW_t, & V_0^\varepsilon = v_0 > 0, \end{cases}$$

which leads to Regime 2 in the slow-fast setting of [33, Theorem 2.1], by choosing the time-scale separation parameter equal to ε . The following assumption is needed in order to conform to the conditions in [33].

Assumption 3.3.1.

- (i) f is locally bounded and of the form $f(y) = -\kappa y + \tau(y)$ with τ globally Lipschitz with Lipschitz constant $L_\tau < \kappa$. In addition, the tail condition $\lim_{|y| \uparrow \infty} \frac{\tau(y)y}{|y|^2} = \kappa$ holds.
- (ii) The function g is either uniformly continuous and bounded from above and away from zero or takes the form $g(y) = \xi|y|^{q_g}$ for $q_g \in [\frac{1}{2}, 1)$ with $\xi \neq 0$.

Remark 3.3.1. Together with Assumption 1.4.1, Condition (ii), necessary to ensure ergodicity of the volatility process, collapses g to the form $g(y) = \xi|y|^{\frac{1}{2}}$ (for details we refer to [27, 33]).

3.3.1. *Theoretical results.* In order to apply the methodology from the previous sections to derive the desired changes of measure, we need a large-time MDP. More precisely, we need an MDP for $\{X^\varepsilon, \sqrt{\varepsilon}W, \sqrt{\varepsilon}W^\perp\}$ in the case of deterministic drift. We do not consider the stochastic drift change here, since a rigorous treatment is out of scope of this paper. Similar problem has been studied in [33, Theorem 2.1] and [27, Theorem 3.3], where authors propose fewer conditions, although in a simpler setting (which happens to include the Heston model as well). We now introduce a theorem, which is a direct application of [27, Theorem 3.3] and [33, Theorem 2.1], that provides the desired MDP.

Theorem 3.3.1. *Let \mathcal{L}_V denote the infinitesimal generator of V before rescaling, i.e.*

$$\mathcal{L}_V h = f\dot{h} + \frac{1}{2}g^2\ddot{h}.$$

Under Assumption 1.4.1 and Assumption 3.3.1 the following hold:

- i) *There exists a unique invariant measure μ corresponding to \mathcal{L}_V ;*
- ii) *The process X^ε converges in probability to $-\frac{1}{2}\bar{v}t$, where $\bar{v} = \int_0^\infty y \mu(dy)$;*
- iii) *There exists a unique solution ς with at most polynomial growth to the Poisson equation*

$$\mathcal{L}_V(\varsigma)(y) = \frac{y - \bar{v}}{2}, \quad \text{with} \quad \int_0^\infty \varsigma(y) \mu(dy) = 0.$$

Furthermore, denote $\mathbf{Q} := \int_0^\infty \mathbf{q}(y) \mathbf{q}^\top(y) \mu(dy)$, where

$$\mathbf{q}(y) = \begin{pmatrix} \rho\sqrt{y} + \varsigma(y)g(y) & \bar{\rho}\sqrt{y} \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then the triple $\{X^\varepsilon, \sqrt{\varepsilon}W, \sqrt{\varepsilon}W^\perp\}$ follows an MDP with good rate function $I_{\mathbf{Q}}$, where

$$I_{\mathbf{Q}}(\phi) := \inf \left\{ \frac{1}{2} \int_0^T u_s^\top u_s ds : u \in L^2([0, T]; \mathbb{R}^3), \dot{\phi}^\top \mathbf{Q}^{-1} \dot{\phi} = u^\top u \right\},$$

if $\phi \in \mathcal{AC}$ and infinite otherwise.

Lemma 3.3.1. *The matrix \mathbf{Q} is invertible.*

Proof. Let Y a random variable with distribution μ , the invariant measure. Then

$$\begin{aligned} \det(\mathbf{Q}) &= \mathbb{E}_\mu \left[\left(\rho\sqrt{Y} + \dot{\zeta}(Y)g(Y) \right)^2 + \bar{\rho}^2 Y \right] - \mathbb{E}_\mu \left[\rho\sqrt{Y} + \dot{\zeta}(Y)g(Y) \right]^2 - \mathbb{E}_\mu \left[\bar{\rho}\sqrt{Y} \right]^2 \\ &= \mathbb{E}_\mu \left[\left| \rho\sqrt{Y} + \dot{\zeta}(Y)g(Y) \right|^2 \right] - \mathbb{E}_\mu \left[\rho\sqrt{Y} + \dot{\zeta}(Y)g(Y) \right]^2 + \bar{\rho}^2 \left\{ \mathbb{E}_\mu[Y] - \mathbb{E}_\mu \left[\sqrt{Y} \right]^2 \right\} \geq 0, \end{aligned}$$

by Cauchy-Schwarz, where an equality would imply that V is constant μ -almost surely. This implies $f(\bar{v}) = g(v) = 0$, which is not possible due to Assumption 1.4.1 on f and g . \square

3.3.2. Application to options with path-dependent payoff. Let us again consider an option with a continuous payoff function G and let $F = \log |G|$. Following the approach of the previous sections and with similar notations, we search solutions to the dual problems for deterministic change of drift. Let $\mathbf{x} := (x_1 \ x_2 \ x_3)^\top$ and $\underline{\mathbf{x}} := (x_2 \ x_3)^\top$. By considering a deterministic change of drift similarly as before our problem becomes

$$L(\mathbf{h}) = \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \frac{1}{\varepsilon} \left\{ 2F(X^\varepsilon) + \frac{1}{2} \|\dot{\mathbf{h}}\|_T^2 - \dot{\mathbf{h}} \circ_T (\mathbf{W}^\varepsilon)^\top \right\} \right].$$

Applying the modified Varadhan lemma gives us the target functional

$$\mathfrak{L}(\mathbf{x}, \mathbf{h}) = 2F(x_1) + \frac{1}{2} \|\dot{\mathbf{h}}\|_T^2 - \dot{\mathbf{h}} \circ \underline{\mathbf{x}}^\top - \frac{1}{2} \int_0^T \dot{\mathbf{x}}^\top \mathbf{Q}^{-1} \dot{\mathbf{x}} dt.$$

Using the diagonal operator $\text{diag}(x_1, \dots, x_n) = \sum_{i=1}^n x_i e_i e_i^\top$, where $n \in \mathbb{N}$ and $\{e_i\}_{i=1}^n$ canonical basis of \mathbb{R}^n , we now write the problem in terms of $\tilde{\mathbf{x}} := \dot{\mathbf{x}}$,

$$\begin{cases} \tilde{\mathbf{x}}^* = \arg \max_{\tilde{\mathbf{x}} \in L^2([0, T]; \mathbb{R}^3)} \left\{ F \left(\int_0^\cdot \tilde{x}_1(t) dt \right) - \frac{1}{4} \int_0^T \tilde{\mathbf{x}}^\top(t) \mathbf{A} \tilde{\mathbf{x}}(t) dt \right\}, \\ \dot{\mathbf{h}} = \tilde{\mathbf{x}}^*, \end{cases}$$

where $\mathbf{A} := \mathbf{Q}^{-1} + \text{diag}(0, 1, 1)$, which is equivalent to

$$(3.8) \quad \sup_{x_1 \in L^2([0, T]; \mathbb{R})} \left\{ F \left(\int_0^\cdot x_1(t) dt \right) - \frac{1}{4} \int_0^T \inf_{\mathbf{x} \in L^2([0, T]; \mathbb{R}^2)} \mathbf{x}^\top(t) \mathbf{A} \mathbf{x}(t) dt \right\}.$$

This substantially simplifies the main optimisation problem, which can now be solved explicitly, since $\mathbf{A} = \begin{pmatrix} a_{11} & \mathbf{a}_{21}^\top \\ \mathbf{a}_{21} & \mathbf{A}_{22} \end{pmatrix}$, with $a_{11} \in \mathbb{R}$, $\mathbf{a}_{21} \in \mathbb{R}^2$, $\mathbf{A}_{22} \in \mathbb{R}^{2 \times 2}$, so that the infimum in (3.8) yields¹ $\underline{\mathbf{x}}^* = -\mathbf{A}_{22}^{-1} \mathbf{a}_{21} x_1 =: \mathbf{B} x_1$. The quadratic form in fact reduces to $\mathbf{x}^\top \mathbf{A} \mathbf{x} = (a_{11} - \mathbf{a}_{21}^\top \mathbf{A}_{22}^{-1} \mathbf{a}_{21}) x_1^2 := \nu x_1^2$, where both constants

(3.9)

$$\begin{aligned} \nu &= \mathbb{E}_\mu \left[\left| \rho\sqrt{Y} + \dot{\zeta}(Y)g(Y) \right|^2 \right] + \bar{\rho}^2 \mathbb{E}_\mu[Y] - \frac{\mathbb{E}_\mu \left[\rho\sqrt{Y} + \dot{\zeta}(Y)g(Y) \right]^2}{2} - \frac{\bar{\rho} \mathbb{E}_\mu[\sqrt{Y}]^2}{2}, \\ \mathbf{B} &= -\frac{1}{2} \begin{pmatrix} \mathbb{E}_\mu[\rho\sqrt{Y} + \dot{\zeta}(Y)g(Y)] \\ \bar{\rho} \mathbb{E}_\mu[\sqrt{Y}] \end{pmatrix}, \end{aligned}$$

are obtained from the definition of \mathbf{Q} . Finally, we have

$$(3.10) \quad \begin{cases} x_1^* = \arg \max_{x_1 \in L^2([0, T], \mathbb{R})} \left\{ F \left(\int_0^\cdot x_1 dt \right) - \frac{\nu}{4} \int_0^T x_1^2 dt \right\}, \\ \dot{\mathbf{h}}^* = \mathbf{B} x_1^*. \end{cases}$$

¹Since \mathbf{Q} is positive-definite, both \mathbf{A} and \mathbf{A}_{22} are positive-definite and thus invertible.

Remark 3.3.2. *This problem is similar to the problem in Section 4.1, with deterministic volatility.*

Moderate deviations, though less studied than large deviations in mathematical finance, have been applied to a wide variety of settings beyond Heston, predominantly in multiscale systems [33, 34]. A unifying treatment of pathwise moderate deviations for Markovian models commonly used in financial applications can be consulted in [27], and further developments in the context of (rough) fractional volatility models were studied in [2] and in [26].

4. EXAMPLES

Our setting allows for a variety of stochastic volatility models. We show some examples below, and give special attention to the Heston model [24], the focus of our numerical experiments.

4.1. Black-Scholes with deterministic volatility. We first consider the Black-Scholes model with time-dependent deterministic volatility, namely

$$dX_t = -\frac{1}{2}\sigma^2(t)dt + \sigma(t)dW_t, \quad X_0 = 0,$$

where $\sigma \in \mathcal{C}_T$ is a deterministic function adhering to Assumption 1.4.1. We provide full details of this specific case, which we will often refer to later on. We consider options with payoffs that are continuous functions of X . Let $G : \mathcal{C}_T \rightarrow \mathbb{R}^+$ and $F := \log |G|$, which is then a continuous function with values in $[-\infty, +\infty)$. Our objective is to find an almost surely positive random variable Z with $\mathbb{E}[Z] = 1$ minimising $\mathbb{E}[G^2(X)Z^{-1}] = \mathbb{E}[e^{2F(X)}Z^{-1}]$.

4.1.1. Small-noise. The small-noise approximation (1.3) reads

$$dX_t^\varepsilon = -\frac{1}{2}\sigma^2(t)dt + \sqrt{\varepsilon}\sigma(t)dW_t, \quad X_0^\varepsilon = 0.$$

Deterministic change of drift. With the change of measure

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_T} := \exp \left\{ -\frac{1}{2} \int_0^T \dot{h}_t^2 dt + \int_0^T \dot{h}_t dW_t \right\},$$

with $h \in \mathbb{H}_T^0$, we can write

$$L(h) := \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \frac{1}{\varepsilon} \left\{ 2F(X^\varepsilon) + \frac{1}{2} \int_0^T \dot{h}_t^2 dt - \dot{h} \circ_T W^\varepsilon \right\} \right].$$

To apply Varadhan's lemma and estimate $L(h)$, we assume there exists $\gamma > 1$ such that

$$(4.1) \quad \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ 4\gamma \frac{F(X^\varepsilon)}{\varepsilon} \right\} \right] < \infty.$$

This condition is satisfied if, for example, the following assumption holds (for details see the proof of [20, Theorem 3.6]):

Assumption 4.1.1. *Let $F : \mathcal{C}_T \rightarrow \mathbb{R}$ such that $F(x) \leq K_1 + K_2 \sup_{t \in [0, T]} |x_t|^\alpha$ for some $\alpha \in (0, 2)$ and $K_1, K_2 > 0$.*

Lemma 4.1.1. *If \dot{h} is of finite variation, the function*

$$\mathfrak{L}(\cdot; h) : x \in \mathcal{C}_T \mapsto 2F \left(-\frac{1}{2} \int_0^\cdot \sigma^2(t)dt + \int_0^\cdot \sigma(t)dx_t \right) + \frac{1}{2} \left(\int_0^T \dot{h}_t^2 dt - \int_0^T \dot{x}_t^2 dt \right) - \dot{h} \circ_T x$$

is well-defined and continuous and for every $\gamma > 0$,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{\gamma}{\varepsilon} \left(\frac{1}{2} \int_0^T \dot{h}_t^2 dt - \dot{h} \circ_T W^\varepsilon \right) \right\} \right] \quad \text{exists in } \mathbb{R}.$$

Proof. If \dot{h} is of finite variation and x is continuous with $x(0) = 0$, then $\int_0^\cdot \dot{h}_t dx_t$ is well defined as a Riemann-Stieltjes integral and

$$\left| \dot{h} \circ_T x \right| = \left| \dot{h}(T)x(T) - x \circ_T \dot{h} \right| \leq \|x\|_\infty \left(|\dot{h}(T)| + \text{TV}(\dot{h}) \right),$$

where $\text{TV}(\cdot)$ denotes the total variation. Similarly, $\int_0^\cdot \sigma(t) dx_t$ is well defined and

$$|\sigma(\cdot) \circ_t x| \leq \|x\|_\infty (\|\sigma\|_\infty + \text{TV}(\sigma)), \quad \text{for all } t \in [0, T].$$

Since F is continuous, the first statement on existence and continuity holds. The second one follows from the computation of exponential Gaussian moments: for every $\gamma > 0$,

$$\varepsilon \log \mathbb{E} \left[\exp \frac{\gamma}{\varepsilon} \left\{ \frac{1}{2} \int_0^T \dot{h}_t^2 dt - \dot{h} \circ_T W^\varepsilon \right\} \right] = \frac{\gamma(1+\gamma)}{2} \int_0^T \dot{h}_t^2 dt.$$

□

Lemma 4.1.2. *If \dot{h} is of finite variation, then $L(h) = \sup_{x \in \mathbb{H}_T^0} \mathfrak{L}(x; h)$.*

Proof. Let $\gamma > 1$ such that (4.1) holds. Then by Cauchy-Schwarz and Lemma 4.1.1,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \frac{\gamma}{\varepsilon} \left\{ 2F \left(-\frac{1}{2} \int_0^\cdot \sigma^2(t) dt + \int_0^\cdot \sigma(t) dW_t^\varepsilon \right) + \frac{1}{2} \int_0^T \dot{h}_t^2 dt - \int_0^T \dot{h}_t dW_t^\varepsilon \right\} \right] < \infty,$$

so that the conditions of Theorem B.0.1 are verified. The continuity has already been shown. □

Since $\mathcal{G} : W \mapsto X$ is continuous by continuity of the Itô integral, we can introduce $\tilde{F} = F \circ \mathcal{G}$ and the existence of a minimum to the dual version of $\inf_{h \in \mathbb{H}_T^0} \sup_{x \in \mathbb{H}_T^0} \mathfrak{L}(x; h)$, namely

$\sup_{x \in \mathbb{H}_T^0} \inf_{h \in \mathbb{H}_T^0} \mathfrak{L}(x; h)$, can be proved as in [20] under Assumption 4.1.1 by choosing $M = 0$ in [20,

Lemma 7.1]. The minimum is then attained for $h = x$ and equal to

$$(4.2) \quad \sup_{x \in \mathbb{H}_T^0} 2F \left(-\frac{1}{2} \int_0^\cdot \sigma^2(t) dt + \int_0^\cdot \sigma(t) \dot{x}_t dt \right) - \int_0^T \dot{x}_t^2 dt.$$

Furthermore, it immediately follows from [20, Theorem 3.6] that, if $h^* \in \mathbb{H}_T^0$ is of finite variation and is a solution to (4.2), then it is asymptotically optimal if

$$L(h^*) = 2F \left(-\frac{1}{2} \int_0^\cdot \sigma^2(t) dt + \int_0^\cdot \sigma(t) \dot{h}_t^* dt \right) - \int_0^T |\dot{h}_t^*|^2 dt.$$

Therefore, in order to derive a change of measure, we search for $h^* \in \mathbb{H}_T^0$ such that

$$h^* = \arg \max_{x \in \mathbb{H}_T^0} F \left(-\frac{1}{2} \int_0^\cdot \sigma^2(t) dt + \int_0^\cdot \sigma(t) \dot{x}_t dt \right) - \frac{1}{2} \int_0^T \dot{x}_t^2 dt.$$

Simplified deterministic change of drift. We consider a simplified version of the problem: since $\varepsilon \int_0^\cdot \sigma^2(t) dt$ is exponentially equivalent to 0, the previous results remain valid when replacing $F(-\frac{1}{2} \int_0^\cdot \sigma^2(t) dt + \int_0^\cdot \sigma(t) \dot{x}_t dt)$ with $F(\int_0^\cdot \sigma(t) \dot{x}_t dt)$. The problem then becomes

$$h^* = \arg \max_{x \in \mathbb{H}_T^0} F \left(\int_0^\cdot \sigma(t) \dot{x}_t dt \right) - \frac{1}{2} \int_0^T \dot{x}_t^2 dt.$$

4.1.2. Application to options with path-dependent payoff. Consider the payoff $G(\vartheta \circ_T X)$ with $G : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ differentiable and ϑ a positive continuous function of finite variation. The optimisation problem reads

$$h^* = \arg \max_{x \in \mathbb{H}_T^0} F \left(\int_0^T \vartheta_t \sigma(t) \dot{x}_t dt \right) - \frac{1}{2} \int_0^T \dot{x}_t^2 dt,$$

with $F = \log |G(\cdot - \frac{1}{2} \int_0^T \vartheta_t \sigma^2(t) dt)|$ in the log-price small-noise approximation. The mapping $\dot{y}_t = \vartheta_t \sigma(t) \dot{x}_t$ yields the Euler-Lagrange equation $\frac{d}{dt} \frac{\dot{y}_t}{(\vartheta_t \sigma(t))^2} = 0$, or in terms of the original

problem $\frac{d}{dt} \frac{\dot{x}_t}{\vartheta_t \sigma(t)} = 0$. We thus search for a solution $\dot{x}_t = \lambda \vartheta_t \sigma(t)$ for some $\lambda \in \mathbb{R}$. The optimisation problem becomes

$$\begin{cases} \lambda^* &= \arg \max_{\lambda \in \mathbb{R}} \left\{ F \left(\lambda \int_0^T |\vartheta_t \sigma(t)|^2 dt \right) - \frac{1}{2} \lambda^2 \int_0^T |\vartheta_t \sigma(t)|^2 dt \right\}, \\ h^* &= \lambda^* \int_0^T \vartheta_t \sigma(t) dt. \end{cases}$$

For example, if $F(x) = \log(e^x - e^c)^+$ for $x, c \in \mathbb{R}$, $\tilde{\lambda}$ will be the unique solution on $(1, \infty)$ of $v\lambda + \log(\lambda - 1) - \log(\lambda) - c = 0$. The following payoffs are standard and satisfy this setup:

- (European Call option): $G(x) = (e^x - K)^+$ and $\vartheta_t = 1$ with $x^+ = \max(0, x)$.
- (Geometric Asian Call option): $G(x) = (e^x - K)^+$ and $\vartheta_t = 1 - t/T$.

4.1.3. *Small-time.* The small-time approximation (1.4) reads

$$dX_t^\varepsilon = -\frac{1}{2} \varepsilon \sigma^2(\varepsilon t) dt + \sigma(t\varepsilon) dW_t^\varepsilon,$$

with $X_0^\varepsilon = 0$ and $W^\varepsilon = \sqrt{\varepsilon} W$. The couple $\{\sigma(t\varepsilon), W_t^\varepsilon\}$ is exponentially equivalent to $\{\sigma(0), W_t^\varepsilon\}$. Since I^W is the good rate function of the LDP verified by $\{W^\varepsilon\}$ by Schilder's Theorem [8, Theorem 5.2.3], $\{\sigma(t\varepsilon), W_t^\varepsilon\}$ verifies an LDP with good rate function $I(s, w) = I^W(w)$ if $s = \sigma(0)$ and is infinite otherwise. By Theorem 2.1.3, $\{\int_0^\cdot \sigma(t\varepsilon) dW_t^\varepsilon, W^\varepsilon\}$ satisfies an LDP with the same good rate function as $\{\int_0^\cdot \sigma(0) dW_t^\varepsilon, W^\varepsilon\}$. Noticing that $\int_0^\cdot \varepsilon \sigma(t\varepsilon)^2 dt$ is exponentially equivalent to zero, our method then leads to the same solution as the problem for $dX_t^\varepsilon = -\frac{1}{2} \varepsilon \sigma^2(0) dt + \sigma(0) dW_t^\varepsilon$, which was treated above. In this small-time setting, we lose all information on the path of σ , except for its initial value.

4.2. **Ornstein-Uhlenbeck.** Lognormal-based stochastic volatility models, such as [39], read

$$(4.3) \quad \begin{aligned} dX_t &= -\frac{1}{2} V_t + \sqrt{V_t} dB_t, & X_0 &= 0, \\ dV_t &= \kappa(\theta - V_t) dt + \xi V_t dW_t, & V_0 &= v_0 > 0. \end{aligned}$$

This model clearly satisfies all the assumptions in the previous sections (linear drift and diffusion) and therefore the importance sampling results can be applied.

4.3. **Rough volatility.** Our setup above clearly does not include rough volatility models [15], where the instantaneous variance process is driven by a continuous Gaussian process with Hurst index in $(0, \frac{1}{2})$. However, recent developments have proved, with various degrees of generality, pathwise large and moderate deviations for this class of models [16, 21, 22, 25, 26], and our general results above can definitely be tailored to them. We however leave this model-specific analysis to future endeavours.

4.4. **The Heston model.** The Heston model is one of the most widely used stochastic volatility model in Finance, first proposed in [24], in which the log stock price process satisfies

$$(4.4) \quad \begin{aligned} dX_t &= -\frac{1}{2} V_t + \sqrt{V_t} dB_t, & X_0 &= 0, \\ dV_t &= \kappa(\theta - V_t) dt + \xi \sqrt{V_t} dW_t, & V_0 &= v_0 > 0, \end{aligned}$$

for two Brownian motions B and W such that $d\langle B, W \rangle_t = \rho dt$. The coefficient $\kappa > 0$ is the speed of mean reversion of the volatility process, while $\theta > 0$ is the long-term mean and $\xi > 0$ is the volatility of the volatility process. The variance process V admits a unique strong solution by Yamada-Watanabe's conditions [28, Proposition 2.13], and so does X . In the setting of (1.6) the functions f and g take the form

$$f(v) = \kappa(\theta - v) \quad \text{and} \quad g(v) = \xi \sqrt{v},$$

and therefore Assumption 1.4.1 clearly holds.

4.4.1. *Small-noise LDP.* From Section 2.1, Assumption 2.1.2 is clearly satisfied, and thus Theorem 1.4.1 holds. In the Heston model, the system (2.5) becomes

$$\begin{cases} \frac{d}{dt} \left\{ \frac{Z - \rho U}{\vartheta \bar{\rho}^2 \sqrt{\psi}} \right\} = 0, \\ \frac{d}{dt} \left\{ \frac{U}{\xi \sqrt{\psi}} - \frac{\rho}{\bar{\rho}^2} \frac{(Z - \rho U)}{\xi \sqrt{\psi}} \right\} = U \left(-\frac{U}{2\psi} + \frac{\kappa}{\xi \sqrt{\psi}} \right) + \frac{Z - \rho U}{\bar{\rho}^2} \left(-\frac{Z - \rho U}{2\psi} + \frac{\frac{1}{2} - \rho \frac{\kappa}{\xi}}{\sqrt{\psi}} \right). \end{cases}$$

Solving the first ODE and plugging it into the second one gives for all $t \in [0, T]$

$$(4.5) \quad \begin{cases} \frac{Z_t - \rho U_t}{\bar{\rho}^2} = \frac{1}{2} \lambda \vartheta_t \psi_t, \\ \dot{A}_t = -\frac{1}{2} \xi A_t^2 + \kappa A_t + \frac{1}{2} \xi \lambda \vartheta_t \left(\frac{1}{2} - \frac{1}{4} \bar{\rho}^2 \lambda \vartheta_t - \rho \frac{\kappa}{\xi} \right) + \frac{1}{2} \rho \lambda \vartheta_t, \end{cases}$$

where $\lambda > 0$ is an arbitrary constant and $A_t = \frac{U_t}{\sqrt{\psi_t}}$ is the solution to the Riccati equation. The option payoff at the terminal time determines the boundary conditions through the Euler-Lagrange equations so that

$$\lambda = -\frac{2F'_T}{\sqrt{\psi_T}} \quad \text{and} \quad A_T = -\frac{\rho}{\xi} F'_T \vartheta_T,$$

with $F_T = F \left(\int_0^T \vartheta_t (Z \sqrt{\psi} - \frac{1}{2} \psi) dt \right)$. Since both conditions include the same optimising variable, the resulting problem becomes an optimisation in \mathbb{R}^2 over λ and A_T (or equivalently A_0) and is thus much simpler than the original optimisation. After solving for A_t for all $t \in [0, T]$, we then solve for (φ, ψ) by writing U and Z in terms of it using (2.4).

Example 4.4.1. In the case $\vartheta_t = \vartheta > 0$ constant, let $C(\lambda) := \frac{1}{2} \xi \lambda \vartheta \left(\frac{1}{2} - \frac{1}{4} \bar{\rho}^2 \lambda \vartheta - \rho \frac{\kappa}{\xi} \right)$ so that the Riccati equation (4.5) reads

$$\frac{dA_t}{dt} = -\frac{\xi}{2} A_t^2 + \kappa A_t + C(\lambda),$$

the solution to which is

$$A_t = \frac{1}{\xi} \left\{ \sqrt{2\xi C(\lambda) - \kappa^2} \tan \left(-\frac{1}{2} \frac{t - D(A_0, \lambda)}{\sqrt{2\xi C(\lambda) - \kappa^2}} \right) + \kappa \right\},$$

where $D \in \mathbb{R}$ is determined from the initial condition on A . Since $A_t = \frac{U_t}{\sqrt{\psi_t}} = \frac{\dot{\psi}_t - \kappa(\theta - \psi_t)}{\psi_t}$, we are to solve $\dot{\psi}_t + (\kappa - A_t)\psi_t = \kappa\theta$, for $t \in [0, T]$, which is just a non-homogeneous linear ODE with the solution

$$\psi_t = \left(\kappa\theta \int_0^t \exp \left\{ \int_0^s (\kappa - A_u) du \right\} ds + v_0 \right) \exp \left\{ -\int_0^t (\kappa - A_s) ds \right\}.$$

Then the optimisation problem for \mathbf{h} reduces to

$$\begin{cases} h_1(t) = U(\psi^*) & = \min_{A_0, \lambda > 0} \frac{\dot{\psi}_t(A_0, \lambda) - \kappa(\theta - \psi_t(A_0, \lambda))}{\psi_t(A_0, \lambda)} \\ h_2(t) = \frac{Z(\varphi^*, \psi^*) - \rho U(\psi^*)}{\bar{\rho}} & = \frac{\vartheta}{2} \min_{A_0, \lambda > 0} \lambda \psi_t(A_0, \lambda), \end{cases}$$

which is an optimisation over $(A_0, \lambda) \in \mathbb{R}^2$.

4.4.2. *Small-time LDP.* In the context of the small-time LDP, Assumption 2.2.1 is clearly satisfied and therefore Theorem 2.2.1 applies immediately. Consider for example an option with path-dependent payoff $G(\vartheta \circ_T X)$ as in Section 2.1.3. In the deterministic case, we have

$$\begin{cases} \varphi^*, \psi^* = \arg \max_{(\varphi, \psi) \in \mathbb{H}_T^0 \times \mathbb{H}_T^{0,+}} \left\{ F \left(\int_0^T \vartheta_t \dot{\varphi}_t dt \right) - \frac{1}{2} \int_0^T \left\{ U(\psi_t)^2 + \left| \frac{Z(\varphi_t, \psi_t) - \rho U(\psi_t)}{\bar{\rho}} \right|^2 \right\} dt \right\}, \\ \mathbf{h}^* = \left(U(\psi^*), \frac{Z(\varphi^*, \psi^*) - \rho U(\psi^*)}{\bar{\rho}} \right), \end{cases}$$

where $U(\psi) = \frac{\dot{\psi}}{\xi\sqrt{\psi}}$ and $Z(\varphi, \psi) = \frac{\dot{\varphi}}{\sqrt{\psi}}$. Similarly to Example 4.4.1, with $A_t = \frac{U_t}{\sqrt{\psi_t}}$,

$$\begin{cases} Z_t - \rho U_t &= \frac{1}{2}\bar{\rho}^2 \lambda \vartheta_t \psi_t, \\ \dot{A}_t &= -\frac{1}{2}\xi A_t^2 + \frac{1}{2}\xi \lambda \vartheta_t \left(\frac{1}{2} - \frac{1}{4}\bar{\rho}^2 \lambda \vartheta_t \right) + \frac{1}{2}\rho \lambda \vartheta_t, \end{cases}$$

for all $t \in [0, T]$, which is very similar to (4.5), but for the κ term.

4.4.3. Large-time moderate deviations. In the context of Section 3.3, the conditions of Assumption 3.0.1 are clearly satisfied. In this case, the invariant measure μ is a Gamma distribution $\Gamma\left(\frac{2\kappa\theta}{\xi^2}, \frac{2\kappa}{\xi^2}\right)$ as shown in [7, Section 3.]. Therefore, in light of Theorem 3.3.1, $\bar{v} = \theta$ and the solution of the Poisson equation $\dot{\varsigma}(y) = -\frac{1}{2\kappa}$ is constant, which means that $\varsigma(y) = (\theta - y)/(2\kappa)$ because of the constraint $\int_0^\infty \varsigma(y)\mu(dy) = 0$. Therefore

$$\mathbb{E}_\mu[Y] = \theta \quad \text{and} \quad \mathbb{E}_\mu[\sqrt{Y}] = \frac{\Gamma\left(\frac{2\kappa}{\xi^2}\theta + \frac{1}{2}\right)}{\Gamma\left(\frac{2\kappa}{\xi^2}\right)} \frac{\xi}{\sqrt{2\kappa}}.$$

From this, noting that $\dot{\varsigma}(y)g(y) = -\xi\sqrt{y}/(2\kappa)$ we can calculate the constants (3.9):

$$\nu = \left(\left(\rho - \frac{\xi}{2\kappa} \right)^2 + 1 - \rho^2 \right) \left(\theta - \mathbb{E}_\mu[\sqrt{Y}] \right) \quad \text{and} \quad B = -\frac{1}{2} \left(\rho - \frac{\xi}{2\kappa} \right) \mathbb{E}_\mu[\sqrt{Y}].$$

4.4.4. Other. Results in Heston which are not mentioned in this section follow directly from their corresponding optimisation problems in Sections 2 and 3 by choosing f and g as in (4.4).

5. NUMERICAL RESULTS

In all the different settings we studied, the final form of the optimisation problem is

$$\sup_{(\varphi, \phi) \in \mathbb{H}_T^0 \times \mathbb{H}_T^{v0}} F(\varphi) - \frac{1}{2} \int_0^T \ell(\varphi_t, \phi_t) dt,$$

where F was linked to the payoff, ℓ to the rate function of the (log-)price process and φ and ϕ are absolutely continuous paths that arise from Varadhan's lemma of the (log-)price and volatility processes respectively. In the tables below, we summarise all problems considered so far. As one can see, in the deterministic drift setting, the MDP problem is usually as simple as solving the problem under the Black-Scholes (BS) model or at least by approximating the model with a Black-Scholes model. Furthermore, the variance reduction for geometric Asian options are also similar under the MDP with deterministic drift change, meaning advantage over a simple BS model is not significant. However, when it comes to the stochastic change of drift of the form $\int_0^T \dot{h}_t \sqrt{V_t} dt$, the MDP problems are slightly harder than in BS approximation and the variance reduction results are in fact significantly better. [We study stochastic change of drift in all cases except in the MDP large-time case, where the invariant distribution of \$\int_0^T \dot{h}_t \sqrt{V_t} dt\$ is not immediate and would require a careful study, which is outside the scope of this paper.](#)

| Method used | Optimisation with general payoff | Optimisation with payoff $\vartheta \circ_T X$ |
|---|--|---|
| Deterministic volatility (BS) approximation | Optimisation in \mathbb{H}_T^0 with ℓ a quadratic function of the (log-)price process φ | Optimisation on \mathbb{R} with simple function to optimise |
| LDP small-(noise/time) | System of ODEs | Optimisation in \mathbb{R}^2 + solving a Riccati equation |
| MDP small-noise | Simple optimisation; the ODE can be precomputed | Straightforward optimisation in \mathbb{R}^2 |
| MDP small-time | Similar complexity to the BS case | Similar complexity to the BS case |
| MDP large-time | Similar complexity to the BS case | Similar complexity to the BS case |

TABLE 1. Summary of the optimisations with *deterministic* change of drift.

| Method used | Optimisation with general payoff | Optimisation with payoff $\vartheta \circ_T X$ |
|--|--|---|
| Deterministic volatility approximation (BS) ^a | Optimisation on \mathbb{H}_T^0 with ℓ a quadratic function of φ | Optimisation on \mathbb{R} with simple function to optimise |
| LDP small-(noise/time) | Same as in the deterministic case with a simple additional ODE | Optimisation in \mathbb{R}^2 and solving a Riccati equation |
| MDP small-(noise/time) | Similar as in the deterministic case with an additional ODE | Straightforward optimisation in \mathbb{R}^2 |

^aHere h_t in the drift change $\int \dot{h}_t \sqrt{V_t} dt$ is computed using the BS approximation at time t .

TABLE 2. Summary of the optimisations with *stochastic* change of drift.

5.1. Numerical results for Black-Scholes. We first provide numerical evidence in the Black-Scholes model (Section 4.1) with $S_0 = 50$ in the log-price small-noise approximation. In order to compare estimators, we look at Asian Arithmetic Call options, with payoff $(\frac{1}{T} \int_0^T \exp\{X_t\} dt - K)^+$. The form of the solution of the optimisation problem studied previously can be found in [20]. **We compare the naive Monte-Carlo estimator to:**

- antithetic Monte-Carlo estimator;
- control estimator based on the price of Geometric Asian options with payoff $(\exp\{\frac{1}{T} \int_0^T X_t dt\} - K)^+$, that can be computed explicitly;
- LDP-based importance sampling estimator above.

Instead of simulating $\int_0^T \exp(X_t) dt$, we consider a discretised payoff on $n = 252$ dates and draw 10^5 paths. For the LDP-based estimator, the law of W after the change of measure is given by Girsanov theorem. In what follows, when we refer to variance reduction we mean the ratio of variance of the classical Monte-Carlo estimator over the variance of estimator in question. As we can see in Table 3 and in Figure 2, even in non-rare events, the estimator derived using LDP provides good variance reduction. However, it is mainly in the context of rare events that it performs best and outperforms the other estimators (Figure 1 and Table 3), revealing the true power of LDP-based importance sampling estimators.

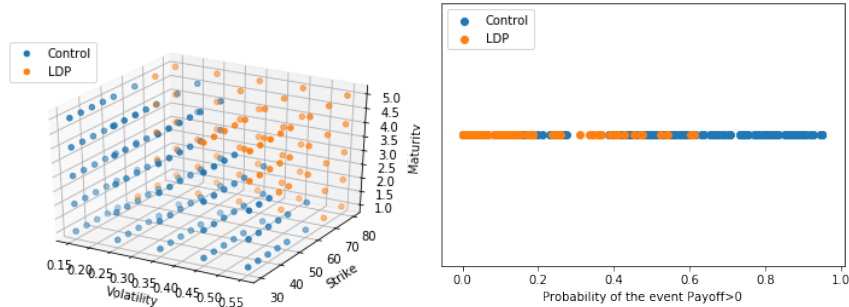


FIGURE 1. Left: Estimator with the highest variance reduction among antithetic, control and LDP estimators (Section 5.1) for different values of σ , K , T . Right: Best performing estimator in terms of variance reduction among antithetic, control and LDP estimators (Section 5.1) as a function of probability of a positive payoff (estimated using Monte-Carlo). Clearly, the LDP estimator performs best when probability of exercising an option is low.

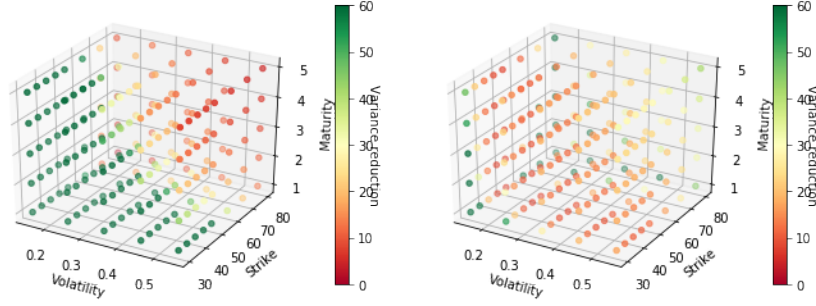


FIGURE 2. Variance reduction of the control (left) and the LDP (right) estimators

| Strike K | Antithetic | Control | LDP | Probability of positive Payoff |
|------------|------------|---------|------|--------------------------------|
| 30 | 64 | 769 | 53 | 0.95 |
| 35 | 59 | 775 | 21 | 0.95 |
| 40 | 31 | 744 | 10 | 0.9 |
| 45 | 10 | 575 | 7.9 | 0.75 |
| 50 | 3.8 | 336 | 8.6 | 0.5 |
| 60 | 2.2 | 69 | 22 | 0.11 |
| 70 | 2.0 | 16 | 123 | 0.013 |
| 80 | 2.3 | 6.9 | 1445 | 0.0011 |

TABLE 3. Variance reduction for several estimators for Arithmetic Call options in Black-Scholes with $S_0 = 50$, $r = 0.05$, $\sigma = 0.25$ and $T = 1$.

5.2. Numerical results for the Heston model.

5.2.1. *Asian option in Heston.* To compare variance reduction results in the Heston model, we look at Asian Geometric Call options, with payoffs of the form

$$\left(\exp \left\{ \frac{1}{T} \int_0^T X_t dt \right\} - K \right)^+ = \left(S_0 \exp \left\{ \frac{rT}{2} \int_0^T \frac{T-t}{T} dX_t \right\} - K \right)^+,$$

where $x^+ = \max\{x, 0\}$. We consider the following model parameters (realistic on Equity markets) for the Heston model recalled in Section 4.4:

$$S_0 = 50; \quad r = 0.05; \quad v_0 = 0.04; \quad \rho = -0.5; \quad \kappa = 2; \quad \theta = 0.09; \quad \xi = 0.2.$$

To simulate the paths (X, V) on $[0, T]$, we use a standard Euler-Maruyama scheme for X , but use the scheme [30] for the CIR process in the volatility, which is upward biased², however nevertheless converges strongly in L^1 to the true process V , which is enough for the purpose of pricing. For $n \in \mathbb{N}$, $\Delta = \frac{T}{n}$ and the increments of the Brownian motion $\{\Delta W_i^n\}_{i=0}^{n-1}$ the discretisation scheme over $[0, T]$ for the variance process $\{V_i^n\}_{i=0}^n$ reads:

$$\begin{cases} \tilde{V}_0^n &= v_0 > 0, \\ \tilde{V}_{i+1}^n &= \tilde{V}_i^n + \kappa \left(\theta - \tilde{V}_i^{n,+} \right) \Delta + \xi \sqrt{\tilde{V}_i^{n,+}} \Delta W_i^n, \quad \text{for all } i \in \{0, \dots, n-1\}, \\ V_i^n &= \tilde{V}_i^{n,+}. \end{cases}$$

In what follows, we compare different LDP and MDP, with $n = 252$ trading days per year. All the results are computed for $T = 1$ using $N_{MC} = 500,000$ Monte-Carlo samples. We also consider an antithetic estimator and an LDP estimator derived assuming a deterministic volatility (denoted by BS). Furthermore, since LDP-based deterministic changes of drift in the BS setting (or in cases where the final form of the optimisation problem is similar) are easy to compute, we also propose a *fully adaptive* scheme based on the BS estimator:

$$h_t = \sum_{i=1}^n h_t^i \mathbb{1}_{[(i-1)\Delta, i\Delta]}(t),$$

²There are many discretisation schemes for the Heston model. Since the objective of this paper is not to study the effects of different schemes we satisfy ourselves with [30].

where h_t^i is the best deterministic change of drift up to the i -th discretisation step.³ We shall refer to *deterministic* schemes to mean changes of law with deterministic changes of drift and to *adaptive* changes of drift for changes of law with drift of the form $\int_0^t h_t \sqrt{V_t} dt$.

5.2.2. LDP results in different settings. We now look at the results of LDP based estimators in *small-noise*, *small-time* and *large-time* setting. Figure 3 indicates that the estimators derived in small-time setting provide good results, but are outperformed by small-noise estimators. Although not apparent in the figure, looking at Table 4, the adaptive estimators provide slightly better results, as a matter of fact, they are notably better for small strikes. However, the computation time is also higher for adaptive estimators, which balances out the slight increase in variance reduction for higher strikes. It thus seems that LDP adaptive estimators (in this case) imply a higher computational cost (given in Table 6) which is not justified by the variance reduction they can provide.

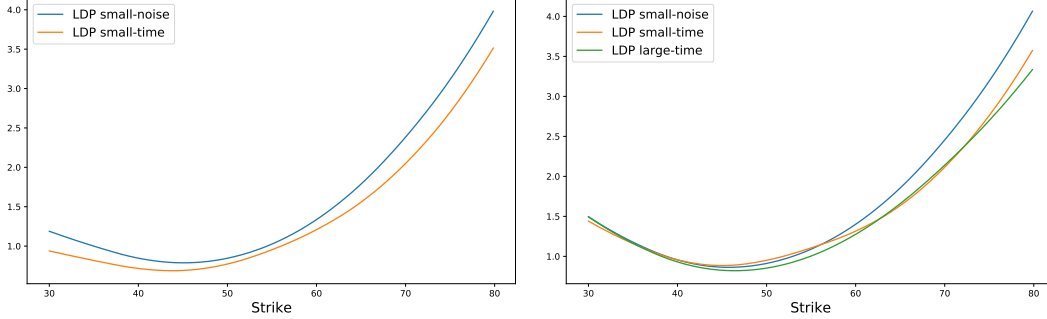


FIGURE 3. Variance reduction for LDP based estimators in log-scale. *Left*: deterministic change of drift. *Right*: adaptive changes of drift.

5.2.3. MDP results in different settings. In the deterministic case, all considered estimators have similar variance reduction (see Figure 4). To be more precise, the BS estimator has a very similar variance reduction or even slightly outperforms the MDP based estimators (left plot of Figure 4, where the lines are almost indistinguishable except for very high strikes, *outside the ‘moderate’ domain, which hence cause numerical problems*). Therefore, in that aspect, MDP based estimators do not justify their higher computational cost compared to the simple LDP-BS estimator. In the adaptive case, the BS estimator performs slightly better than before, whereas the MDP based estimators significantly outperform their results from the deterministic case *and* those of the BS estimator. Moreover, as it will be discussed in the next section, their variance reduction is in fact even close to that of the LDP based estimators.

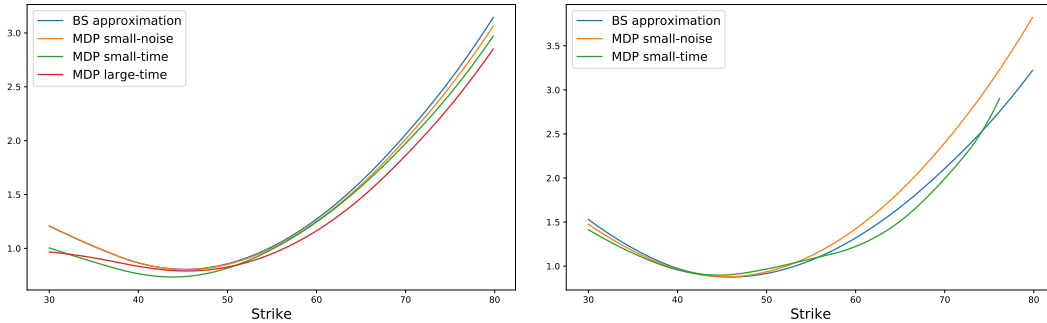


FIGURE 4. Variance reduction for MDP based estimators for in log-scale. *Left*: deterministic change of drift. *Right*: adaptive changes of drift. Because of computational problems, the adaptive small-time MDP estimator was not computed correctly for a strikes greater than 75. Nevertheless, it looks to be outperformed significantly by other MDP estimators.

³Fully adaptive schemes are computationally very heavy, therefore we only consider it in the Black-Scholes setting. The change of law is computed $n \times N_{MC}$ times ($252 \times 5E5 = 1.26E8$ in our case).

5.2.4. *Overall comparison.* Looking at Figure 5, as expected the LDP small-noise adaptive estimators perform best, although MDP small-noise and large-time adaptive estimators are not far behind. As for the computation times, Table 6 indicates MDP estimators are on average about 10% and LDP estimators approximately 15% slower than the corresponding standard BS estimators. The fully adaptive BS estimator provides interesting results, especially for near-the-money strikes, where it performs much better than MDP and LDP estimators. Although the estimator is time-consuming, it can still provide a good balance between variance reduction and computation time for certain strikes (Tables 4-5-6 and Figure 6). In the following three tables and plots, we use the following short notations (with in brackets their precise meanings, used in the graphs):

- **Prob**: Probability of having a positive payoff.
- **LDPsn** (*LDP small-noise*): Deterministic estimator based on LDP in small-noise log-price setting.
- **LDPsn A** (*LDP small-noise adaptive*): Adaptive estimator based on LDP in small-noise log-price setting.
- **BS** (*BS approximation*): Deterministic BS estimator.
- **BS A** (*BS approximation adaptive*): Adaptive BS estimator.
- **MDPsn A** (*MDP small-noise adaptive*): Adaptive estimator based on MDP in small-noise log-price setting.
- **BS A2** (*BS fully adaptive*): Fully adaptive BS estimator.
- **Ant** (*antithetic*): Antithetic estimator.
- **Classic**: Classic Monte-Carlo estimator.

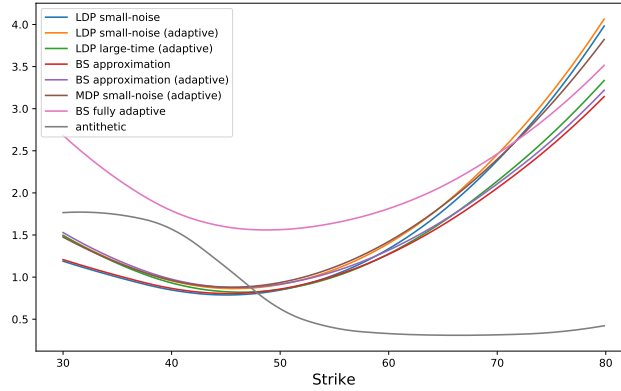


FIGURE 5. Variance reduction for different estimators in log-scale. The antithetic estimator offers almost no variance reduction for OTM options, because with high strikes very few paths end up in-the-money, thus reducing the effect of antithetic samples.

| Strike | Prob. | LDPsn | LDPsn A | BS | BS A | MDPsn A | BS A2 | Ant |
|--------|---------|-------|--------------|------|------|---------|------------|-----|
| 30 | 0.95 | 14 | 26 | 16 | 33 | 29 | 470 | 58 |
| 35 | 0.94 | 9.4 | 13 | 10 | 15 | 14 | 150 | 55 |
| 40 | 0.9 | 6.6 | 8.2 | 7.3 | 9.3 | 9.1 | 60 | 36 |
| 45 | 0.76 | 5.8 | 6.7 | 6.4 | 7.5 | 7.5 | 39 | 13 |
| 50 | 0.52 | 6.6 | 7.5 | 7.1 | 8.2 | 8.5 | 36 | 4.2 |
| 55 | 0.26 | 10 | 11 | 10 | 11 | 13 | 43 | 2.5 |
| 60 | 0.096 | 20 | 23 | 18 | 20 | 26 | 64 | 2.1 |
| 65 | 0.025 | 58 | 65 | 41 | 46 | 69 | 120 | 2.0 |
| 70 | 0.005 | 220 | 250 | 110 | 120 | 240 | 280 | 1.9 |
| 75 | 0.00078 | 1100 | 1200 | 310 | 350 | 960 | 750 | 1.9 |
| 80 | 0.0001 | 5700 | 6800 | 860 | 990 | 4000 | 2000 | 1.7 |
| 85 | 1.1e-05 | 35000 | 43000 | 2400 | 2800 | 18000 | 5900 | 2.7 |

TABLE 4. Variance reduction for different estimators and probability of positive payoff.

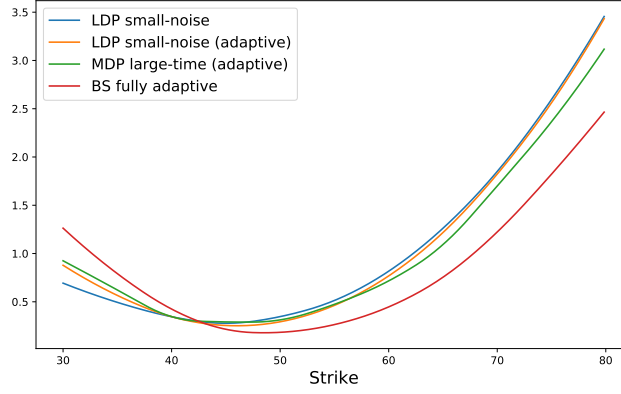


FIGURE 6. Ratio of variance reduction over computation time for selected estimators in log-scale.

| Strike | Prob. | LDPsn | LDPsn A | BS | BS A | MDPsn A | BS A2 | Ant |
|--------|---------|-------------|---------|-----|------|-----------|------------|-------------|
| 30 | 0.95 | 3.3 | 4.0 | 3.2 | 5.6 | 4.6 | 19 | 14.5 |
| 35 | 0.94 | 2.3 | 2.0 | 2.2 | 2.9 | 2.4 | 5.0 | 13.8 |
| 40 | 0.9 | 1.5 | 1.3 | 1.4 | 1.5 | 1.4 | 2.7 | 9.0 |
| 45 | 0.76 | 1.4 | 1.1 | 1.2 | 1.4 | 1.3 | 1.9 | 3.25 |
| 50 | 0.52 | 1.6 | 1.1 | 1.3 | 1.2 | 1.2 | 1.7 | 1.1 |
| 55 | 0.26 | 2.3 | 1.5 | 2.0 | 1.9 | 1.8 | 2.1 | 0.41 |
| 60 | 0.096 | 4.7 | 3.6 | 3.5 | 3.7 | 4.2 | 3.2 | 0.35 |
| 65 | 0.025 | 11 | 9.8 | 7.2 | 7.4 | 11 | 6.0 | 0.33 |
| 70 | 0.005 | 48 | 33 | 16 | 21 | 31 | 16 | 0.32 |
| 75 | 0.00078 | 240 | 180 | 48 | 63 | 160 | 52 | 0.27 |
| 80 | 0.0001 | 1300 | 960 | 180 | 180 | 660 | 160 | 0.28 |
| 85 | 1.1e-05 | 7300 | 6600 | 490 | 500 | 3000 | 510 | 0.45 |

TABLE 5. Ratio of variance reduction over computation time for different estimators.

| Strike | Classic | LDPsn | LDPsn A | BS | BS A | MDPsn A | BS A2 | Ant |
|--------|---------|-------|---------|----|------|---------|-------|-----|
| 30 | 11 | 12 | 14 | 12 | 13 | 14 | 31 | 11 |
| 35 | 11 | 11 | 14 | 12 | 13 | 13 | 36 | 11 |
| 40 | 11 | 12 | 13 | 13 | 14 | 14 | 30 | 11 |
| 45 | 11 | 11 | 13 | 13 | 13 | 13 | 28 | 11 |
| 50 | 11 | 11 | 14 | 13 | 14 | 15 | 28 | 11 |
| 55 | 11 | 12 | 15 | 12 | 13 | 15 | 28 | 11 |
| 60 | 11 | 12 | 14 | 12 | 13 | 13 | 28 | 11 |
| 65 | 11 | 12 | 14 | 13 | 13 | 13 | 27 | 11 |
| 70 | 11 | 12 | 15 | 14 | 13 | 15 | 24 | 11 |
| 75 | 11 | 12 | 14 | 14 | 13 | 13 | 21 | 11 |
| 80 | 11 | 12 | 14 | 12 | 13 | 13 | 20 | 11 |
| 85 | 11 | 12 | 14 | 12 | 13 | 13 | 19 | 11 |

TABLE 6. Computation time (in seconds) for different estimators.

5.2.5. *Variance swaps in Heston.* The methodology can also be applied to options with payoffs depending on volatility, for example for options with payoffs of the form $\int_0^T V_t \mathbb{1}_{\{S_t \geq K\}} dt$. Consider $T = 1$ and different strikes $K > 0$ under the Heston model with the same parameters as above. The results for different estimators are summarised in Table 7. For small strikes, the LDP estimators based on the BS approximation are not performing well, which is not surprising since in the BS approximation the payoff of the option in question is almost constant for small strikes. On the other

hand, LDP and MDP give good results. We also notice a clear difference in favour of non-adaptive changes of drift. For high strikes, we have the same behaviour as before: adaptive MDP estimators give intermediate results between LDP (performing best) and BS. [At this stage though, we do not have a clear explanation for the observed differences in the performance of non-adaptive and adaptive changes of drift for different strikes. Further research is required to understand the underlying dynamics better.](#)

| Strike | LDPsn | LDPsn A | MDPsn | MDPsn A | BS | BS A | Ant |
|--------|------------|------------|------------|-----------|-----|------|------------|
| 10 | 220 | 43 | 190 | 44 | 1.0 | 1.0 | 45 |
| 20 | 160 | 38 | 140 | 38 | 1.0 | 1.0 | 46 |
| 30 | 8.2 | 6.3 | 8.6 | 6.5 | 1.0 | 1.0 | 14 |
| 40 | 1.1 | 1.1 | 1.1 | 1.1 | 1 | 1 | 2.5 |
| 45 | 0.86 | 0.85 | 0.88 | 0.87 | 1.0 | 1.0 | 2.9 |
| 50 | 0.96 | 0.96 | 0.86 | 0.76 | 1.0 | 1.0 | 12 |
| 55 | 2.4 | 2.5 | 2.6 | 2.4 | 1.8 | 1.9 | 6.8 |
| 60 | 3.1 | 3.3 | 4.4 | 3.0 | 3.8 | 4.1 | 3.2 |
| 70 | 8.4 | 9.1 | 7.4 | 6.9 | 6.4 | 7.0 | 2.1 |
| 80 | 16 | 19 | 15 | 22 | 11 | 11 | 2.0 |
| 90 | 56 | 68 | 26 | 54 | 17 | 19 | 1.9 |
| 100 | 200 | 240 | 36 | 190 | 42 | 54 | 2.3 |

TABLE 7. Variance reduction for different estimators.

REFERENCES

- [1] P. BALDI AND L. CAMELLINO, *General Freidlin-Wentzell large deviations and positive diffusions*, Statistics & Probability Letters, 81 (2011), pp. 1218–1229.
- [2] C. BAYER, P. K. FRIZ, A. GULISASHVILI, B. HORVATH, AND B. STEMPER, *Short-time near-the-money skew in rough fractional volatility models*, Quantitative Finance, 19 (2018), pp. 779–798.
- [3] S. BIAGINI, T. PENNANEN, AND A.-P. PERKKIÖ, *Duality and optimality conditions in stochastic optimization and mathematical finance*, Journal of Convex Analysis, 25 (2018), pp. 403–420.
- [4] K. BINDER, D. M. CEPERLEY, J.-P. HANSEN, M. KALOS, D. LANDAU, D. LEVESQUE, H. MUELLER-KRUMBHAAR, D. STAUFFER, AND J.-J. WEIS, *Monte Carlo methods in statistical physics*, vol. 7, Springer Science & Business Media, 2012.
- [5] A. CHIARINI AND M. FISCHER, *On large deviations for small-noise Itô processes*, Advances in Applied Probability, 46 (2014), pp. 1126–1147.
- [6] G. CONFORTI, S. DE MARCO, AND J.-D. DEUSCHEL, *On small-noise equations with degenerate limiting system arising from volatility models*, in Large Deviations and Asymptotic Methods in Finance, Springer, 2015, pp. 473–505.
- [7] J. COX, J. INGERSOLL, AND S. ROSS, *A theory of the term structure of interest rates*, Econometrica, 53 (1985), pp. 385–407.
- [8] A. DEMBO AND O. ZEITOUNI, *Large deviations techniques and applications*, Springer, 2010.
- [9] C. DONATI-MARTIN, A. ROUAULT, M. YOR, AND M. ZANI, *Large deviations for squares of Bessel and Ornstein-Uhlenbeck processes*, Probability Theory and Related Fields, 129 (2004), pp. 261–289.
- [10] P. DUPUIS AND D. JOHNSON, *Moderate deviations-based importance sampling for stochastic recursive equations*, Advances in Applied Probability, 49 (2017), pp. 981–1010.
- [11] P. DUPUIS, K. SPILIOPOULOS, AND H. WANG, *Importance sampling for multiscale diffusions*, Multiscale Modeling & Simulation, 10 (2012), pp. 1–27.
- [12] P. DUPUIS AND H. WANG, *Importance sampling, large deviations, and differential games*, Stochastics: An International Journal of Probability and Stochastic Processes, 76 (2004), pp. 481–508.
- [13] M. I. FREIDLIN AND A. D. WENTZELL, *Random perturbations of dynamical systems*, Springer, 2012.
- [14] J. GARCIA, *A large deviation principle for stochastic integrals*, Journal of Theoretical Probability, 21 (2007), pp. 476–501.
- [15] J. GATHERAL, T. JAISSON, AND M. ROSENBAUM, *Volatility is rough*, Quantitative Finance, 18 (2018), pp. 933–949.
- [16] S. GERHOLD, A. JACQUIER, M. PAKKANEN, H. STONE, AND T. WAGENHOFER, *Pathwise large deviations for the rough Bergomi model: Corrigendum*, Journal of Applied Probability, 58 (2021), pp. 849–850.
- [17] P. GLASSERMAN, *Monte Carlo methods in Financial Engineering*, vol. 53, Springer, 2004.
- [18] P. GLASSERMAN AND Y. WANG, *Counterexamples in importance sampling for large deviations probabilities*, The Annals of Applied Probability, 7 (1997), pp. 731–746.
- [19] Z. GRBAC, D. KRIEF, AND P. TANKOV, *Long-time trajectorial large deviations and importance sampling for affine stochastic volatility models*, Advances in Applied Probability, 53 (2021), pp. 220–250.
- [20] P. GUASONI AND S. ROBERTSON, *Optimal importance sampling with explicit formulas in continuous time*, Finance and Stochastics, 12 (2007), pp. 1–19.

- [21] A. GULISASHVILI, *Large deviations principle for Volterra type fractional stochastic volatility models*, SIAM Journal on Financial Mathematics, 9 (2018), pp. 1102–1136.
- [22] ———, *Time-inhomogeneous Gaussian stochastic volatility models: Large deviations and super roughness*, Stochastic Processes and their Applications, 139 (2021), pp. 37–79.
- [23] C. HARTMANN, C. SCHÜTTE, M. WEBER, AND W. ZHANG, *Importance sampling in path space for diffusion processes with slow-fast variables*, Probability Theory and Related Fields, 170 (2018), pp. 177–228.
- [24] S. L. HESTON, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, Review of Financial Studies, 6 (1993), pp. 327–343.
- [25] A. JACQUIER, M. S. PAKKANEN, AND H. STONE, *Pathwise large deviations for the rough Bergomi model*, Journal of Applied Probability, 55 (2018), pp. 1078–1092.
- [26] A. JACQUIER AND A. PANNIER, *Large and moderate deviations for stochastic Volterra systems*, Stochastic Processes and their Applications, 149 (2022), pp. 142–187.
- [27] A. JACQUIER AND K. SPILIOPOULOS, *Pathwise moderate deviations for option pricing*, Mathematical Finance, 30 (2019), pp. 426–463.
- [28] I. KARATZAS AND S. E. SHREVE, *Brownian Motion and Stochastic Calculus*, Springer New York, 1998.
- [29] A. KLENKE, *Probability Theory*, Springer London, 2014.
- [30] R. LORD, R. KOEKKOEK, AND D. VAN DIJK, *A comparison of biased simulation schemes for stochastic volatility models*, Quantitative Finance, 10 (2009), pp. 177–194.
- [31] J. LYNCH AND J. SETHURAMAN, *Large deviations for processes with independent increments*, The Annals of Probability, 15 (1987), pp. 610–627.
- [32] B. F. MANLY, *Randomization, bootstrap and Monte Carlo methods in biology*, Chapman and Hall, 2018.
- [33] M. R. MORSE AND K. SPILIOPOULOS, *Moderate deviations for systems of slow-fast diffusions*, Asymptotic Analysis, 105 (2017), pp. 97–135.
- [34] M. R. MORSE AND K. SPILIOPOULOS, *Importance sampling for slow-fast diffusions based on moderate deviations*, Multiscale Modeling & Simulation, 18 (2020), pp. 315–350.
- [35] T. PENNANEN, *Convex duality in stochastic optimization and mathematical finance*, Mathematics of Operations Research, 36 (2011), pp. 340–362.
- [36] S. ROBERTSON, *Sample path large deviations and optimal importance sampling for stochastic volatility models*, Stochastic Processes and their Applications, 120 (2010), pp. 66–83.
- [37] R. T. ROCKAFELLAR, *Conjugate duality and optimization*, SIAM, 1974.
- [38] M. SCHILDER, *Asymptotic formulas for Wiener integrals*, Transactions of the AMS, 125 (1966), pp. 63–85.
- [39] A. SEPP AND P. KARASINSKI, *Beta stochastic volatility model*, Risk Magazine, (2012), pp. 66–71.
- [40] D. SIEGMUND, *Importance sampling in the Monte Carlo study of sequential tests*, The Annals of Statistics, (1976), pp. 673–684.
- [41] S. VARADHAN, *Diffusion processes in a small time interval*, Communications on Pure and Applied Mathematics, 20 (1967), pp. 659–685.
- [42] T. YAMADA AND S. WATANABE, *On the uniqueness of solutions of stochastic differential equations*, Kyoto Journal of Mathematics, 11 (1971).

APPENDIX A. TECHNICAL PROOFS

A.1. Proof of Lemma 3.1.1. Let $\beta > 1$, (\mathcal{X}, d) be a metric space and the random variable $Z^\varepsilon : \Omega \rightarrow \mathcal{X}$. We first show that Z^{ε^β} is exponentially equivalent to $x_0 \in \mathcal{X}$. For $\delta > 0$, let $\Gamma_\delta := \{x \in \mathcal{X} : d(x, x_0) > \delta\}$ and note that $\{\omega \in \Omega : Z^{\varepsilon^\beta}(\omega) \in \Gamma_\delta\}$ is measurable since Γ_δ is an open set. Since \tilde{I} is a good rate function and hence lower semi-continuous, it achieves the infimum over a closed set. It thus follows that $\inf_{x \in \overline{\Gamma}_\delta} \tilde{I}(x) > 0$, since $\tilde{I}(x) = 0$ if and only if $x = x_0$. Recall also that Z^ε satisfies an LDP with the good rate function \tilde{I} , then $\limsup_{\varepsilon \downarrow 0} \varepsilon^\beta \log \mathbb{P}[Z^{\varepsilon^\beta} \in \Gamma_\delta] \leq -\inf_{x \in \overline{\Gamma}_\delta} \tilde{I}(x)$ and therefore

$$\begin{aligned}
 (A.1) \quad \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}[Z^{\varepsilon^\beta} \in \Gamma_\delta] &= \limsup_{\varepsilon \downarrow 0} \varepsilon^{1-\beta} \left(\varepsilon^\beta \log \mathbb{P}[Z^{\varepsilon^\beta} \in \Gamma_\delta] \right) \\
 &\leq - \left\{ \inf_{x \in \overline{\Gamma}_\delta} \tilde{I}(x) \right\} \limsup_{\varepsilon \downarrow 0} \varepsilon^{1-\beta} = -\infty.
 \end{aligned}$$

Thus proving the exponential equivalence. Next, we show that $x_0 \in \mathcal{X}$ satisfies an LDP with the good rate function I . Let $\Gamma_\delta \subset \mathcal{X}$ such that $x_0 \notin \Gamma_\delta$, then $\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}[x_0 \in \Gamma] = -\infty$ for the upper bound and thus $-\inf_{x \in \Gamma^\circ} I(x) = -\infty$. Now let $\Gamma_{x_0} \subset \mathcal{X}$ be such that $x_0 \in \Gamma_{x_0}$, then $-\inf_{x \in \Gamma_\delta} I(x) = 0$ for the upper bound and $\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}[x_0 \in \Gamma_{x_0}] = 0$ for the lower bound. By the exponential equivalence the stated result follows. The second statement is immediate: for any $\delta > 0$, the inequality (A.1) implies exponential equivalence (the term on the right-hand side) by the large deviations assumption with $\beta > 1$, used in the inequality.

A.2. Proof of Lemma 3.1.2. Let (\mathcal{X}, d) be a metric space, $Z^\varepsilon : \Omega \rightarrow \mathcal{X}$ and $M > 0$ and let $B_M := \{x \in \mathcal{X} : |x| \leq M\}$. Since Z^ε is uniformly integrable, there exists $\delta > 0$ such that

$$\sup_{\varepsilon > 0} \mathbb{E} [|Z^\varepsilon| \mathbb{1}_{\{Z^\varepsilon \notin B_M\}}] < \delta,$$

and hence

$$\sup_{\varepsilon>0} \mathbb{E}[|Z^\varepsilon|] \leq \sup_{\varepsilon>0} \mathbb{E}[|Z^\varepsilon| \mathbb{1}_{\{Z^\varepsilon \in B_M\}}] + \sup_{\varepsilon>0} \mathbb{E}[|Z^\varepsilon| \mathbb{1}_{\{Z^\varepsilon \notin B_M\}}] \leq M + \delta.$$

Since x_0 is the (unique) zero of the rate function, then $\{Z^\varepsilon\}_{\varepsilon>0}$ clearly converges in probability to x_0 , and so $\{|Z^\varepsilon|\}_{\varepsilon>0}$ converges to $|x_0|$ by continuity. There exists then a subsequence $\{Z^{\vartheta_\varepsilon}\}_{\varepsilon>0}$ such that $\{|Z^{\vartheta_\varepsilon}|\}_{\varepsilon>0}$ converges almost surely to $|x_0|$. Fatou's lemma then implies

$$|x_0| = \mathbb{E}[|x_0|] \leq \liminf_{\varepsilon \downarrow 0} \mathbb{E}[|Z^{\vartheta_\varepsilon}|] < \sup_{\varepsilon>0} [|Z^{\vartheta_\varepsilon}|].$$

On the other hand, for $\Gamma_\delta := \{x \in \mathcal{X} : d(x, x_0) > \delta\}$ it holds by Bonferroni's (in)equality

$$\mathbb{E}[|Z^{\vartheta_\varepsilon}|] = \mathbb{E}[|Z^{\vartheta_\varepsilon}| \mathbb{1}_{\{(Z^{\vartheta_\varepsilon} \in \Gamma_\delta) \cap (Z^{\vartheta_\varepsilon} \notin B_M)\}}] + \mathbb{E}[|Z^{\vartheta_\varepsilon}| \mathbb{1}_{\{(Z^{\vartheta_\varepsilon} \in \Gamma_\delta) \cap (Z^{\vartheta_\varepsilon} \in B_M)\}}] + \mathbb{E}[|Z^{\vartheta_\varepsilon}| \mathbb{1}_{\{Z^{\vartheta_\varepsilon} \notin \Gamma_\delta\}}].$$

Since

$$|Z^{\vartheta_\varepsilon}| \mathbb{1}_{\{Z^{\vartheta_\varepsilon} \notin \Gamma_\delta\}} \leq \left(d(Z^{\vartheta_\varepsilon}, x_0) + |x_0|\right) \mathbb{1}_{\{Z^{\vartheta_\varepsilon} \notin \Gamma_\delta\}},$$

then $\mathbb{E}[|Z^{\vartheta_\varepsilon}|] \mathbb{1}_{\{Z^{\vartheta_\varepsilon} \notin \Gamma_\delta\}} \leq \delta + |x_0|$ and finally

$$\limsup_{\varepsilon \downarrow 0} \mathbb{E}[|Z^{\vartheta_\varepsilon}|] \leq \limsup_{\vartheta_\varepsilon \downarrow 0} \mathbb{E}[|Z^{\vartheta_\varepsilon}| \mathbb{1}_{\{Z^{\vartheta_\varepsilon} \notin B_M\}}] + M \lim_{\varepsilon \downarrow 0} \mathbb{P}[Z^{\vartheta_\varepsilon} \in \Gamma_\delta] + \delta + |x_0|.$$

Since $Z^{\vartheta_\varepsilon}$ follows an LDP with rate function I and $I(\Gamma_\delta) > 0$ by assumption, then $\lim_{\varepsilon \downarrow 0} \mathbb{P}[Z^{\vartheta_\varepsilon} \in \Gamma_\delta] = 0$. This, together with $Z^{\vartheta_\varepsilon}$ being uniformly integrable, gives the result by taking M to infinity.

A.3. Proof of Lemma 3.1.3. A key step in proving Lemma 3.1.3 is the tightness of the rescaled variance process with an unique strong solution

$$dV_t^\varepsilon = f(V_t^\varepsilon)dt + h(\varepsilon)g(V_t^\varepsilon)dW_t,$$

adhering to Assumption 1.4.1, where h is such as defined in the beginning of Section 3, i.e., $h(\varepsilon)$ tends to infinity and $\sqrt{\varepsilon}h(\varepsilon)$ to zero as ε goes to zero.

Lemma A.3.1. *The family of random variables $\{V^\varepsilon\}_{\varepsilon>0}$ in (3.3) is tight.*

Proof of Lemma A.3.1. By Kolmogorov-Chentsov [29, Theorem 21.42] we need to show there exist $\alpha, \beta, M > 0$ such that for every $\varepsilon > 0$ and $0 \leq s < t \leq T$

$$\mathbb{E}[|V_t^\varepsilon - V_s^\varepsilon|^\alpha] \leq M |t - s|^{1+\beta},$$

therefore using the obvious inequality $|a + b|^\alpha \leq 2^{\alpha-1}(|a|^\alpha + |b|^\alpha)$ for $\alpha \geq 1$ we have

$$\begin{aligned} \mathbb{E}\left[\left|\int_0^t f(V_u^\varepsilon)du - \int_0^s f(V_u^\varepsilon)du + \int_0^t g(V_u^\varepsilon)dW_u - \int_0^s g(V_u^\varepsilon)dW_u\right|^\alpha\right] \\ \leq 2^{\alpha-1} \left(\mathbb{E}\left[\left|\int_s^t f(V_u^\varepsilon)du\right|^\alpha\right] + \mathbb{E}\left[\left|\int_s^t g(V_u^\varepsilon)dW_u\right|^\alpha\right]\right). \end{aligned}$$

Consider for now only the drift term

$$\mathbb{E}\left[\left|\int_s^t f(V_u^\varepsilon)du\right|^\alpha\right] \leq (t-s)^{\alpha-1} \mathbb{E}\left[\int_s^t |f(V_u^\varepsilon)|^\alpha du\right] \leq K_1^\alpha (t-s)^{\alpha-1} \int_s^t \mathbb{E}[1 + |V_u^\varepsilon|^\alpha] du,$$

where we used Jensen's inequality for the first inequality. The second follows from the linear growth condition in Assumption 1.4.1. For the diffusion term

$$\begin{aligned} \mathbb{E}\left[\left|\int_s^t g(V_u^\varepsilon)dW_u\right|^\alpha\right] &\leq \left\{\frac{\alpha(\alpha-1)}{2}\right\}^{\frac{\alpha}{2}} \mathbb{E}\left[\left|\int_s^t |g(V_u^\varepsilon)|^2 du\right|^{\frac{\alpha}{2}}\right] \\ &\leq \left\{\frac{\alpha(\alpha-1)}{2}\right\}^{\frac{\alpha}{2}} (t-s)^{\frac{\alpha}{2}-1} \mathbb{E}\left[\int_s^t |g(V_u^\varepsilon)|^\alpha du\right] \\ &\leq K_2^\alpha \left\{\frac{\alpha(\alpha-1)}{2}\right\}^{\frac{\alpha}{2}} (t-s)^{\frac{\alpha}{2}-1} \int_s^t \mathbb{E}[1 + |V_u^\varepsilon|^{\alpha H}] du, \end{aligned}$$

where the first line follows from the Burkholder-Davis-Gundy inequality (with $\alpha \geq 2$) and the last one from the H -polynomial growth ($H \geq \frac{1}{2}$) condition on the diffusion. Adding both terms together

and applying the Gronwall lemma to the integrands yields

$$\begin{aligned}
\mathbb{E}[|V_t^\varepsilon - V_s^\varepsilon|^\alpha] &\leq 2^{\alpha-1} \left(K_1^\alpha (t-s)^{\alpha-1} \int_s^t \mathbb{E}[1 + |V_u^\varepsilon|^\alpha] du \right. \\
&\quad \left. + K_2^\alpha \left\{ \frac{\alpha(\alpha-1)}{2} \right\}^{\frac{\alpha}{2}} (t-s)^{\frac{\alpha}{2}-1} \int_s^t \mathbb{E}[1 + |V_u^\varepsilon|^{\alpha H}] du \right) \\
&\leq 2^{\alpha-1} \left(K_1^\alpha (t-s)^{\alpha-1} \int_s^t 2^{\frac{\alpha}{2}-1} (1 + v_0^\alpha) e^{2\alpha(u-s)} du \right. \\
&\quad \left. + K_2^\alpha \left\{ \frac{\alpha(\alpha-1)}{2} \right\}^{\frac{\alpha}{2}} (t-s)^{\frac{\alpha}{2}-1} \int_s^t 2^{\frac{\alpha}{2}-1} (1 + v_0^\alpha) e^{2\alpha(u-s)} du \right) \\
&\leq 2^{\alpha-1} \left(2^{\frac{\alpha}{2}-1} K_1^\alpha (1 + v_0^\alpha) e^{2\alpha HT} (t-s)^\alpha + 2^{\frac{\alpha H}{2}-1} K_2^\alpha (1 + v_0^{\alpha H}) e^{2\alpha HT} (t-s)^{\frac{\alpha}{2}} \right) \\
&\leq 2^{\alpha-1} (C_1 T^{\frac{\alpha}{2}} + C_2) (t-s)^{\frac{\alpha}{2}} = M |t-s|^{\frac{\alpha}{2}},
\end{aligned}$$

where the constants are

$$C_1 := 2^{\frac{\alpha}{2}-1} K_1^\alpha (1 + v_0^\alpha) e^{2\alpha T}, \quad C_2 := 2^{\frac{\alpha H}{2}-1} K_2^\alpha (1 + v_0^{\alpha H}) e^{2\alpha HT}, \quad M := 2^{\alpha-1} \max\{C_1 T^{\frac{\alpha}{2}}, C_2\}.$$

Choosing any $\alpha > 2$ and $\beta = \frac{\alpha}{2} - 1$ completes the proof. \square

Similarly as in [5], define the bounded map $\Phi_t : \mathcal{C}_T \rightarrow \mathbb{R}$ for each $t \in [0, T]$ as

$$\Phi_t(\psi) = \left| \psi_t - v_0 - \int_0^t f(\psi_s) ds \right| \wedge 1.$$

It is also continuous. Indeed, let $\psi^n \rightarrow \psi$ in \mathcal{C}_T , then we have by the [second](#) triangle inequality

$$|\Phi_t(\psi^n) - \Phi_t(\psi)| \leq |\psi^n - \psi| + \int_0^t |f(\psi^n) - f(\psi)| ds.$$

Since $\mathcal{A} := \{\psi^n \in \mathcal{C}_T : n \in \mathbb{N}\} \cup \{\psi\}$ is a compact subset of \mathcal{C}_T and f is Lipschitz continuous on \mathcal{A} by Assumption 1.4.1, there exists a Lipschitz constant L , such that

$$|f(\varphi_t) - f(\phi_t)| \leq L \sup_{t \in [0, T]} |\varphi_t - \phi_t|,$$

for all $t \in [0, T]$ and $\varphi, \phi \in \mathcal{A}$. Therefore, since

$$|\Phi_t(\psi^n) - \Phi_t(\psi)| \leq \sup_{t \in [0, T]} |\psi_t^n - \psi_t| + tL \sup_{t \in [0, T]} |\psi_t^n - \psi_t|,$$

then Φ_t is continuous. Since $\{V^\varepsilon\}_{\varepsilon>0}$ is tight as a family of random variables by Lemma A.3.1, we know that by taking a subsequence, $\{V^\varepsilon\}_{\varepsilon>0}$ converges in distribution to some random variable, call it $\bar{\psi}$, on the same probability space. Then, observing that Φ_t is continuous and bounded, we have by the Continuous mapping theorem that $\lim_{\varepsilon \downarrow 0} \mathbb{E}[\Phi_t(V^\varepsilon)] = \mathbb{E}[\Phi(\bar{\psi})]$. Next, by definition of Φ_t , $\bar{\psi} = v_0 + \int_0^t f(\bar{\psi}_s) ds$ will be the limiting function if we are able to show the above limit is zero for that particular choice of $\bar{\psi}$. Indeed, by Hölder's inequality

$$\mathbb{E}[\Phi_t(V^\varepsilon)] = \mathbb{E} \left[\left| h(\varepsilon) \int_0^t g(V_s^\varepsilon) dW_s \right| \right] \leq \left(|h(\varepsilon)|^2 \mathbb{E} \left[\left| \int_0^t g(V_s^\varepsilon) dW_s \right|^2 \right] \right)^{\frac{1}{2}} \leq h(\varepsilon) M,$$

since the integral can be shown to be bounded by $M > 0$ by the same arguments as in the proof of Lemma A.3.1. Therefore $\lim_{\varepsilon \downarrow 0} \mathbb{E}[\Phi_t(V^\varepsilon)] = 0$ and the limiting function indeed solves the ODE $\bar{\psi} = v_0 + \int_0^t f(\bar{\psi}_s) ds$ almost surely for all $t \in [0, T]$.

A.4. Proof of Theorem 3.1.1. We first prove the following version of Gronwall's lemma:

Lemma A.4.1. *Let $0 \leq a < b$ and $\varphi : [a, b] \rightarrow \mathbb{R}^+$ continuous with $\varphi(t) \leq M + \int_a^t f(s)g(\varphi(s))ds$ on $[a, b]$ for some $M > 0$. If $f : [a, b] \rightarrow \mathbb{R}$ and $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ satisfy Assumption 1.4.1, then $G(\varphi(t)) \leq G(M) + \int_a^t f(s)ds$ for all $t \in [a, b]$, where $G(u) = \int_1^u g^{-1}(s)ds$ and $G(0) = -\infty$.*

Proof. Let $\vartheta(t) := \int_a^t g(\varphi(s))f(s)ds$, so that $\dot{\vartheta}(t) = g(\varphi(t))f(t) \leq g(M + \vartheta(t))f(t)$ on $[a, b]$, since g is increasing. By integration,

$$\int_a^t f(s)ds \geq \int_a^t \frac{\dot{\vartheta}(s)}{g(M + \vartheta(s))} ds = \int_{M + \vartheta(a)}^{M + \vartheta(t)} \frac{du}{g(u)} = \int_M^{\vartheta(t) + M} \frac{du}{g(u)},$$

with $u := M + \vartheta(s)$, since $\vartheta(a) = 0$ by definition. The function G can be decomposed into:

$$G(u) = \int_1^u g^{-1}(s)ds = -\mathbb{1}_{\{u \in (0, 1)\}} \int_u^1 g^{-1}(s)ds + \mathbb{1}_{\{u \in [1, \infty)\}} \int_1^u g^{-1}(s)ds.$$

Second term is clearly increasing, since the inverse of an increasing function is an increasing function. Similarly, for the first term, since g^{-1} is increasing and positive for $u > 0$, the map $u \mapsto \int_u^1 g^{-1}(s)ds$ is decreasing and the term is increasing. Therefore, since G is increasing, then, for any $t \in [a, b]$,

$$G(\varphi(t)) \leq G(\vartheta(t) + M) \leq \int_1^{M+\vartheta(t)} \frac{ds}{g(s)} \leq \left(\int_1^M + \int_M^{M+\vartheta(t)} \right) \frac{ds}{g(s)} \leq G(M) + \int_a^t f(s)ds.$$

□

Let $Y^\varepsilon = \int_0^\cdot g(V_t^\varepsilon) dW_t^\varepsilon$. By Lemma 3.1.1 and Lemma 3.1.3, the process V^ε is exponentially equivalent to ψ (with speed ε), so that $\{V^\varepsilon, W^\varepsilon\}$ is exponentially equivalent to $\{\psi, W^\varepsilon\}$. Therefore, by Theorem 2.1.3, the triple $\{V^\varepsilon, W^\varepsilon, Y^\varepsilon\}$ satisfies an LDP with good rate function

$$I^{V, W, Y}(v, w, y) = \frac{1}{2} \int_0^T \dot{w}_t^2 dt, \quad \text{if } w \in \mathbb{H}_T^0, \dot{y} = g(\psi)\dot{w} \text{ and } v = \psi,$$

and is infinite otherwise. Now let

$$(A.2) \quad \tilde{\eta}_\varepsilon := \eta^\varepsilon - \int_0^\cdot f'(\psi_t) \eta^\varepsilon(t) dt = \int_0^\cdot \{f(V_t^\varepsilon) - f(\psi_t) - f'(\psi_t)(V_t^\varepsilon - \psi_t)\} \varepsilon^{-\beta} dt + Y^\varepsilon.$$

and suppose the first term is exponentially equivalent to 0. Then, by the Contraction principle [8, Theorem 4.2.1] we deduce an LDP for $\{V^\varepsilon, W^\varepsilon, \tilde{\eta}_\varepsilon\}$. Moreover, since $\tilde{\eta}_\varepsilon \mapsto \eta^\varepsilon$ is a continuous function⁴, the good rate function given in the statement is obtained using the Contraction principle once more. Therefore the rest of the proof relies on proving that

$$\{f(V^\varepsilon) - (f(\psi) - f'(\psi)(V^\varepsilon - \psi))\} \varepsilon^{-\beta}$$

is exponentially equivalent to 0. We start by showing η^ε is bounded. To that end we consider a Taylor expansion of $\ell(u) := f(\psi + u(V^\varepsilon - \psi))$ for $u \in \mathbb{R}$ around zero evaluated at $u = 1$:

$$\ell(1) = \ell(0) + \ell'(0) + \int_0^1 \ell''(u)(1-u) du,$$

so that, since $\ell(0) = f(\psi)$ and $\ell(1) = f(V^\varepsilon)$, we have

$$\{f(V^\varepsilon) - f(\psi) - f'(\psi)(V^\varepsilon - \psi)\} \varepsilon^{-\beta} = \varepsilon^\beta |\eta^\varepsilon|^2 \int_0^1 (1-u) f''(\psi_u + u(V_u^\varepsilon - \psi_u)) du.$$

Now let $R > 0$ and $\omega \in \Omega$ such that

$$\sup_{t \in [0, T]} |V_t^\varepsilon(\omega) - \psi_t| \leq R \quad \text{and} \quad \sup_{t \in [0, T]} |Y_t^\varepsilon(\omega)| \leq \varepsilon^{-\frac{\beta}{4}}.$$

Then for all $t \in [0, T]$ we have

$$\begin{aligned} |\eta_t^\varepsilon(\omega)| &= \left| \int_0^t \{f(V_s^\varepsilon(\omega)) - (f(\psi_s) - f'(\psi_s)(V_s^\varepsilon(\omega) - \psi_s))\} \varepsilon^{-\beta} ds + Y_t^\varepsilon(\omega) \right| \\ &= \left| \int_0^t f'(\psi_s) \eta_s^\varepsilon ds + \int_0^t |\eta_s^\varepsilon|^2 \varepsilon^\beta \int_0^1 (1-u) f''(\psi_u + u(V_u^\varepsilon(\omega) - \psi_u)) du ds + Y_t^\varepsilon(\omega) \right| \\ &\leq \int_0^t |f'(\psi_s)| |\eta_s^\varepsilon| ds + \int_0^t |\eta_s^\varepsilon|^2 \varepsilon^\beta \int_0^1 (1-u) |f''(\psi_u + u(V_u^\varepsilon(\omega) - \psi_u))| du ds + |Y_t^\varepsilon(\omega)| \\ &\leq \alpha \int_0^t (|\eta_s^\varepsilon| + |\eta_s^\varepsilon|^2 \varepsilon^\beta) ds + \varepsilon^{-\frac{\beta}{4}}, \end{aligned}$$

where

$$\alpha := 1 + \max \left\{ \sup_{x \in \mathcal{I}_\psi} |f'(x)|, \sup_{x \in \mathcal{I}_\psi} |f''(x)| \right\} > 0 \quad \text{and} \quad \mathcal{I}_\psi := \left[- \sup_{u \in [0, 1]} \{|\psi_u| + R\}, \sup_{u \in [0, 1]} \{|\psi_u| + R\} \right].$$

It is clear that α is finite, since continuous functions admit a maximizer on compact sets. Then by Lemma A.4.1 $G(|\eta_t^\varepsilon(\omega)|) \leq G(\varepsilon^{-\frac{\beta}{4}}) + \alpha t$ for all $t \in [0, T]$ and $u > 0$ with

$$G(u) = \int_1^u \frac{1}{x + \varepsilon^\beta x^2} dx = \log \left(u \frac{1 + \varepsilon^\beta}{1 + \varepsilon^\beta u} \right),$$

we have

$$G(|\eta_t^\varepsilon(\omega)|) \leq \log \left(\varepsilon^{-\frac{\beta}{4}} \frac{1 + \varepsilon^\beta}{1 + \varepsilon^{\frac{3}{4}\beta}} \right) + \alpha T, \quad \text{and} \quad |\eta_t^\varepsilon| \leq \varepsilon^{-\beta} \left\{ \frac{1}{1 - C(\varepsilon) \frac{\varepsilon^\beta}{1 + \varepsilon^\beta}} - 1 \right\},$$

where

$$C(\varepsilon) = e^{\alpha T} \varepsilon^{-\frac{\beta}{4}} \frac{(1 + \varepsilon^\beta)}{1 + \varepsilon^{\frac{3}{4}\beta}} < e^{\alpha T} \varepsilon^{-\frac{\beta}{4}}, \quad \text{for all } 0 < \varepsilon < 1.$$

⁴It is easy to show that $\eta^\varepsilon = e^{\int_0^\cdot f(\psi_s) ds} \int_0^\cdot e^{-\int_0^t f(\psi_s) ds} \tilde{\eta}_\varepsilon(t) dt$ solves the ODE in (A.2).

Therefore there exists $0 < \varepsilon_0 < 1$ such that

$$\sup_{t \in [0, T]} |\eta_t^\varepsilon(\omega)| \leq 2e^{\alpha T} \varepsilon^{-\frac{\beta}{4}}, \quad \text{for all } 0 < \varepsilon < \varepsilon_0.$$

We now finally prove that $\varepsilon^\beta |\eta^\varepsilon|^2 \int_0^1 (1-u) f''(\psi + u(V^\varepsilon - \psi)) du$ is exponentially equivalent to zero. Let $\delta > 0$ and $0 < \varepsilon < \varepsilon_0$, and define

$$\begin{aligned} \Gamma_\varepsilon &:= \left\{ \omega \in \Omega : \sup_{t \in [0, T]} \varepsilon^\beta |\eta_t^\varepsilon|^2 \left| \int_0^1 (1-u) f''(\psi_t + u(V_t^\varepsilon - \psi_t)) du \right| \geq \delta \right\}, \\ A_\varepsilon &:= \left\{ \omega \in \Omega : \sup_{t \in [0, T]} |V_t^\varepsilon - \psi_t| \leq R \right\}, \quad B_\varepsilon := \left\{ \omega \in \Omega : \sup_{t \in [0, T]} |Y_t^\varepsilon| \leq \varepsilon^{-\frac{\beta}{4}} \right\}, \end{aligned}$$

then

$$\begin{aligned} \varepsilon \log \mathbb{P}[\Gamma_\varepsilon] &\leq \varepsilon \log \left\{ \mathbb{P}[\Gamma_\varepsilon \cap A_\varepsilon \cap B_\varepsilon] + \mathbb{P}[A_\varepsilon^c] + \mathbb{P}[B_\varepsilon^c] \right\} \\ &\leq \varepsilon \log 3 + \max \left\{ \varepsilon \log \mathbb{P}[\Gamma_\varepsilon \cap A_\varepsilon \cap B_\varepsilon], \varepsilon \log \mathbb{P}[A_\varepsilon^c], \varepsilon \log \mathbb{P}[B_\varepsilon^c] \right\}. \end{aligned}$$

Using the previous bound on η^ε , which holds under condition $\{A_\varepsilon \cap B_\varepsilon\}$, we have that

$$\begin{aligned} \varepsilon \log \mathbb{P}[\Gamma_\varepsilon \cap A_\varepsilon \cap B_\varepsilon] &\leq \varepsilon \log \mathbb{P} \left[\left\{ \omega \in \Omega : \sup_{t \in [0, T]} \varepsilon^\beta |\eta_t^\varepsilon|^2 \geq \frac{1}{\alpha} \delta \right\} \cap A_\varepsilon \cap B_\varepsilon \right] \\ &\leq \varepsilon \log \mathbb{P} \left[\varepsilon^{\frac{\beta}{2}} \geq \frac{1}{4\alpha} e^{-2\alpha T} \delta \right] \xrightarrow{\varepsilon \downarrow 0} -\infty. \end{aligned}$$

Next, since V^ε is exponentially equivalent to ψ with speed ε , then $\lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}[A_\varepsilon^c] = -\infty$. Finally, looking at the LDP with speed $\varepsilon^{1+\frac{\beta}{2}}$, V^ε is still exponentially equivalent to ψ by the same argument as before, thus similarly $\varepsilon^{\frac{1}{2}+\frac{\beta}{4}} \int_0^\cdot g(V_t^\varepsilon) dW_t = \varepsilon^{\frac{\beta}{4}} Y^\varepsilon$ satisfies an LDP with good rate function having a unique minimum at zero and with speed $\varepsilon^{1+\frac{\beta}{2}}$. Therefore, $\varepsilon^{\frac{\beta}{4}} Y^\varepsilon$ is exponentially equivalent to zero and $\lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}[B_\varepsilon^c] = -\infty$, which completes the proof.

APPENDIX B. VARIATIONS AROUND VARADHAN'S LEMMA AND EXPONENTIAL EQUIVALENCE

Varadhan's integral lemma is a generalisation of Laplace's method. It gives the asymptotic behaviour of $\mathbb{E}[e^{\frac{\varphi(Z_\varepsilon)}{\varepsilon}}]$ on a log scale for a family of random variables Z_ε and a continuous function φ . The Laplace's method states that under some conditions the following relation holds

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \int_a^b e^{nf(x)} dx = \sup_{x \in [a, b]} f(x).$$

One notable application of Varadhan's integral lemma is finding a good change of measure in importance sampling, as we will see it in the following sections. For these applications, we need a slightly more general formulation of the lemma than the one found in [8, Theorem 4.3.1]. Nevertheless, the proof in [8, Theorem 4.3.1] can be easily adapted as done in [36, Lemma 4.4.].

Theorem B.0.1. (*Varadhan's Integral Lemma*) *Let \mathcal{X} be a metric space, Z^ε a family of \mathcal{X} -valued random variables satisfying an LDP with good rate function $I : \mathcal{X} \rightarrow [0, +\infty]$ and $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ a continuous function. Assume further either the tail condition*

$$\lim_{M \uparrow \infty} \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{\varphi(Z^\varepsilon)}{\varepsilon} \right\} \mathbb{1}_{\{\varphi(Z^\varepsilon) \geq M\}} \right] = -\infty,$$

or the moment condition for some $\gamma > 1$ (because it implies the previous tail condition)

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \gamma \frac{\varphi(Z^\varepsilon)}{\varepsilon} \right\} \right] < \infty.$$

Then

$$\lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{\varphi(Z^\varepsilon)}{\varepsilon} \right\} \right] = \sup_{x \in \mathcal{X}} \{\varphi(x) - I(x)\}.$$

The below-modified theorem was proven in [36] and allows the above function φ to reach $-\infty$ and accounts for cases where the problem is written in term of a family of functions $\{\varphi_\varepsilon\}_{\varepsilon > 0}$.

Theorem B.0.2. (*Modified Varadhan's Integral Lemma*) *Let \mathcal{X} and \mathcal{Y} be two metric spaces, Z^ε a family of \mathcal{X} -valued random variables that satisfies an LDP with good rate function $I : \mathcal{X} \rightarrow [0, +\infty]$ and let $\varphi : \mathcal{Y} \rightarrow [-\infty, +\infty]$ and $\psi : \mathcal{X} \rightarrow \mathbb{R}$ be two continuous functions. Let $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$ be*

a continuous map and $\{\Lambda_\varepsilon : \mathcal{X} \rightarrow \mathcal{Y}\}$ be a family of measurable functions such that $\Lambda_\varepsilon(Z^\varepsilon)$ is exponentially equivalent to $\Lambda(Z^\varepsilon)$. Suppose that there exists $\gamma > 1$ such that

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \gamma \frac{\varphi(\Lambda_\varepsilon(Z^\varepsilon)) + \psi(Z^\varepsilon)}{\varepsilon} \right\} \right] < \infty.$$

Then

$$\lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[\exp \left\{ \frac{\varphi(\Lambda_\varepsilon(Z^\varepsilon)) + \psi(Z^\varepsilon)}{\varepsilon} \right\} \right] = \sup_{x \in \mathcal{X}} \left\{ \varphi(\Lambda(x)) + \psi(x) - I(x) \right\}.$$

Definition B.0.1. (*Exponential equivalence*) Let (\mathcal{X}, d) be a metric space, Z^ε and $\widetilde{Z}^\varepsilon$ two families of \mathcal{X} -valued random variables defined on some probability spaces $\{(\Omega, B_\varepsilon, P_\varepsilon)\}$. Z^ε and $\widetilde{Z}^\varepsilon$ are called exponentially equivalent if for every $\delta > 0$, $\{w : d(Z^\varepsilon, \widetilde{Z}^\varepsilon) > \delta\}$ is B_ε measurable and $\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}_\varepsilon(d(Z^\varepsilon, \widetilde{Z}^\varepsilon) > \delta) = -\infty$.

ORFE DEPARTMENT, PRINCETON UNIVERSITY
Email address: mgeha@princeton.edu

DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE LONDON, AND ALAN TURING INSTITUTE
Email address: a.jacquier@imperial.ac.uk

DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE LONDON
Email address: z.zuric19@imperial.ac.uk