Model independent hedging strategies for variance swaps

David Hobson Martin Klimmek

December 14, 2011

Abstract

A variance swap is a derivative with a path-dependent payoff which allows investors to take positions on the future variability of an asset. In the idealised setting of a continuously monitored variance swap written on an asset with continuous paths it is well known that the variance swap payoff can be replicated exactly using a portfolio of puts and calls and a dynamic position in the asset. This fact forms the basis of the VIX contract.

But what if we are in the more realistic setting where the contract is based on discrete monitoring, and the underlying asset may have jumps? We show that it is possible to derive model-independent, no-arbitrage bounds on the price of the variance swap, and corresponding sub- and super-replicating strategies. Further, we characterise the optimal bounds. The form of the hedges depends crucially on the kernel used to define the variance swap.

1 Introduction

The purpose of this article is to construct hedging strategies which super-replicate the payoff of a variance swap for *any* price path of the underlying asset, including price paths with jumps. The idea is that at initiation time 0, an agent purchases a portfolio of puts and calls which she holds until time T. In addition, she follows a simple, dynamic investment strategy in the underlying over $[0,T]$. Then, for every possible path of the underlying, the sum of the payoff from the vanilla portfolio plus the gains from trade from the dynamic strategy is (more than) sufficient to cover the obligation from the variance swap. Implicit in this set-up is the idea that the super-hedge does not rely on any modelling assumptions. Instead, the super-hedge succeeds path-wise and is robust to the presence of jumps.

The problem of finding the cheapest super-hedging strategy can be seen as the dual of a primal problem which is to bound the prices for variance swaps over the class of all models for the asset price process which are consistent with the traded prices of puts and calls. If the variance swap is sold for the price upper-bound and hedged with the corresponding super-replicating strategy then the seller will not lose money under any scenario.

In addition to super-hedges and upper bounds on the price of the variance swap we also give sub-hedges and lower bounds. Moreover, our analysis is not restricted to any particular definition of the variance swap, nor is it based on a mathematical idealisation of a continuous time limit of the swap contract, but rather on a discrete set of observations. Nonetheless, the sub- and super-replicating hedges which work for discretely sampled variance swaps continue to work in the continuous time limit. As long as the price path has a quadratic variation, these limits exist by Föllmer's path-wise Itô formula [16].

This article shares the model-independent ethos for the pricing of variance swaps implicit in Neuberger [24] and Dupire [15] in the setting of continuous price processes. In those articles, it was shown that if we assume that the asset price process is a continuous forward price, then the continuously monitored variance swap based on either squared log returns or squared simple returns is perfectly replicated by the following strategy: synthesise $-2\log$ contracts using put and call options and trade continuously in the asset to hold a number of shares equal to twice the reciprocal of the current asset price at all times. We will refer to this strategy as the classical continuous hedge. It follows that in the setting of a continuous forward price, the unique no-arbitrage price for the variance swap is equal to the price of the contract with payoff equal to $-2\log$ contracts. This result holds independently of any modelling assumptions beyond path continuity. The hedging strategies in this paper are of the same character, consisting of a static position in calls and puts and dynamic trading in the underlying. However, the underlying setup is considerably more general, and the results more powerful since the hedges continue to super-replicate the variance swap for discontinuous price-paths and discrete monitoring over arbitrary time partitions. Nonetheless, this increase in generality comes at a cost in that instead of a replicating strategy we get sub- and super-replicating strategies and instead of a unique no-arbitrage price we get a no-arbitrage interval of prices.

As is well known, the model-independent analysis of derivative prices is related to the construction of extremal solutions for the Skorokhod embedding problem. This relationship was first developed in Hobson [17], see Hobson [18] for a recent survey, and exploits the idea that the classification of martingales with a given terminal law is equivalent to the classification of stopping times for Brownian motion, such that the stopped process has that given law. As we shall see, the monotone function which is associated with the cheapest super-hedging strategy arises in the Perkins solution [26] of the Skorokhod embedding problem [27]. For another example of model independent pricing and the connection between derivatives and the Skorokhod embedding problem in the context of variance options, see Cox and Wang [10]. In the setting of continuous price paths Cox and Wang [10] give bounds on the prices of call options on realised variance by exploiting a connection with the Root solution of the Skorokhod embedding problem.

In a recent paper [22], Kahalé shows how to derive a tight sub-replicating strategy and corresponding model-independent lower bound for the price of a variance swap based on the squared log return kernel. The paper by Kahalé was an inspiration for our study which grew from an attempt to relate his work to the previous literature on model-independent bounds and the Skorokhod embedding problem. By framing the problem in this way we extend the results of Kahalé [22] to other kernels, and give upper bounds as well as lower bounds. Moreover, in the case of squared returns where the connection is particularly explicit, we explain the origin of the extremal models, and we give a natural interpretation for some of the quantities appearing in [22] in terms of the Perkins embedding of the Skorokhod embedding problem. The analysis of the squared returns kernel motivates our general approach to variance swap bounds and links this work to previous results of the authors (Hobson and Klimmek [19]) on characterising solutions of the Skorokhod embedding problem with particular optimality properties.

One of the features of our analysis is that we study the variance swap under a variety of definitions for the contract. Early definitions of the variance swap were based on squared simple daily returns. Accordingly, the first analysis of the discrepancy between the classical continuous hedge and realised variance in the presence of jumps, which is due to Demeterfi et. al. [12], focused on this kernel. Later, the finance industry switched to a standardised definition based on log-returns. (These contracts are typically sold OTC, and therefore any specification of the contract, and any observation frequency is possible.) In the presence of discrete monitoring or jumps (but not in the case of continuous monitoring and continuous price processes) each kernel lends different characteristics to variance swap values. Partly for this reason a variety of kernels have been proposed in the literature. Bondarenko [3] introduces a kernel which lies between the squared log return and squared simple return definitions. Bondarenko's proposal is motivated by the fact that variance swaps based on this kernel can be replicated perfectly in the presence of jumps and in discrete time, see also Neuberger [25]. The kernel proposed by Carr and Corso [7] in the context of commodity markets, which is based on squared price differences, belongs to the same class. Recently Martin [23] has proposed yet another definition which is similar to the squared-return kernel but involves both the forward and the asset price. Our analysis covers all these kernels (though the kernel in [23] is only covered for the case of zero interest rates), and emphasises that the impact of jumps depends crucially on the nature of the kernel. We find that kernels split into two classes - below we name them increasing and decreasing kernels — and the special properties of the Bondarenko kernel come from the fact that it lies in the intersection of these classes.

Apart from asset price jumps, a further issue in the pricing and hedging of variance swaps is that the idealised continuous time limit may be a poor approximation to the traded contract which is based on discrete monitoring. For example, in [5] Broadie and Jain show that when the price path has negative jumps the value of the discretely monitored (log-return) variance swap can differ significantly from the value of the continuously monitored variance swap. Similarly, Bondarenko [3] investigates the hedging error that develops if the strategy of the classical continuous approach is approximated discretely, and reports replication errors of around 30 percent. From a theoretical perspective, Jarrow et. al. [21] show that we may have that the price of the continuously monitored variance swap is finite, whilst simultaneously the discretely sampled analogue may has an infinite price, an observation which raises fundamental questions about the validity of using the continuous time integrated variance as an approximation for the discretely monitored quantity. These previous studies underscore the importance of a model-independent analysis, especially one based on a finite number of monitoring points.

Recognising the importance of the jump contribution to variance swap values, Carr, Lee and Wu [9] show how it is possible to price and hedge a variance swap based on log returns if the asset price follows a Lévy model. The analysis is extended to a more general class of variation swaps in Carr and Lee [8]. Given a particular Lévy model for the dynamics of the price path, Carr and Lee show that there exists a model-dependent adjustment to the multiplier 2 appearing in the classical continuous hedge such that the value of the variance swap is given by the new multiplier times the price of a log-contract. In general, this price is not enforceable through a hedging strategy. Moreover, since all models are wrong and since the adjustment of the multiplier depends on specifying a particular model, this approach may still significantly mis-price realised variance, even if the Lévy model calibrates well to options prices.

The appeal of the classical continuous hedge of Neuberger and Dupire is that, apart from price-path continuity, the only necessary assumption is that a log contract can be synthesised from put and call options, and then the option payoff can be replicated perfectly along each path. In this article, we continue to assume that regular payoffs can be replicated with vanilla options, but relax the continuity assumption. The prices of variance swaps are highly sensitive to the presence of jumps, and so this is an important advance.

2 Main results

In this section we outline the main results of this article. Precise statements follow in the main body of the text, but here we describe the main theorems, and explain how they can be interpreted to give a model-independent bound on the price of a discretely monitored variance swap. Broadly speaking the results show firstly how to construct a family of model-independent sub-hedges for the variance swap (consisting of a static European position which may be synthesised with calls, and a simple dynamic trading position in the underlying) each of which is associated with a lower bound on the price of the variance swap; secondly how to choose the best (most expensive) strategy of this class; and thirdly that there are no other strategies of any class which outperform this best strategy without exposing the agent to model-risk. Fourthly we show how the results extend to super-hedges and upper bounds, and how they can be modified to cover other specifications of a variance swap.

Result 2.1. *(See Theorem 5.10) Let $f_0, f_1, \dots f_N$ be a sequence of positive real numbers.*

(a) There exist a pair of functions $\psi : (0,\infty) \to \mathbb{R}$ and $\delta : (0,\infty) \to \mathbb{R}$ such that $\psi(f_0) = 0$ and

$$\sum_{k=0}^{N-1} \left(\frac{f_{k+1} - f_k}{f_k} \right)^2 \geq \psi(f_N) + \sum_{k=0}^{N-1} \delta(f_k)(f_{k+1} - f_k). \tag{1}$$

(b) There exists a family of such pairs of functions where $\psi = \psi_{\kappa,R}$ is given by

$$\psi_{\kappa,R} = \left\{ \begin{array}{l} \psi_{\kappa,R}(x) \\ \psi_{\kappa,R}(z) \end{array} \right\} = \left\{ \begin{array}{ll} 2\int_{f_0}^{x} (x-u)\kappa(u)u^{-3}du & x \geq f_0 \\ \psi_{\kappa,R}(k(z)) + \psi'_{\kappa,R}(k(z))(z - k(z)) + ((z/k(z)) - 1)^2 & z < f_0. \end{array} \right. \tag{2}$$

and $\delta = -\psi'_+$, the right-derivative of ψ. Here κ is a decreasing function $\kappa : [f_0,\infty) \mapsto (0, f_0]$, and k is its right-inverse. Note that ψ is continuously differentiable on (f_0,∞).

Result 2.2. *(See Theorem 6.3) Let μ be a given probability measure on $(0,\infty)$ with mean f_0 and consider*

$$\sup_{\kappa} \left\{ \int \psi_{\kappa,R}(f)\mu(df) \right\}$$

where the supremum is taken over decreasing functions $\kappa : [f_0,\infty) \mapsto (0, f_0]$. Then the supremum is attained by $\kappa = \alpha(\mu)$ where α is a quantity which arises in the Perkins solution of the Skorokhod embedding problem for μ in Brownian motion.

Result 2.3. *(See Theorem 7.7) Suppose μ is as above and let $\psi = \psi_{\alpha(\mu),R}$. Then there exists a stochastic model $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{0\le t\le T})$ supporting a stochastic process $X = (X_t)_{0\le t\le T}$ such that $X_0 = f_0$, X_T has law μ, X is a martingale, and*

$$\int_0^T \frac{d[X]_t}{X_{t-}^2} = \psi(X_T) - \int_0^T \psi'_+(X_{s-})dX_s. \tag{3}$$

The sequence $f_0, f_1, \dots f_N$ should be interpreted as the realised values of the forward price of an asset evaluated on a grid of points $0 = T_0 < T_1 < \dots < T_N = T$. Then the payoff of the floating leg of a discretely monitored variance swap (based on squared returns) is $\sum_{k=0}^{N-1}((f_{k+1} - f_k)/f_k)^2$. Now consider a trading strategy which involves holding a static position in European calls with payoff ψ, coupled with a dynamic position in the forward whereby $\delta(f_k)$ units of the forward are held over the time-interval $(T_k, T_{k+1}]$ for each $k = 0, 1, \dots, N-1$. The final payoff of such a strategy is $\psi(f_N) + \sum_{k=0}^{N-1} \delta(f_k)(f_{k+1} - f_k)$. Then the inequality (1) above shows that it is possible to choose the pair (ψ, δ) such that the payoff of the variance swap dominates the terminal value of the trading strategy on a path-by-path basis. Consequently, no-arbitrage arguments give that the price of the variance swap is bounded below by the price of the European contract with payoff ψ. (Note that the sum $\sum_k \delta(f_k)(f_{k+1} - f_k)$ of forward contracts is costless by definition.) The payoff $\psi \equiv \psi(f_N)$ can be synthesised using calls, and if we assume that calls with maturity T are traded, this contact can be priced in a model-independent fashion in terms of those traded calls. Hence, given the prices of call options, the first part of Result 2.1 can be used to give a model-independent bound on the price of a variance swap.

The second part of Result 2.1 gives a family of suitable hedging strategies, parameterised by a decreasing function κ. Different choices of κ will lead to different lower bounds on the price of the variance swap. Given call prices, (or equivalently given the law of the forward price at time T as represented by μ) Result 2.2 shows how to choose the best contract based on functions $(\psi, \delta = -\psi'_+)$ of the given form by maximising the cost of the hedging portfolio, thus obtaining the best (highest) lower bound. Result 2.2 also relates the optimal contract to certain quantities which arise in the Perkins construction of the Skorokhod embedding problem.

Finally, Result 2.3 shows that the bounds we get from considering hedges of the given form are best possible, in the sense that in the case of continuous monitoring (or more generally in the limit as the partition becomes dense) there is a stochastic model for the forward price which is consistent with no-arbitrage and with the prices of traded calls and for which the payoff of the variance swap is equal, path-wise, to the payoff of the model-independent hedging strategy. Hence also, the model-based price of the variance swap is equal to the model-independent lower bound.

It is also possible to consider super-hedges and price upper bounds, and other definitions of the variance swap contract.

Result 2.4. *(See Theorems 5.10, 6.3 and 7.7)*

(a) There are analogous versions of Results 2.1, 2.2 and 2.3 for upper bounds for variance swaps based on the inequality

$$\sum_{k=0}^{N-1}\left(\frac{f_{k+1} - f_k}{f_k}\right)^2 \le \psi(f_N) + \sum_{k=0}^{N-1} \delta(f_k)(f_{k+1} - f_k). \tag{4}$$

This time the associated family of functions is based on a decreasing function $\ell : (0, f_0] \mapsto [f_0, \infty)$, where the best choice of ℓ is $\ell = \beta(\mu)$ where β is a quantity which arises in a 'reversed' Perkins solution of the Skorokhod embedding problem.

(b) There are analogous versions of Results 2.1, 2.2 and 2.3 for upper and lower bounds for variance swaps based on squared logarithmic returns, in which the left-hand-side of (1) or (4) is replaced by $\sum_{k=0}^{N-1}(\log f_{k+1} - \log f_k)^2$, and the left-hand-side of (3) is replaced by $[\log X]_T - [\log X]_0$. Although the families of functions (ψ, δ) change, they are again based on monotonic functions and the same functions $\kappa = \alpha(\mu)$ and $\ell = \beta(\mu)$ yield the upper and lower bounds. However, for the squared logarithmic return payoff $\kappa = \alpha(\mu)$ is now associated with the upper bound, and $\ell = \beta(\mu)$ is associated with the lower bound.

Our analysis is inspired by the work of Kahalé [22]. Kahalé considers variance swap contracts based on squared logarithmic returns, and lower bounds only, and so one of our contributions is to extend his work to upper bounds and to other specifications of the variance swap. Kahalé's key contribution is to observe that an inequality of the form

$$(\log y - \log x)^2 \geq \psi(y) - \psi(x) + \delta(x)(y - x) \tag{5}$$

can be summed along sequences $(f_k)_{0\leq k\leq N}$ to yield $\sum_{k=0}^{N-1}(\log f_{k+1} - \log f_k)^2 \geq \psi(f_N) - \psi(f_0) + \sum_{k=0}^{N-1}\delta(f_k)(f_{k+1} - f_k)$ which is the analogue of (1). Hence in order to find lower bounds for the prices of variance swaps it is sufficient to find solutions of (5).

Our primary goal in this paper is to prove theorems corresponding to the above results for a wide variety of definitions of the variance swap, including both squared returns and squared logarithmic returns. Our secondary goal is to explain why the model which is associated with the lower bound for squared log-return based contract is associated with the upper bound for the squared return contract and vice-versa (it turns out that the effect of jumps is opposite for these two contracts) and to explain why the optimal martingale is related to the Perkins solution of the Skorokhod embedding problem.

The remainder of the paper is structured as follows. In the next section we formally introduce the variance swap, and show how the definition depends on the form of the kernel. We also introduce the notions of model independent hedging, and consistent models. In Section 4 we study the problem in the setting of continuous monitoring for a process with jumps. We use this section to develop intuition and to explain why slightly different kernels can lead to opposite results. We also explain why for the squared return kernel the payoff of the variance swap is directly linked to a Skorokhod embedding problem.

The understanding we develop in Section 4 will motivate much of the subsequent analysis. Section 5 contains the first main theorem, and shows how to construct a class of sub-hedging strategies. In Sections 6 we find the most expensive sub-hedge of this class for a given set of call prices, and thus we derive a model independent bound on the price of a variance swap. Then in Section 7 we show this bound is best possible, by showing that in the continuous time limit it can be attained. Here we rely on Föllmer's work ([16]) on constructing Itô calculus without probability. In Section 8 we extend our results from contracts written on forwards to include the case of contracts written on undiscounted prices. The penultimate section gives some numerical results and in Section 10 we describe more precisely the contribution of this paper relative to the seminal paper by Kahalé, [22].

3 Variance Swap Kernels and Model-Independent Hedging

3.1 Variation swaps

We begin by defining the payoff of a variance swap on a path-by-path basis. The payoff will depend on a kernel, on the times at which the kernel is evaluated and on the asset price at these times.

Definition 3.1. *(i) A variation swap kernel is a continuously differentiable bi-variate function $H : (0,\infty) \times (0,\infty) \to [0,\infty)$ such that for all $x \in (0,\infty)$, $H(x,x) = 0 = H_y(x,x)$. We say that the swap kernel is regular if it is three times continuously differentiable.*

A variance swap kernel is a regular variation swap kernel H such that $H_{yy}(x,x) = x^{-2}$.

(ii) A partition P on $[0,T]$ is a set of times $0 = t_0 < t_1 < ... < t_N = T$. A partition is uniform if $t_k = \frac{kT}{N}$, $k = 0,1,...N$. A sequence of partitions $\mathcal{P} = (P^{(n)})_{n\geq 1} = (\{t_k^{(n)}; 0 \leq k \leq N^{(n)}\})_{n\geq 1}$ is dense if $\lim_{n\uparrow\infty} \sup_{k\in\{0,...,N^{(n)}-1\}} |t_{k+1}^{(n)} - t_k^{(n)}| = 0$.

(iii) A price realisation $f = (f(t))_{0\leq t\leq T}$ is a càdlàg function $f : [0,T] \to (0,\infty)$.

(iv) The payoff of a variation swap with kernel H for a partition P and a price realisation f is

$$V_H(f,P) = \sum_{k=0}^{N-1} H(f(t_k), f(t_{k+1})). \tag{6}$$

(v) Let $\mathcal{P} = (P^{(n)})_{n\geq 1}$ be a dense sequence of partitions. If $\lim_{n\uparrow\infty} V_H(f,P^{(n)})$ exists, then the limit is denoted $V_H(f,P_\infty)$ and is called the continuous time limit of $V_H(f,P^{(n)})$ on $\mathcal{P}$.

Remark 3.2. *(i) Our main focus in this article is on variance swap kernels but we will discuss variation swap kernels $H^S(x,y) = (y-x)^3$ and $H^Q(x,y) = (y-x)^2$ briefly, see Remark 4.1 and Example 7.10. (Strictly speaking H^S is not a variation swap kernel since it is not non-negative, but most of our analysis still applies in this case.) A regular variation swap kernel is a variance swap kernel if $H(x, x(1+\delta)) = \delta^2 + o(\delta^2)$ for δ small. Examples of variance swap kernels include $H^R(x,y) = \left(\frac{y-x}{x}\right)^2$, $H^L(x,y) = (\log(y) - \log(x))^2$ and $H^B(x,y) = -2\left(\log(y/x) - \left(\frac{y-x}{x}\right)\right)$.*

(ii) The price realisations f should be interpreted as realisations of the forward price of the asset with maturity T. Later we will extend the analysis to cover undiscounted price processes, rather than forward prices.

(iii) Large parts of the subsequent analysis can be extended to allow for price processes which can take the value zero, provided we also define $H(0,0) = 0$, or equivalently truncate the sum in (6) at the first time in the partition that f hits 0. In this case we must have that zero is absorbing, so that if $f(s) = 0$, then $f(t) = 0$ for all $s \leq t \leq T$.

(iv) In practice the variance swap contract is an exchange of the quantity $V = V_H(f,P)$ for a fixed amount K. However, since there is no optionality to the contract, and since the contract paying K can trivially be priced and hedged, we concentrate solely on the floating leg.

(v) In many of the earliest academic papers, and in particular in Demeterfi et. al [12, 13], but also in some very recent papers, e.g. Zhu and Lian [28], the variance swap is defined in terms of the kernel H^R. However, it has become market practice to trade variance swaps based on the kernel H^L. Nonetheless these contracts are traded over-the-counter and in principle it is possible to agree any reasonable definition for the kernel. Variance swaps defined using the variance kernel H^B were introduced by Bondarenko [3], see also Neuberger [25]. As we shall see, the contract based on this kernel has various desirable features. For continuous paths, in the limit of a dense partition the contract does not depend on the chosen kernel, see Example 7.10 and Lemma 7.9, but this is not the case in general.

(vi) The labels $\{S,Q,R,L,B\}$ on the variation swap kernels denote {Skew, Quadratic, Returns, Logarithmic returns, Bondarenko} respectively.

An important concept will be the quadratic variation of a path. For a dense sequence of partitions $\mathcal{P}$, the *quadratic variation* $[f]$ of f on $\mathcal{P}$ is defined to be $[f]_t = \lim_{n\uparrow\infty} \sum_{t_k^{(n)} \leq t} (f(t_{k+1}^{(n)}) - f(t_k^{(n)}))^2$, provided the limit exists. We split the function into its continuous and discontinuous parts, $[f]_t = [f]_t^c + \sum_{u\leq t}(\Delta f(u))^2$. Later we will relate this definition to that introduced by Föllmer [16], which is used to develop a path-wise version of Itô calculus. Föllmer's non-probabilistic Itô calculus has been used elsewhere in mathematical finance, most notably by Bick and Willinger [2], and helps emphasise the fact that the gains from trade have an interpretation as (the limit of) Riemann sums.

3.2 Model independent pricing

Our goal is to discuss how to price the variance swap contract, or more generally any path-dependent claim, under an assumption that European call and put (vanilla) options with maturity T are traded and can be used for hedging, but without any assumption that a proposed model is a true reflection of the real dynamics. In this sense the strategies and prices we derive are model independent.

Let call prices for maturity T be given by $C(K)$, written as a function of strike and expressed in units of cash at time[1] T. We assume that a continuum (in K) of calls are traded, and to preclude arbitrage we assume that C is a decreasing convex function such that $C(0) = f(0)$, $C(K) \geq (f(0) - K)^+$ and $\lim_{K\uparrow\infty} C(K) = 0$, see e.g. Davis and Hobson [11]. We exclude the case where $C(f(0)) = 0$ for then $C(K) = (f(0) - K)^+$ and the situation is degenerate: the forward price must remain constant and upper and lower bounds on the price of the variance swap are zero. Although we assume that calls are traded today (time 0), we do not make any assumption on how call prices will behave over time, except that they will respect no-arbitrage conditions and that on expiry they will be worth the intrinsic value.

Definition 3.3. *A synthesisable payoff is a function $\psi : (0,\infty) \mapsto \mathbb{R}$ which can be represented as the difference of two convex functions (so that $\psi''(x)$ exists as a measure).*

[1]Throughout this paper, all option contracts have maturity T unless stated otherwise.

Let $\Psi = \{\psi : \psi \in \Psi\}$ be the set of synthesisable payoffs $\psi : (0,\infty) \mapsto \mathbb{R}$. Then the left- and right-derivatives $\psi'_{\pm}$ (or $\psi'(x\pm)$) exist and we have

$$\psi(f) = \psi(f(0)) + \psi'(f(0)+)(f - f(0)) + \int_{(0,f(0)]} (x-f)^+ \psi''(x)dx + \int_{(f(0),\infty)} (f-x)^+ \psi''(x)dx. \tag{7}$$

Thus we can represent the payoff of any sufficiently regular European contingent claim as a constant plus the gains from trade from holding a fixed quantity of forwards, plus the payoff of a static portfolio of vanilla calls and puts.

Let $D[0,t]$ denote the space of càdlàg functions on $[0,t]$.

Definition 3.4. *A dynamic strategy for a fixed partition P is a collection of functions $\Delta = (\delta_{t_0}, \dots, \delta_{t_{N-1}})$, where $\delta_{t_j} : D[0,t_j] \to \mathbb{R}$. The payoff of a dynamic strategy along a price realisation f is*

$$\sum_{k=0}^{N-1} \delta_{t_k}((f(t))_{0\le t\le t_k})(f(t_{k+1}) - f(t_k)). \tag{8}$$

Let $\bar{\Delta}(P)$ be the set of dynamic strategies.

Definition 3.5. *$\Delta \in \bar{\Delta}(P)$ is a Markov dynamic strategy if $\delta_{t_j}(f(t)_{0\le t\le t_j}) = \delta_{t_j}(f(t_j))$ for all j. A Markov dynamic strategy is a time homogeneous Markov dynamic strategy (THMD-strategy) if $\delta_{t_j}(f(t_j)) = \delta(f(t_j))$ for all j.*

The quantity δ_{t_j} represents the quantity of forwards to be held over the interval $(t_j, t_{j+1}]$. In principle this quantity may depend on the current time and on the price history $(f(t))_{0\le t\le t_j}$. However, as we shall see, for our purposes it is sufficient to work with a much simpler set of strategies where the quantity does not explicitly depend on time, nor on the price history except through the current value. We call this the Markov property, but note there are no probabilities involved here yet.

Definition 3.6. *A semi-static hedging strategy (ψ, Δ) is a function $\psi \in \Psi$ and a dynamic strategy $\Delta \in \bar{\Delta}(P)$. The terminal payoff of a semi-static hedging strategy for a price realisation f is*

$$\psi(f(T)) + \sum_{k=0}^{N-1} \delta_{t_k}((f(t))_{0\le t\le t_k})(f(t_{k+1}) - f(t_k)). \tag{9}$$

Without loss of generality we may assume that $\psi'(f(0)+) = 0$. If not then we simply adjust each δ_{t_k} by the quantity $\psi'(f(0)+)$ and the payoff in (9) is unchanged. In the sequel, we will concentrate on the case when Δ is a THMD strategy. Then we identify $\Delta \in \bar{\Delta}(P)$ with $\delta : (0,\infty) \to \mathbb{R}$ and write (ψ, δ) instead of (ψ, Δ).

Given that investments in the forward market may be assumed to be costless, the dynamic strategy has zero price. Thus, in order to define the price of a semi-static hedging strategy it is sufficient to focus on the price associated with the payoff function ψ. The last two terms in (7) are expressed in terms of the payoffs of calls and puts. Thus we can identify the price of $\psi(f(T))$ with the price of a corresponding portfolio of vanilla objects. We also use put-call parity[2] to express the cost of the penultimate term in (7) in terms of call prices. Let $\Psi_0 = \{\psi \in \Psi : \psi'_+(f(0)) = 0\}$, and let $\Psi_c \subseteq \Psi_0$ be the subset of Ψ_0 consisting of the continuously differentiable functions.

Definition 3.7. *The price of a semi-static hedging strategy $(\psi \in \Psi_0, \Delta \in \bar{\Delta}(P))$ is*

$$\psi(f(0)) + \int_{(0,f(0)]} \psi''(x)(C(x) - f(0) + x)dx + \int_{(f(0),\infty)} \psi''(x)C(x)dx.$$

The idea we wish to capture is that the agent holds a static position in calls together with a dynamic position in the underlying such that in combination they provide sub- and super-hedges for the claim.

[2] This means that we do not need to introduce a notation for the put price, which is convenient since P is already in use for the partition. Put-call parity for the forward says that the price of a put with strike x is the price of a call with the same strike minus $f(0) - x$

Definition 3.8. *Let $G = G((f(t_k))_{k=0,...N})$ be the payoff of a path-dependent option. Suppose that there exists a semi-static hedging strategy (ψ,Δ) such that on the partition P*

$$G \leq \textit{(respectively} \geq\textit{)} \ \psi(f(T)) + \sum_{k=0}^{N-1} \delta_{t_k}((f(t))_{0\leq t\leq t_k})(f(t_{k+1}) - f(t_k)).$$

Then (ψ,Δ) is called a semi-static super-hedge (respectively semi-static sub-hedge) for G.

Given a semi-static sub-hedge (respectively super-hedge) we say that the price of the sub-hedge (respectively super-hedge) is a *model independent lower (respectively upper) bound* on the price of the path-dependent claim G.

3.3 Consistent models

The aim of the agent is to construct a hedge which works path-wise, and does not depend on an underlying model. Nonetheless, sometimes it is convenient to introduce a probabilistic model and a stochastic process, and to interpret $f(t)$ as a realisation of that stochastic process. In that case we work with a probability space $(\Omega,\mathcal{F},\mathbb{F} = (\mathcal{F}_t)_{0\leq t\leq T},\mathbb{P})$ supporting the stochastic process $X = (X_t)_{0\leq t\leq T}$.

Definition 3.9. *A model $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ and associated stochastic process $X = (X_t)_{0\leq t\leq T}$ is consistent with the call prices $(C(K))_{K\geq 0}$ if $(X_t)_{t\geq 0}$ is a non-negative $(\mathbb{F},\mathbb{P})$-martingale and if $\mathbb{E}[(X_T - K)^+] = C(K)$ for all $K > 0$.*

In the setting of a stochastic model $V_H(X,P) : \Omega \to \mathbb{R}^+$ is a random variable, and for $\omega \in \Omega$, $V_H(X(\omega),P)$ is a realised value of a variance swap. From a pricing perspective we are interested in getting upper and lower bounds on $\mathbb{E}[V_H(X(\omega),P)]$ as we range over consistent models. Knowledge of call prices is equivalent to knowledge of the marginal law of X_T under a consistent model (Breeden and Litzenberger [4]). If we write μ for the law of X_T and if $C_\mu(K) = \mathbb{E}[(Z_\mu - K)^+]$ where Z_μ is a random variable with law μ, then X is consistent for the call prices C if $C_\mu(K) = C(K)$. We write $m = \int_0^\infty x\mu(dx)$ and we assume, using the martingale property, that $X_0 = m$. Then the problem of characterising consistent models is equivalent to the problem of characterising all martingales with a given distribution at time T.

4 Motivation

4.1 The continuous case

In the situation where both the monitoring and the price-realisations are continuous the theory for the pricing of variance swaps is complete and elegant. We will use this setting to develop intuition for the jump case.

Suppose that the price realisation f is continuous, and possesses a quadratic variation $[f] : [0,T] \to \mathbb{R}^+$ on a dense sequence of partitions $\mathcal{P}$. Dupire [15] and Neuberger [24] independently made the observation that the continuity assumption implies that a variance swap with payoff $\int_0^T f(t)^{-2} d[f]_t$ can be replicated perfectly by holding a static portfolio of log contracts and trading dynamically in the underlying asset. Both Dupire and Neuberger assume $f \equiv X$ is a realisation of a semi-martingale, but in our setting, the observation follows from a path-wise application of Itô's formula in the sense of Föllmer [16], see Section 7. Applying Itô's formula to $-2\log(f(t))$ we have

$$-2\log(f(T)) + 2\log(f(0)) = -2\int_0^T \frac{1}{f(t)} df(t) + \int_0^T \frac{1}{f(t)^2} d[f]_t. \tag{10}$$

Then, as we show in Section 7 below, down a dense sequence of partitions $V_H(f,P_\infty) = \lim_n V_H(f,P^{(n)})$ exists and

$$V_H(f,P_\infty) = \int_0^T \frac{1}{f(t)^2} d[f]_t = -2\log(f(T)) + 2\log(f(0)) + \int_0^T \frac{2}{f(t)} df(t). \tag{11}$$

Provided it is possible to trade continuously and without transaction costs, the rightmost term of this identity has a clear interpretation as the sum of a European contingent claim with maturity T and payoff $-2\log(f(T)/f(0))$ and the gains from trade from a dynamic investment of $2/f(t)$ in the underlying. Alternatively, the rightmost element of (11) can be viewed as the payoff of a semi-static hedging strategy in the continuous time limit for the choice $\psi(x) = -2\log(x/f(0)) + 2(x - f(0))/f(0)$ and $\Delta = (\delta_t)_{0\le t\le T}$ where $\delta_t((f(u))_{0\le u\le t}) = (2/f(t)) - (2/f(0))$. Note that there is equality in (11) so that (ψ, δ) is both a sub- and super-hedge for $V_H(f, P_\infty)$. In particular, under a price continuity assumption, the variance swap has a model-independent price and an associated riskless hedge.

4.2 The effect of jumps on hedging with the classical continuous hedge

Even if the continuity assumption cannot be justified, the associated replication strategy is nevertheless a reasonable candidate for a hedging strategy in the general case. Let us focus on the discrepancy between the payoff of the variance swap and the gains from trade resulting from using the hedge derived in the continuous case. The path-by-path Itô formula continues to apply in the case with jumps, see [16] and Section 7 below. Hence

$$\begin{aligned} -2\log(f(T)) + 2\log(f(0)) &= -2\int_0^T \frac{1}{f(t-)}df(t) + \int_0^T \frac{1}{f(t-)^2}d[f]_t^c \\ &\quad + \sum_{0\le t\le T} 2\left\{\left(\frac{\Delta f(t)}{f(t-)}\right) - \log\left(1 + \frac{\Delta f(t)}{f(t-)}\right)\right\}. \end{aligned}$$

Note that $d[\log(f)]_t = d[f]_t^c/f(t-)^2 + (\Delta\log(f(t)))^2$. By adding and subtracting the discontinuous part of the quadratic variation of $\log(f)$ on the right-hand-side of the above expression, we find

$$-2\log(f(T)) + 2\log f(0) = -2\int_0^T \frac{1}{f(t-)}df(t) + [\log(f)]_T - \sum_{0\le t\le T} J_L(\Delta f(t)/f(t-)) \tag{12}$$

where

$$J_L(\eta) = -2\eta + 2\log(1+\eta) + \log(1+\eta)^2 = -\eta^3/3 + O(\eta^4).$$

It is intuitively clear, but see also Corollary 7.5, that $V_{H^L}(f, P_\infty) \equiv [\log(f)]_T$. Then it follows by re-arrangement of equation (12) that the discrepancy between the realised value of the variance swap $V_{H^L}(f, P_\infty)$ and the return generated by the classical continuous hedging strategy can be represented as the sum of the jump contributions:

$$V_{H^L}(f, P_\infty) - \left(-2\log(f(T)) + 2\log f(0) + 2\int_0^T \frac{1}{f(t-)}df(t)\right) = \sum_{0\le t\le T} J_L\left(\frac{\Delta f(t)}{f(t-)}\right).$$

We call this the hedging error with the convention that if the hedge sub-replicates the variance swap then the hedging error is positive.

Now consider the kernel H^R and define $V_{H^R}(f, P_\infty) = \int_0^T d[f]_t/f(t-)^2$, again, see Corollary 7.5 for justification. By a similar analysis (see also [12, 13]), but adding and subtracting $\left(\frac{\Delta f(t)}{f(t-)}\right)^2$ instead of the discontinuous part of the quadratic variation of $\log(f)$, we have

$$V_{H^R}(f, P_\infty) - \left(-2\log(f(T)) + 2\log(f(0)) + 2\int_0^T \frac{1}{f(t-)}df(t)\right) = \sum_{0\le t\le T} J_R\left(\frac{\Delta f(t)}{f(t-)}\right).$$

where

$$J_R(\eta) = -2\eta + 2\log(1+\eta) + \eta^2 = 2\eta^3/3 + O(\eta^4).$$

In the continuous case, under some mild regularity conditions on f and $\mathcal{P}$, the variance swap value is independent of the chosen kernel. In contrast, the value of a variance swap in the general case is highly dependent on the chosen kernel.

To see that this is the case, and to examine the impact of jumps on the hedging error for the kernels H^L and H^R we consider the shapes of the functions J_R and J_L, see Figure 1. For the kernel H^L, a downward jump results in a positive contribution to the hedging error. Thus, if all jumps are downwards, then the classical continuous hedging strategy sub-replicates $V_{H^L}(f, P_\infty)$. Conversely, upward jumps result in a negative contribution to the hedging error. The story is reversed for the kernel H^R.

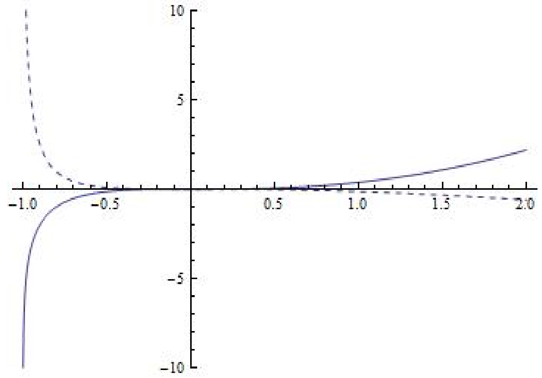

Figure 1: J_L (as represented by the dashed line) is convex decreasing for $x \leq 0$ and concave decreasing for $x \geq 0$. In contrast J_R (solid line) is first concave increasing and then convex increasing. The different shapes of these two curves explains the different nature of the dependence of the payoff of the variance swap on upward and downward jumps for different kernels.

It follows from the argument in the previous paragraph that for the kernel H^L the hedging error will be maximised under scenarios for which the price realisation has downward jumps, but no upward jumps. Paths with this feature might arise as realisations of $-N$ where $N = (N_t)_{t\geq 0}$ is a compensated Poisson process. Moreover, from the convexity of J_L on $(-1, 0)$, it is plausible that the scenarios in which the hedging error is maximised are those in which price realisations have a single large downward jump, rather than a series of small jumps. Again if we wish to minimise the hedging error we should expect a single large upward jump, and the story is reversed for the kernel H^R.

We will use the analysis of this section to give us intuition about the extremal models which will lead to the price bounds on variance swaps derived in the Section 5. The bounds will depend crucially on the kernel. Models under which the variance swap with kernel H^L has highest price (assuming consistency with a given set of call prices) will be characterised by a single downward jump and no upward jumps.

Remark 4.1. *We will see later that the model which minimises the price for variance swaps with kernel H^R also minimises the price for variation swaps with kernel H^S. If f has a quadratic variation, then in the continuous limit $V_{H^S}(f, P_\infty) = \sum_{0<t\leq T}(\Delta f(t))^3$. This payoff will be smallest if all jumps are downwards and we will see that if the call prices are given for expiry time T, then the model that produces the lowest price is one under which the price path has a single downward jump.*

4.3 A related Skorokhod embedding problem

In this section we relate the problem of finding extremal prices for the variance swap to a Skorokhod embedding problem. Let μ be a measure on $\mathbb{R}^+$ with mean m and let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space supporting a right-continuous martingale $X = (X_t)_{0\leq t\leq T}$ such that $X_0 = m$ and $X_T \sim \mu$. Suppose there exists Brownian motion B started at m and a time-change $t \to A_t$, null at 0, such that $X_t = B_{A_t}$. (If X is a continuous martingale, then this is guaranteed by the Dambis-Dubins-Schwarz theorem.) Here B is defined relative to a filtration $\mathbb{G} = (\mathcal{G}_u)_{0\leq u\leq A_T}$ and $\mathcal{G}_{A_t} \supseteq \mathcal{F}_t$. Let A^c be the continuous part of A. Note that $dA^c_t = (dX^c_t)^2 = d[X]^c_t$. Let $S^X = (S^X_t)_{t\geq 0}$ (respectively S) be the process of the running maximum of X (respectively B) so that $S^X_t = \sup_{u\leq t} X_u$. Clearly $X_t \leq S^X_t \leq S_{A_t}$ and then, path-by-path with $\Delta B_{A_t} = B_{A_t} - B_{A_{t-}}$, we have

$$V_{H^R}(X, P_\infty) = \int_0^T \frac{d[X]_t^c}{(X_{t-})^2} + \sum_{0\le t\le T} \left(\frac{\Delta X_t}{X_{t-}}\right)^2 \geq \int_0^T \frac{d[X]_t^c}{(S_{t-}^X)^2} + \sum_{0\le t\le T} \left(\frac{\Delta X_t}{S_{t-}^X}\right)^2 \tag{13}$$

$$\geq \int_0^T \frac{dA_t^c}{(S_{A_{t-}})^2} + \sum_{0\le t\le T} \left(\frac{\Delta B_{A_t}}{S_{A_{t-}}}\right)^2. \tag{14}$$

We suppose, for the moment, that μ has a second moment. Then $(X_t)_{0\le t\le T}$ is a square-integrable martingale and we find that,

$$\begin{aligned}
\mathbb{E}\left[\int_0^T \frac{dA_t^c}{(S_{A_{t-}})^2} + \sum_{0\le t\le T} \left(\frac{\Delta B_{A_t}}{S_{A_{t-}}}\right)^2\right] &= \mathbb{E}\left[\int_0^T \frac{dA_t^c + \Delta A_t}{(S_{A_{t-}})^2}\right] \\
&= \mathbb{E}\left[\int_0^T \frac{dA_t}{(S_{A_{t-}})^2}\right] \\
&\geq \mathbb{E}\left[\int_0^{A_T} \frac{du}{(S_u)^2}\right].
\end{aligned}$$

This motivates looking at the following problem:

$$\min_{\tau\in UI(\mu)} \mathbb{E}\left[\int_0^\tau \frac{du}{S_u^2}\right], \tag{15}$$

where $UI(\mu)$ is the class of stopping times such that $B_\tau \sim \mu$ and $B_{t\wedge\tau}$ is uniformly integrable. This problem is a special case of a problem considered in Hobson and Klimmek [19], where it is proved that the minimum is attained by the Perkins embedding, which we will denote τ_μ^P. Note that the Perkins solution of the Skorokhod embedding problem is generally defined for centred probability measures, but the translation to measures with non-zero mean equal to the non-zero starting point is trivial.

Let $I = (I_t)_{t\ge0}$ denote the infimum process $I_t = \inf_{u\le t} B_u$.

Theorem 4.2. *[Perkins [26], Hobson and Pedersen [20]] Given ν a probability measure with support on $\mathbb{R}^+$, with mean m let Z_ν denote a random variable with law ν and define $C_\nu(z) = \mathbb{E}[(Z_\nu - z)^+]$ and $P_\nu(z) = \mathbb{E}[(z - Z_\nu)^+]$. Define also $\alpha = \alpha_\nu : (m,\infty) \mapsto [0,m)$ and $\beta = \beta_\nu : (0,m) \mapsto (m,\infty]$ by*

$$\alpha(z) = \arg\max_{y<m} \frac{C_\nu(z) - P_\nu(y)}{z-y}, \qquad \beta(z) = \arg\min_{y>m} \frac{P_\nu(z) - C_\nu(y)}{y-z}. \tag{16}$$

Note that α and β are decreasing functions, and where the $\arg\max$ *or* $\arg\min$ *is not uniquely defined we make α and β right-continuous by convention.*

Let B be Brownian motion started at m, with maximum process S and minimum process I. Suppose μ has no atom at m. Then $\tau_\nu^P := \inf\{u > 0 : B_u < \alpha_\nu(S_u)$ or $B_u > \beta_\nu(I_u)\}$ solves the Skorokhod embedding problem for ν in the sense that $B_{\tau_\nu^P} \sim \nu$ and $(B_{t\wedge\tau_\nu^P})_{t\ge0}$ is uniformly integrable.

If ν has an atom at m then we assume $\mathcal{F}_0$ is sufficiently rich as to support a uniform random variable $\tilde{Z}_U$, which is independent of B. Then

$$\tau_\nu^P := \begin{cases} 0 & \tilde{Z}_U \le \nu(\{m\}) \\ \inf\{u > 0 : B_u < \alpha_\nu(S_u) \text{ or } B_u > \beta_\nu(I_u)\} & \tilde{Z}_U > \nu(\{m\}) \end{cases}$$

solves the Skorokhod embedding for ν.

The salient characteristic of the Perkins embedding which results in optimality in (15) is that either $B_{\tau_\nu^P} = S_{\tau_\nu^P}$ or $B_{\tau_\nu^P} = \alpha_\nu(S_{\tau_\nu^P})$, and then at τ_ν^P the Brownian motion B is either at a new maximum, or a new minimum.

Now consider the problem of finding the consistent model for which $V_{H^R}(X, P_\infty)$ has lowest possible price, and recall that knowledge of call prices is equivalent to knowledge of the marginal law μ of X_T. To obtain the lowest possible price we might expect equality in each of (13)-(14), and thus that just before a jump, the process is at its current maximum. Moreover, the model should be related to the Perkins embedding.

Lemma 4.3. *Let B be Brownian motion started at m. Let $\overline{H}_b = \inf\{u \geq 0 : B_u = b\}$ be the first hitting time of level b by Brownian motion. Let $\Lambda(t)$ be a strictly increasing, continuous function such that $\Lambda(0) = m$ and $\lim_{t\uparrow T}\Lambda(t)$ is infinite.*

Define the process $\tilde{Q}^\mu = (\tilde{Q}^\mu_t)_{0\leq t\leq T}$ by

$$\tilde{Q}^\mu_t = B_{\overline{H}_{\Lambda(t)}\wedge\tau^P_\mu}, \tag{17}$$

and let Q^μ be the right-continuous modification of $\tilde{Q}^\mu$.

Then, Q^μ is a martingale such that $Q^\mu_T \sim \mu$. Moreover, the paths of Q^μ are continuous and increasing, except possibly at a single jump time. Finally, either $Q^\mu_T \equiv B_{\tau^P_\mu} = S_{\tau^P_\mu}$ or $Q^\mu_T \equiv B_{\tau^P_\mu} = \alpha_\mu(S_{\tau^P_\mu})$.

Proof. Since τ^P_μ is finite almost surely we have that $Q^\mu_T \equiv B_{\tau^P_\mu} \sim \mu$. Moreover, for $\Lambda(t) < \tau^P_\mu$, $Q^\mu_t = \Lambda(t) = B_{\overline{H}_{\Lambda(t)}} = S_{\overline{H}_{\Lambda(t)}}$. □

The martingale Q^μ will be used in Section 7 to show that in the continuous-time limit, the bounds we obtain are tight. The martingale Q^μ is the related to the Perkins embedding in the same way that the Dubins-Gilat [14] martingale is related to the Azéma-Yor [1] embedding.

We can also consider a reflected version of the martingale Q^μ based on the infimum process rather than the maximum process.

Lemma 4.4. *Let $\lambda(t)$ be a strictly decreasing, continuous function such that $\lambda(0) = m$ and $\lim_{t\uparrow T}\lambda(t)$ is zero.*

Define the process $\tilde{R}^\mu = (\tilde{R}^\mu_t)_{0\leq t\leq T}$ by

$$\tilde{R}^\mu_t = B_{\overline{H}_{\lambda(t)}\wedge\tau^P_\mu}, \tag{18}$$

and let R^μ be the right-continuous modification of $\tilde{R}^\mu$.

Then, R^μ is a martingale such that $R^\mu_T \sim \mu$. Moreover, the paths of R^μ are continuous and decreasing, except possibly at a single jump time. Finally, either $R^\mu_T \equiv B_{\tau^P_\mu} = I_{\tau^P_\mu}$ or $R^\mu_T \equiv B_{\tau^P_\mu} = \beta_\mu(I_{\tau^P_\mu})$.

Remark 4.5. *In this section we have exploited a connection between the problem of finding bounds on the prices of variance swaps and the Skorokhod embedding problem. This link is one of the recurring themes of the literature on the model-independent bounds, see Hobson [18]. We exhibit this link for the kernel H^R, and in this sense at least, it seems that variance swaps defined via H^R are the more natural mathematical object. Nonetheless, the intuition developed via H^R and the Skorokhod embedding problem is valid more widely.*

5 Path-wise Bounds for Variance Swaps

Previous sections have defined notation and developed intuition for the problem. Now we begin the construction of path-wise hedging strategies. We do this by defining a class of synthesisable payoffs with a useful extra property which can be exploited to give sub-hedges. Then, motivated by the results of Section 4.3 on the relation between the variance swap and the Perkins solution of the Skorokhod embedding problem, we define a further sub-class of payoffs which are based on decreasing functions.

To construct a sub-hedge for a variation swap with kernel H for any price realisation f, suppose that there exists a pair of functions (ψ, δ) such that for $x, y \in \mathbb{R}$

$$H(x, y) \geq \psi(y) - \psi(x) + \delta(x)(y - x). \tag{19}$$

Then we may interpret (ψ, δ) as a semi-static hedging strategy (for a time-homogeneous Markov dynamic strategy δ) and then for any price realisation f and partition P,

$$V_H(f, P) \geq \psi(f(T)) - \psi(f(0)) + \sum_k \delta(f(t_k))(f(t_{k+1}) - f(t_k)). \tag{20}$$

By Definition 3.8 we have constructed a sub-hedge for the variation swap with kernel H.

If (ψ, δ) satisfies (19) then so does $(\psi + a + by, \delta - b)$ for any constants a, b. Earlier we argued that without loss of generality for a semi-static hedging strategy we could assume $\psi'_+(f(0)) = 0$. Now we may restrict attention further to ψ with $\psi(f(0)) = 0$. Let $\Psi_{0,0} = \{\psi \in \Psi : \psi(f(0)) = 0 = \psi'(f(0)+)\}$.

Suppose now that H is a variance swap kernel, and that ψ is continuously differentiable. Recall that $H_y(x,x) = 0$. Dividing both sides of (19) by $y-x$ and letting $y \downarrow x$, we find that $\delta(x) \le -\psi'(x)$. Similarly letting $y \uparrow x$, $\delta(x) \ge -\psi'(x)$. Thus if (19) is to hold we must have that $\delta \equiv -\psi'$ and in the continuously differentiable case our search for pairs of functions satisfying (19) is reduced to finding functions ψ satisfying

$$H(x,y) \ge \psi(y) - \psi(x) - \psi'(x)(y-x), \tag{21}$$

or equivalently, $\psi(y) \le H(x,y) + \psi(x) + \psi'(x)(y-x)$. Note that there is equality in (21) at $y = x$.

It remains to show how to choose ψ solving (21). Using the intuition developed in the previous section for the kernel H^R we expect optimal sub-hedging strategies to be associated with the martingale Q defined in (17). For realisations of Q, either the path has no jump, or there is a single jump, and if the jump occurs when the process is at x then the jump is to $\alpha(x)$, where α is a decreasing function.

With this in mind let $\kappa : [f(0),\infty) \to (0,f(0)]$ be a decreasing function with a continuously differentiable inverse k. Fix $y < f(0)$ and consider varying x over $x < f(0)$. We want there to be equality in (21) at $x = k(y)$, and then also the x-derivatives of both sides of (21) must match. Then ψ must satisfy

$$\psi(y) = H(k(y),y) + \psi(k(y)) + \psi'(k(y))(y-k(y)), \tag{22}$$

and moreover if ψ' is differentiable, we must have $H_x(k(y),y) + \psi''(k(y))(y-k(y)) = 0$ or equivalently

$$\psi''(x) = H_x(x,\kappa(x))/(x-\kappa(x)). \tag{23}$$

This suggests that we can define candidate sub-hedge payoffs ψ via (23) on $(f(0),\infty)$ and via (22) on $(0,f(0))$.

Now we wish to extend these arguments to the case where ψ and κ need not be so regular. Suppose that left- and right-derivatives of ψ exist. By the arguments above we find that if (19) is to hold then $-\psi'(x-) \le \delta(x) \le -\psi'(x+)$. It does not matter which δ we choose in this interval, but for definiteness we take $\delta = -\psi'_+$. Recall the discussion after (21).

Definition 5.1. *$\psi \in \Psi$ is a candidate sub-hedge payoff if for all $y \in (0,\infty)$,*

$$\psi(y) = \inf_x \{H(x,y) + \psi'(x+)(y-x) + \psi(x)\}. \tag{24}$$

Given a candidate sub-hedge payoff ψ we can generate a candidate semi-static hedge (ψ,δ) by taking $\delta(x) = -\psi'(x+)$. We will say that ψ is the root of the semi-static sub-hedge $(\psi,-\psi'_+)$.

Let $\mathcal{K} = \mathcal{K}(f(0))$ be the set of monotone decreasing right-continuous functions $\kappa : [f(0),\infty) \to (0,f(0)]$, with $\kappa(f(0)) = f(0)$ and let k denote the right-continuous inverse to κ.

Define $\Phi(u,y) = H_x(u,y)/(u-y)$. Write $\Phi^R(u,y) = H^R_x(u,y)/(u-y)$, and similarly for other kernels.

Definition 5.2. *For $\kappa \in \mathcal{K}$ with inverse k, define $\psi_{\kappa,H} \equiv \psi_\kappa : (0,\infty) \mapsto \mathbb{R}^+$, by $\psi_\kappa(f(0)) = 0$ and*

$$\psi_\kappa = \left\{ \begin{array}{c} \psi_\kappa(x) \\ \psi_\kappa(z) \end{array} \right\} = \left\{ \begin{array}{ll} \int_{f(0)}^x (x-u)\Phi(u,\kappa(u))du & x > f(0) \\ \psi_\kappa(k(z)) + \psi'_\kappa(k(z))(z-k(z)) + H(k(z),z) & z < f(0) \end{array} \right. \tag{25}$$

We call such a function a candidate payoff of Class $\mathcal{K}$.

By convention we use the variable x on $(f(0),\infty)$ and z on $(0,f(0))$, to reflect the fact that ψ is defined explicitly on the former set, but only implicitly on the latter.

Remark 5.3. *For $x > f(0)$ we have $\psi'_\kappa(x) = \int_{f(0)}^x \Phi(u,\kappa(u))du$. Note that ψ'_κ is continuous on $[f(0),\infty)$.*

For $z < f(0)$ we have $\psi'_\kappa(z+) = \psi'_\kappa(k(z)) + H_y(k(z),z)$ so that ψ'_κ is continuous at z if k is continuous there.

If H is a regular variation kernel then it is straightforward to show that ψ_κ defined via (25) is the difference of two convex functions, and therefore that $\psi_\kappa \in \Psi_{0,0}$.

For the present we fix κ and we write simply ψ for ψ_κ. Note that the value of $\psi(x)$ does not depend on the right-continuity assumption for κ. Further, as we now argue, nor does it depend on the right-continuity assumption of the inverse k. Observe that if κ is not injective and there is an interval $A_z \equiv \{x : \kappa(x) = z\} \subseteq (f(0),\infty)$ over which κ takes the value z, then k has a jump at z. Nonetheless, the value of $\psi(z)$ does not depend on the choice of $k(z)$ within the interval A_z. To see this, for $x \in A_z$ consider $\Psi(x) := \psi(x) + \psi'(x)(z-x) + H(x,z)$. Then, on A_z, $d\Psi/dx = \psi''(x)(z-x) + H_x(x,z) \equiv 0$, using (23).

Motivated by the results of Section 4.3 we have defined ψ relative to the set of decreasing functions $\mathcal{K}$ with the aim of constructing a sub-hedge. However, there are analogous definitions based on constructing super-hedges or using the martingale R or both.

Definition 5.4. *$\psi \in \Psi$ is a candidate super-hedge payoff if for all $y \in (0,\infty)$,*

$$\psi(y) = \sup_x \{H(x,y) + \psi'_+(x)(y-x) + \psi(x)\}. \tag{26}$$

Define $\mathcal{L} = \mathcal{L}(f(0))$ be the set of monotone decreasing functions $\ell : (0, f(0)) \to (f(0),\infty)$, with $\ell(f(0)) = f(0)$. Let l be inverse to ℓ.

Definition 5.5. *For $\ell \in \mathcal{L}$ with inverse l, define $\psi_\ell : (0,\infty) \mapsto \mathbb{R}^+$, the candidate payoff of Class $\mathcal{L}$ by $\psi_\ell(f(0)) = 0$ and*

$$\psi_\ell = \left\{ \begin{array}{c} \psi_\ell(x) \\ \psi_\ell(z) \end{array} \right\} = \left\{ \begin{array}{ll} \int_x^{f(0)} (u-x)\Phi(u,\ell(u))du & x < f(0); \\ \psi_\ell(l(z)) + \psi'_\ell(l(z))(z-l(z)) + H(l(z),z) & z > f(0). \end{array} \right.$$

Our next aim is to give conditions which guarantee that the semi-static strategy $(\psi, -\psi'_+)$ satisfies equation (19).

Definition 5.6. *A variation swap kernel H is an* increasing *(a* decreasing*) kernel if it is a regular variation swap kernel and*

(i) $\Phi(u,y)$ is monotone increasing (decreasing) in y,

(ii) $H(a,b) + H_y(a,b)(c-b) \geq (\leq) H(a,c) - H(b,c)$ for all $a > b$.

Remark 5.7. *A sufficient condition for the second condition in Definition 5.6 is that $H_{yy}(x,y)$ is decreasing (increasing) in its first argument.*

Example 5.8. *H^R and H^S are increasing kernels and H^L is a decreasing kernel. The kernels H^B and H^Q are simultaneously both increasing and decreasing since $\Phi^B(u,y) = 2u^{-2}$ and $\Phi^Q(u,y) = 2$ do not depend on y and Condition (ii) in Definition 5.6 is satisfied with equality in both cases.*

Example 5.9. *Consider the kernels $H^{G-}(u,y) = uH^R(u,y)$ and $H^{G+}(u,y) = yH^R(u,y)$. In the first case, variance is weighted by the pre-jump value of the price realisation and in the second case the variance is weighted by the post-jump value. Swaps of this type are known as Gamma swaps, see, for example, Carr and Lee [8]. Both H^{G-} and H^{G+} are increasing kernels.*

Theorem 5.10. *(i) (a) If H is an increasing kernel then every candidate payoff of Class $\mathcal{K}$ is the root of a semi-static sub-hedge for the kernel H. In particular, if H is an increasing variance swap kernel, and if ψ_κ is as given in (25) then we can construct a model-independent sub-hedge for the variance swap in the sense of an identity (20).*

(b) If H is an increasing kernel then every candidate payoff of Class $\mathcal{L}$ is the root of a semi-static super-hedge for the kernel H.

(ii) (a) If H is a decreasing kernel then every candidate payoff of Class $\mathcal{L}$ is the root of a semi-static sub-hedge for the kernel H.

(b) If H is a decreasing kernel then every candidate payoff of Class $\mathcal{K}$ is the root of a semi-static super-hedge for the kernel H.

Proof. We will prove the theorem in the case (i)(a). The proofs in the other cases are similar.

Fix $\kappa \in \mathcal{K}$ let $L_\kappa(x,y) = \psi_\kappa(x) + \psi'_\kappa(x+)(y-x) + H(x,y) - \psi_\kappa(y)$. The result will follow if we can show that $L_\kappa(x,y) \geq 0$ for all $(x,y) \in (0,\infty)^2$. Since κ is fixed we drop the subscript κ in what follows.

Suppose that $x,z > f(0)$ and $y \in (0,\infty)$. Since ψ is continuously differentiable on $(f(0),\infty)$ and since $\psi(x)+\psi'(x)(y-x) = \int_{f(0)}^{x}(y-u)\Phi(u,\kappa(u))du$ we have that

$$\begin{aligned}
L(x,y) - L(z,y) &= \psi(x)+\psi'(x)(y-x)+H(x,y)-\psi(z)-\psi'(z)(y-z)-H(z,y) \\
&= \int_z^x \{(y-u)\Phi(u,\kappa(u))+H_x(u,y)\}\,du \\
&= \int_z^x \{\Phi(u,y)-\Phi(u,\kappa(u))\}(u-y)du.
\end{aligned}$$

If $y \geq f(0)$, then set $z = y$ to find that

$$L(x,y) = \int_y^x \{\Phi(u,y)-\Phi(u,\kappa(u))\}(u-y)du.$$

Since $y \geq f(0) \geq \kappa(u)$, $\Phi(u,y) \geq \Phi(u,\kappa(u))$ for all u. Hence $L(x,y) \geq 0$ with equality at $y = x$.

For $y < f(0)$ we have $L(k(y),y) = 0$ and

$$L(x,y) = \int_{k(y)}^x \{\Phi(u,y)-\Phi(u,\kappa(u))\}(u-y)du.$$

If $k(y) < x$ then $y > \hat{y}$, for all $\hat{y} \in [\kappa(x+),\kappa(x-)]$. Then for $u \in (k(y),x)$, $\kappa(u) \leq y$ and since $\Phi(u,z)$ is increasing in z, the integrand is positive.

If $x < k(y)$, then $y < \hat{y}$ for all $\hat{y} \in [\kappa(x+),\kappa(x-)]$. Then for $u \in (x,k(y))$ we have $\kappa(u) > y$. Then again $L(x,y) \geq 0$.

Finally, we show that $L(x,y) \geq 0$ when $x < f(0)$. Note that since, by what we have shown above, $L(k(x),y) \geq 0$ it will suffice to show that $L(x,y) \geq L(k(x),y)$. But,

$$\begin{aligned}
L(x,y) - L(k(x),y) &= \psi(x)+\psi'(x+)(y-x)+H(x,y)-\psi(k(x))-\psi'(k(x))(y-k(x))-H(k(x),y) \\
&= \psi(k(x))+\psi'(k(x))(x-k(x))+H(k(x),x)+\psi'(k(x))(y-x) \\
&\quad +H_y(k(x),x)(y-x)+H(x,y)-\psi(k(x))-\psi'(k(x))(y-k(x))-H(k(x),y) \\
&= H(k(x),x)+H(x,y)+H_y(k(x),x)(y-x)-H(k(x),y) \\
&\geq 0,
\end{aligned}$$

where the last inequality follows from Definition (5.6). □

Remark 5.11. *Suppose we use a different convention for $\delta(x)$ whilst still respecting the inequality $-\psi'(x-) \leq \delta(x) \leq -\psi'(x+)$. Recall that ψ' can only have jumps on $(0, f(0)]$. Then the proof of Theorem 5.10(i)(a) goes through unchanged except that in the last step we choose $v \in [k(z),k(z-)]$ such that $-\delta(z) = \psi'(v)+H_y(v,z)$ (such a choice is possible by the intermediate value theorem). Then $L(v,y) \geq 0$ and*

$$L(z,y) - L(v,y) = \psi(z) - \delta(z)(y-z) + H(z,y) - \psi(v) - \psi'(v)(y-v) - H(v,y).$$

Substituting for $\psi(z) = \psi(v)+\psi'(v)(z-v)+H(v,z)$ and $\delta(z)$ we again obtain $L(z,y)-L(v,y) \geq 0$ using Definition (5.6).

6 The most expensive sub-hedge

In the next three sections we concentrate on lower bounds and increasing variance kernels, but there are equivalent results for upper bounds and/or decreasing variance kernels.

In this section we fix the call prices and attempt to identify the most expensive sub-hedge from the set of sub-hedges generated by candidate payoffs of Class $\mathcal{K}$. The price of this sub-hedge provides a highest model-independent lower bound on the price of the variance swap in a sense which we will explain in the section on continuous limits.

Associated with the set of call prices $C(k)$ (and put prices $C(k) - f(0) + k$ given by put-call parity) there is a measure μ on $\mathbb{R}^+$ with mean m. Since f is a forward price we must have $f(0) = m$. Write $C = C_\mu$

to emphasise the connection between these quantities. Then $C(k) = C_\mu(k) = \int_k^\infty (x-k)\mu(dx)$. Recall that C_μ is convex so that $\mu(dx) = C''_\mu(x)dx$ with the right-hand-side to be interpreted in a distributional sense as necessary. We wish to calculate the cost of the European claim which forms part of the semi-static sub-hedge. By construction this is equal to $\int_{\mathbb{R}^+} \psi(x)\mu(dx) = \int_0^m \psi''(z)(C_\mu(z) - m + z)dz + \int_m^\infty \psi''(x)C_\mu(x)dx$.

Proposition 6.1. *For H a variance swap kernel and $\kappa \in \mathcal{K}(m)$,*

$$\int_0^\infty \psi_\kappa(x)\mu(dx) = \int_0^m \mu(dz)H(m,z) + \int_m^\infty du \Sigma_\mu^{(u)}(\kappa(u)) \tag{27}$$

where, for $v < m < u$,

$$\Sigma_\mu^{(u)}(v) = \Phi(u,v)C_\mu(u) + \int_{(0,v]} \mu(dz)(u-z)\{\Phi(u,z) - \Phi(u,v)\}.$$

Proof. Let $\psi = \psi_\kappa$. Note that by definition $\psi(m) = 0$, so there is no contribution from mass at m and we can divide the integral on the left of (27) into intervals $(0,m)$ and (m,∞). For the latter,

$$\begin{aligned}\int_m^\infty \psi(x)\mu(dx) &= \int_m^\infty \mu(dx)\int_m^x (x-u)\Phi(u,\kappa(u))du \\ &= \int_{u=m}^\infty du\Phi(u,\kappa(u))\int_u^\infty (x-u)\mu(dx) \\ &= \int_{u=m}^\infty du\Phi(u,\kappa(u))C_\mu(u) =: I_1.\end{aligned}$$

Now consider $\int_0^m \psi(z)\mu(dz)$. For this, using $H(k,z) = H(m,z) + \int_m^k H_x(u,z)du$ and $\psi(x) + \psi'(x)(z-x) = \int_m^x du(z-u)\Phi(u,\kappa(u))$ we have

$$\begin{aligned}\int_0^m \psi(z)\mu(dz) &= \int_0^m \mu(dz)H(m,z) + \int_0^m \mu(dz)\int_m^{k(z)} du(u-z)\{\Phi(u,z) - \Phi(u,\kappa(u))\} \\ &=: I_2 + I_3.\end{aligned}$$

Note that I_2 depends on H but not on κ. Moreover, I_3 does not depend on the particular values chosen for the inverse taken over intervals of constancy of κ. (If $x < \tilde{x}$ are a pair of possible values for $k(z)$ then $\int_x^{\tilde{x}} du(u-z)\{\Phi(u,z) - \Phi(u,\kappa(u))\} = 0$ since over this range $\kappa(u) = z$.) Changing the order of integration we have

$$I_3 = \int_m^\infty du \int_{(0,\kappa(u)]} \mu(dz)(u-z)\{\Phi(u,z) - \Phi(u,\kappa(u))\},$$

and then $I_1 + I_3 = \int_m^\infty du\Sigma_\mu^{(u)}(\kappa(u))$. □

Our goal is to maximise the expression (27) over decreasing functions $\kappa \in \mathcal{K}$. As noted above, I_2 is independent of κ, and to maximise $\int_m^\infty du\Sigma_\mu^{(u)}(\kappa(u))$ we can maximise $\Sigma_\mu^{(u)}(\kappa)$ separately for each $u > m$, and then check that the maximiser is a decreasing function of u. First we give a useful lemma.

Lemma 6.2. *For $u > m$ and $v \in (0,u)$ define $\Theta_\mu^{(u)}(v) := C_\mu(u) - \int_{(0,v]} \mu(dz)(u-z)$. Define $\overline{\kappa} = \overline{\kappa}(u) = \sup\{v : \Theta_\mu^{(u)}(v) \geq 0\}$. Then $\overline{\kappa}(u) = \alpha(u)$ where $\alpha = \alpha(\mu)$ is the quantity which arises in (16) in the definition of the Perkins solution to the Skorokhod embedding problem.*

Proof. For each u, $\Theta_\mu^{(u)}$ is a strictly decreasing right-continuous function taking both positive and negative values on $(0,m)$. We have $\Theta_\mu^{(u)}(\overline{\kappa}(u-)) \geq 0 \geq \Theta_\mu^{(u)}(\overline{\kappa}(u+))$.

If P_μ is differentiable at v then $\int_{(0,v]} \mu(dz)(u-z)$ is the value at u of the tangent to P_μ at v, and $\alpha(u)$ is the smallest v for which this value is greater than or equal to $C_\mu(u)$.

The definition $\overline{\kappa}(u) = \sup\{v : \Theta_\mu^{(u)}(v) \geq 0\}$ ensures that $\overline{\kappa}$ is right continuous, and then $\overline{\kappa} = \alpha$ as required. □

Theorem 6.3. *Suppose H is an increasing variance swap kernel. Then the set of lower price bounds $\{\int_0^\infty \psi_\kappa(x)\mu(dx); \kappa \in \mathcal{K}$ is maximised over $\kappa \in \mathcal{K}$ by $\kappa = \alpha(\mu)$.*

Proof. Suppose H is an increasing variance swap kernel so that $\Phi(u,y)$ is increasing in y. We want to show that for each u $\Sigma_\mu^{(u)}(v)$ is maximised by $v = \overline{\kappa}(u) = \alpha(u)$. Then since α is decreasing we have that $\overline{\kappa} = \alpha$ maximises $\int_0^\infty \psi_\kappa(x)\mu(dx)$ over $\kappa \in \mathcal{K}$

Suppose $m > v > \overline{\kappa}(u)$. We aim to show that for all $\kappa \in (\overline{\kappa}(u), v)$ we have $\Sigma_\mu^{(u)}(v) \le \Sigma_\mu^{(u)}(\kappa)$. We have

$$\begin{aligned}
\Sigma_\mu^{(u)}(v) - \Sigma_\mu^{(u)}(\kappa) &= \Phi(u,v)C_\mu(u) + \int_0^v \mu(dz)(u-z)\{\Phi(u,z) - \Phi(u,v)\} \\
&\quad -\Phi(u,\kappa)C_\mu(u) - \int_0^\kappa \mu(dz)(u-z)\{\Phi(u,z) - \Phi(u,\kappa)\} \\
&= \int_\kappa^v \mu(dz)(u-z)\{\Phi(u,z) - \Phi(u,v)\} + [\Phi(u,v) - \Phi(u,\kappa)]\,\Theta_\mu^{(u)}(\kappa).
\end{aligned}$$

Since H is an increasing variance kernel, for $z \in (\kappa, v)$, $\Phi(u,z) \le \Phi(u,v)$, and the first integral is non-positive. Furthermore, $\Phi(u,v) \ge \Phi(u,\kappa)$ and $\Theta^{(u)}(\kappa) < 0$. Hence we conclude that $\Sigma_\mu^{(u)}(v) \le \Sigma_\mu^{(u)}(\kappa)$.

Similar arguments show that if $v < \overline{\kappa}(u)$ then $\Sigma_\mu^{(u)}(v) \le \Sigma_\mu^{(u)}(\kappa)$ for any $\kappa \in (v, \overline{\kappa}(u))$, and it follows that $\kappa = \overline{\kappa}(u)$ is a maximiser of $\Sigma_\mu^{(u)}(v)$. □

Corollary 6.4. *Suppose $\kappa_n(x)$ is a sequence of elements of $\mathcal{K}$ with $\kappa_n(x) \downarrow \alpha(x) \equiv \alpha_\mu(x)$. Then $\int_{[0,\infty)} \psi_{\kappa_n}(x)\mu(dx)$ converges monotonically to $\int_{[0,\infty)} \psi_\alpha(x)\mu(dx)$.*

Proof. Recall that $\int_{[0,\infty)} \psi_\kappa(x)\mu(dx) = \int_0^1 \mu(dz)H(1,z) + \int_1^\infty du\Sigma_\mu^{(u)}(\kappa(u))$. By the above arguments we have that $\Sigma_\mu^{(u)}(z)$ is increasing in z for $z > \overline{\kappa}(u)$. Hence the result follows by monotone convergence. □

Example 6.5. *Let $H = H^R$, an increasing variance kernel. Let $\mu = U[0,2]$ and let $\kappa : [1,2] \to [0,1]$ be given by $\kappa(x) = \alpha_\mu(x) = x - 2\sqrt{x-1}$. Similarly we define $\ell : [0,1] \to [1,2]$ by $\ell(x) = \beta_\mu(x) = x + 2\sqrt{1-x}$. Then ψ_κ is the most expensive sub-hedge of class $\mathcal{K}$ and ψ_ℓ is the cheapest super-hedge of class $\mathcal{L}$. Although we cannot calculate the functions ψ_κ, ψ_ℓ explicitly, they can be evaluated numerically, see the left hand side of Figure 2. Now suppose $H = H^L$. The roles of ψ_κ and ψ_ℓ are reversed (see the right hand side of Figure 2) and ψ_κ is the root of a semi-static super-hedge and ψ_ℓ is the root of a semi-static sub-hedge.*

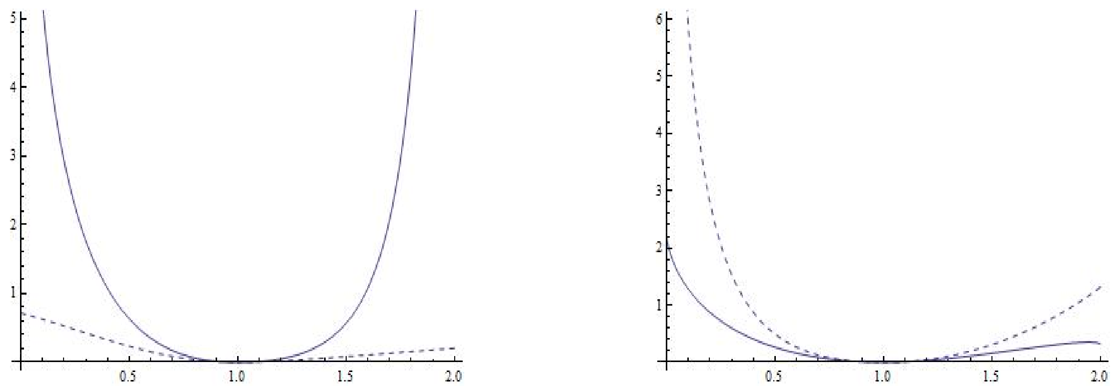

Figure 2: For the two kernels ψ_κ is shown as a dashed line and ψ_ℓ is shown as a solid line. For the kernel H^R (left-hand-side), ψ_κ is associated with a lower bound on the price of the variance swap. For the kernel H^L (right-hand-side) ψ_κ is associated with an upper bound.

7 Continuous limits and the tightness of the bound

The bounds we have constructed based on the functions ψ_κ hold simultaneously across all paths and all partitions. The purpose of this section is to consider the limit as the partition becomes finer. It will turn out that in the continuous limit there is a stochastic model which is consistent with the observed call prices and for which there is equality in the inequality (19) from which we derive the lower bound. In this sense the model-free bound is optimal, and can be attained.

The analysis of this section justifies restricting attention to candidate payoffs of Classes $\mathcal{K}$ and $\mathcal{L}$. Hedges of this type either sub-replicate or super-replicate the payoff of the variance swap depending on the form of the kernel, but there could be other sub- and super-replicating strategies which do not take this form. In principle, for a given partition one of these other sub-hedges could give a tighter model-independent bound than we can derive from our analysis. (As an extreme example, suppose the partition is trivial ($0 = t_0 < t_1 = T$). Then $V_H(f,P) = H(f(0), f(T))$ which can be replicated exactly using call options.) However, in the continuous limit our bound is best possible, so that when the partition is finite, but the mesh size is small we expect our hedge to be close to best possible and relatively simple to implement. In Remark 7.11 we make some remarks on the rate of convergence.

For a finite partition $P^{(n)}$ in the dense sequence $\mathcal{P} = (P^{(n)})_{n\geq 1}$ we have

$$V_H(f,P^{(n)}) = \sum_{k=0}^{N^{(n)}-1} H(f(t_k), f(t_{k+1})) \geq \psi(f(T)) - \psi(f(0)) - \sum_{k=0}^{N^{(n)}-1} \psi'(f(t_k))(f(t_{k+1}) - f(t_k)). \tag{28}$$

We want to conclude that the limits $V_H(f,P_\infty) = \lim_n V_H(f,P^{(n)})$ and

$$\lim_n \sum_{k=0}^{N^{(n)}-1} \psi'(f(t_k))(f(t_{k+1}) - f(t_k)) = \int_0^T \psi'(f(t-))df(t) \tag{29}$$

exist for each path under consideration. Our analysis follows the development of a path-wise Itô's formula in Föllmer [16]. Let ϵ_t denote a point mass at t.

Definition 7.1. *A path realisation f has a quadratic variation on a dense sequence of partitions $\mathcal{P} = (P^{(n)})_{n\geq 1}$ if, when we define the measure*

$$\zeta_n = \sum_{k=0,\ t_k\in P^{(n)}}^{N^{(n)}-1} (f(t_{k+1}) - f(t_k))^2 \epsilon_{t_k},$$

then the sequence ζ_n converges weakly to a Radon measure ζ on $[0,T]$. Then $([f]_t)_{t\geq 0}$ is given by $[f]_t = \zeta([0,t])$.

The atomic part of ζ is given by squared jumps of f. Moreover the quadratic variation $([f]_t)_{t\geq 0}$ is simply the cumulative mass function of ζ.

Theorem 7.2. *(Föllmer [16]) Suppose the price realisation f has a quadratic variation along a dense sequence of partitions $\mathcal{P} = (P^{(n)})_{n\geq 1}$ and G is a twice continuously differentiable function from $\mathbb{R}^+$ to $\mathbb{R}$, then*

$$\int_0^T G'(f(t-))df(t) = \lim_{n\uparrow\infty} \sum_{t=0}^{N^{(n)}-1} G'(f(t_k))(f(t_{k+1}) - f(t_k))$$

exists and

$$\begin{aligned} G(f(T)) - G(f(0)) \quad &= \quad \int_0^T G'(f(s-))df(s) + \frac{1}{2}\int_{(0,T]} G''(f(s))d[f]^c_s \\ &\quad + \sum_{s\leq T} \left[G(f(s)) - G(f(s-)) - G'(f(s-))\Delta f(s)\right], \end{aligned}$$

and the series of jump terms is absolutely convergent.

Hence, provided ψ is twice continuously differentiable on the support of f and f has a quadratic variation along $\mathcal{P}$, it follows immediately that the limit in (29) exists. In our setting $\psi_\kappa''(u) = \Phi(u, \kappa(u))$ for $u > 1$, so that a sufficient condition for $\psi_\kappa''(u)$ to be continuous on $(1, \infty)$ is that κ is continuous. Further, on $u < 1$, provided $k \equiv \kappa^{-1}$ is differentiable and H_y exists, we have $\psi'(z) = \psi'(k(z)) + H_y(k(z), z)$. Hence, sufficient conditions for ψ to be twice continuously differentiable on $(0,1)$ are that k is continuously differentiable, κ is continuous and H_{xy} and H_{yy} are continuous. Let $\mathcal{K}_c$ be the class of decreasing functions $\kappa : (f(0), \infty) \to (0, f(0))$ which are continuous and have an inverse k which is continuously differentiable.

Corollary 7.3. *Suppose that H is an increasing variance kernel, and that f has a quadratic variation along a dense sequence of partitions* $\mathcal{P} = \mathcal{P} = (P^{(n)})_{n\geq 1}$. *Suppose* $\kappa \in \mathcal{K}_c$ *and* $\psi = \psi_\kappa$. *Then the limit in (29) exists.*

Now we want to consider $V_H(f, P_\infty) = \lim_n V_H(f, P^{(n)})$.

Lemma 7.4. *Suppose H is a variance swap kernel. If* $\mathcal{P} = (P^{(n)})_{n\geq 1}$ *is a dense sequence of partitions, and f has a quadratic variation along* $\mathcal{P}$, *then* $\lim_{n\uparrow\infty} V_H(f, P^{(n)})$ *exists and satisfies*

$$V_H(f, P_\infty) = \int_{(0,T]} \frac{1}{f(t-)^2} d[f]_t + \sum_{0<t\leq T} H(f(t-), f(t)) - \sum_{0<t\leq T} \frac{1}{f(t-)^2} (\Delta f(t))^2. \tag{30}$$

Proof. Our proof follows Föllmer [16]. Fix $\epsilon > 0$. Partition $[0, T]$ into two classes: a finite class $C_1 = C_1(\epsilon)$ of jump times and a class $C_2 = C_2(\epsilon)$ such that

$$\sum_{s\in[0,T],\ s\in C_2(\epsilon)} (\Delta f(s))^2 \leq \epsilon^2. \tag{31}$$

Then $\sum_{k=0}^{N^{(n)}-1} H(f(t_k), f(t_{k+1})) = \sum_1 H(f(t_k), f(t_{k+1})) + \sum_2 H(f(t_k), f(t_{k+1}))$, where $\sum_1$ indicates a sum over those $0 \leq k \leq N^{(n)} - 1$ for which $(t_k, t_{k+1}]$ contains a jump of class C_1. It follows that

$$\lim_{n\uparrow\infty} \sum_1 H(f(t_k), f(t_{k+1})) = \sum_{t\in C_1(\epsilon)} H(f(t-), f(t)). \tag{32}$$

On the other hand, using the properties $H(x,x) = 0$, $H_y(x,x) = 0$ we have from Taylor's formula that $H(x,y) = \frac{1}{2} H_{yy}(x,x)(y-x)^2 + r(x,y)$. Using the fact that $(f(t))_{0\leq t\leq T}$ is a compact subset of $(0,\infty)$ we may assume that the remainder term satisfies $|r(x,y)| \leq R(|y-x|)(y-x)^2$ where R is an increasing function on $[0,\infty)$ such that $R(c) \to 0$ as $c \to 0$. Then

$$\begin{aligned}
\sum_2 H(f(t_k), f(t_{k+1})) &= \frac{1}{2} \sum_2 H_{yy}(f(t_k), f(t_k))(f(t_{k+1}) - f(t_k))^2 + \sum_2 r(f(t_k), f(t_{k+1})) \\
&= \frac{1}{2} \sum H_{yy}(f(t_k), f(t_k))(f(t_{k+1}) - f(t_k))^2 \\
&\quad - \frac{1}{2} \sum_1 H_{yy}(f(t_k), f(t_k))(f(t_{k+1}) - f(t_k))^2 \\
&\quad + \sum_2 r(f(t_k), f(t_{k+1})).
\end{aligned} \tag{33}$$

Since $H_{yy}(f,f) = 2/f^2$ is uniformly continuous over the bounded set of values $(f(t))_{0\leq t\leq T}$, by (9) in Föllmer [16], the first term in (33) converges to $\int_{(0,T]} \frac{1}{f(t-)^2} d[f]_t$ and the second term converges to $-\sum_{s\in C_1} \frac{1}{f(t-)^2} (\Delta f(t))^2$. Using (31) and the fact that the remainder term satisfies $|r(x,y)| \leq R(|y-x|)(y-x)^2$ we have that the last term is bounded by $R(\epsilon)[f]_T$. Finally, letting $\epsilon \downarrow 0$ we conclude that $V_H(f, P_\infty) = \lim_n V_H(f, P^{(n)})$ exists and (30) follows. □

Corollary 7.5. *Suppose f has a quadratic variation along a dense sequence of partitions* $\mathcal{P}$. *Then,* $V_{H^R}(f, P_\infty) = \int_{(0,T]} f(t-)^{-2} d[f]_t$ *and* $V_{H^L}(f, P_\infty) = [\log f]_T$.

Combining (28) with Theorem 7.2 and Lemma 7.4 it follows that for a path f with a finite quadratic variation along a dense sequence of partitions P and ψ a twice-continuously differentiable function with $\psi(f(0)) = 0$,

$$V_H(f, P_\infty) \geq \psi(f(T)) - \int_0^T \psi'(f(t-))df(t). \tag{34}$$

The left hand side is the payoff of the variance swap in the continuous limit. The expression on the right can be interpreted as the payoff of a semi-static hedging strategy $(\psi, -\psi')$ under continuous trading. From Definition 3.7 for each of the partitions in the sequence we have that the price of the semi-static hedge is

$$\int_0^\infty \psi(x)\mu(dx) = \int_{f(0)}^\infty \psi''(x)C_\mu(x)dx + \int_0^{f(0)} \psi''(z)(C_\mu(z) - f(0) + z)dz. \tag{35}$$

Since this value does not depend on the partition, in the continuous-time setting we define the price of sub-hedge $(\psi, -\psi')$ to also be the expression given in (35).

Corollary 7.6. *Suppose H is an increasing variance swap kernel. A model-independent lower bound on the price of the continuous time limit of the variance swap with kernel H is*

$$\sup_\kappa \int_0^\infty \psi_\kappa(x)\mu(dx) = \int_0^\infty \psi_{\alpha_\mu}(x)\mu(dx) \tag{36}$$

where α_μ is the quantity arises in the Perkins embedding (Theorem 4.2).

Proof. For any decreasing function $\kappa \in \mathcal{K}_c$ we can construct ψ_κ such that $\int_0^\infty \psi_\kappa(x)\mu(dx)$ is the price of a sub-hedge for V_H for any partition, and this continues to hold in the continuous-time limit. Moreover, by optimising over κ we obtain a bound $\int_0^\infty \psi_{\alpha_\mu}(x)\mu(dx)$ which is the best bound of this form by Theorem 6.3. Note that even if α_μ is not in class $\mathcal{K}_c$, by Corollary 6.4 we can approximate it from above by a sequence of elements of class $\mathcal{K}_c$ such that in the limit we obtain the price $\int_0^\infty \psi_{\alpha_\mu}(x)\mu(dx)$ as a bound. $\square$

Our goal now is to show that this is a best bound in general and not just an optimal bound based on inequalities such as (28) for $\psi \equiv \psi_\kappa$ and κ a decreasing function. We do this by showing that there is a consistent model for which the price of the continuously monitored variance swap is equal to $\int_0^\infty \psi_{\alpha_\mu}(x)\mu(dx)$.

We now assume that no options with maturities which are shorter than T are traded. The bounds we have derived here and in previous sections remain valid without this assumption, but in showing they are tight we construct a model which is consistent with maturity-T options only. This model need not be consistent with the prices of other vanilla options with shorter maturities, should such options be traded.

Theorem 7.7. *Suppose H is an increasing variance swap kernel. Then there exists a consistent model such that for any dense sequence of partitions $\mathcal{P} = (P^{(n)})_{n\geq 1}$ we have that $\lim_n V_H((X_t)_{0\leq t\leq T}, P^{(n)})$ exists and then*

$$V_H((X_t)_{0\leq t\leq T}, P_\infty) = \psi_{\alpha_\mu}(X_T) - \int_0^T \psi'_{\alpha_\mu}(X_{s-}+)dX_s. \tag{37}$$

Proof. Recall Definition 3.9 and note that we are given a set of call prices and that in constructing a consistent model we are free to design an appropriate probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ as well as a stochastic process $(X_t)_{t\geq 0}$.

Suppose we are given call prices $C(x) = C_\mu(x)$ for some μ. Let $(\Omega, \mathcal{G}, \mathbb{G} = (\mathcal{G}_t)_{0\leq t\leq T}, \mathbb{P})$ support a Brownian motion $(W_u)_{u\geq 0}$ with initial value $W_0 = f(0) = \int_{\mathbb{R}^+} x\mu(dx)$ and suppose $\mathcal{G}_0$ contains a $U[0,1]$ random variable which is independent of W. (This last condition is necessary purely to ensure that the Perkins embedding of μ can be defined when μ has an atom at $f(0)$. If μ has no atom at $f(0)$ then we may take $\mathcal{G}_0$ to be trivial.)

Let τ_μ^P be the Perkins embedding of μ in W. Write S for the maximum process of W so that $S_u = \max_{v\leq u} W_v$. Write $\overline{H}_x$ for the first hitting time by W of x. Let $(\Lambda(t))_{0\leq t\leq T}$ be a strictly increasing continuous function with $\Lambda(0) = f(0)$ and $\lim_{t\uparrow T}\Lambda(t) = \infty$. Now define the left-continuous process $\tilde{X} = (\tilde{X}_t)_{0\le t\le T}$ via

$$\tilde{X}_t = \begin{cases} \Lambda(t) & \overline{H}_{\Lambda(t)} \le \tau^P_\mu \\ W_{\tau^P_\mu} & \tau^P_\mu < \overline{H}_{\Lambda(t)}. \end{cases} \tag{38}$$

Note that the condition $\overline{H}_{\Lambda(t)} \le \tau^P_\mu$ can be re-written as $\Lambda(t) \le S_{\tau^P_\mu}$ or equivalently $t \le \Lambda^{-1}(S_{\tau^P_\mu})$. Define also $\tilde{\mathcal{F}}_t = \mathcal{G}_{\bar{H}_{\Lambda(t)}}$. Then $\tilde{X}$ is adapted to the filtration $\tilde{\mathbb{F}} = (\tilde{\mathcal{F}}_t)_{0\le t\le T}$ and $\tilde{X}$ is a $\tilde{\mathbb{F}}$-martingale for which $\tilde{X}_T = W_{\tau^P_\mu} \sim \mu$.

In order to construct a right-continuous martingale with the same properties, for $t < T$ we set $\mathcal{F}_t = \cap_{u>t}\tilde{\mathcal{F}}_t$ and $X_t = \lim_{u\downarrow t}\tilde{X}_u$, and for $t = T$ we set $\mathcal{F}_T = \tilde{\mathcal{F}}_T$ and $X_T = \tilde{X}_T$. Then X is a right-continuous $\mathbb{F}$-martingale such that $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$ is a consistent model.

Now we want to show that for this model $\lim_n V_H((X_t)_{0\le t\le T}, P^{(n)})$ exists and (37) holds path-wise. Clearly $X_t(\omega)$ has a quadratic variation down any dense sequence of partitions, and hence the limit exists by Lemma 7.4. Moreover, writing ψ for ψ_{α_μ} and X_t as shorthand for $X_t(\omega)$, and noting that until the time of the jump (if any) $X_t \ge f(0)$ so that ψ is continuously differentiable at X_t, we have for each ω

$$\begin{aligned}
\psi(X_T) - \int_0^T \psi'(X_{t-})dX_t &= \psi(W_{\tau^P_\mu}) - \int_{t=0}^{\Lambda^{-1}(S_{\tau^P_\mu})} \psi'(\Lambda(t))d\Lambda(t) - \psi'(S_{\tau^P_\mu})(W_{\tau^P_\mu} - S_{\tau^P_\mu}) \\
&= \psi(W_{\tau^P_\mu}) - \int_{f(0)}^{S_{\tau^P_\mu}} \psi'(u)du - \psi'(S_{\tau^P_\mu})(W_{\tau^P_\mu} - S_{\tau^P_\mu}) \\
&= \psi(W_{\tau^P_\mu}) - \psi(S_{\tau^P_\mu}) - \psi'(S_{\tau^P_\mu})(W_{\tau^P_\mu} - S_{\tau^P_\mu}).
\end{aligned}$$

There are two cases. Either $W_{\tau^P_\mu} = S_{\tau^P_\mu}$, in which case this expression is equal to 0 or, $W_{\tau^P_\mu} = \alpha_\mu(S_{\tau^P_\mu})$ and then the expression becomes

$$\psi(\alpha_\mu(s)) - \psi(s) - \psi'(s)(\alpha_\mu(s) - s) \equiv H(s, \alpha(s))$$

at $s = S_{\tau^P_\mu}$, using Definition 5.2. In either case the right hand side of (37) is $H(S_{\tau^P_\mu}, W_{\tau^P_\mu})$. For the left hand side of (37), $[X]^c_T = 0$ and $(\Delta X_u)^2 = (S_{\tau^P_\mu} - W_{\tau^P_\mu})^2 1_{\{u=\Lambda^{-1}(S_{\tau^P_\mu})\}} 1_{\{W_{\tau^P_\mu} \ne S_{\tau^P_\mu}\}}$ so that from (30), $V_H(f, P_\infty) = H(S_{\tau^P_\mu}, W_{\tau^P_\mu})$. Hence (37) holds path-wise.

□

Corollary 7.8. *Suppose H is an increasing variance swap kernel. Then the highest model independent lower bound on the price of a variance swap which is valid across all partitions is given by the expression in (36).*

Corollary 7.9. *If $\Phi(u, y)$ does not depend on y then the corresponding variance swap is perfectly replicable by $(\psi, -\psi'_+)$. For all consistent models the variation swap has price $\int_{\mathbb{R}^+} \psi(x)\mu(dx)$.*

Example 7.10. *Recall the definitions of the kernels H^B and H^Q and Example 5.8. $\Phi^B(u, y) = 2u^{-2}$ and so $\psi'(u) = -2/u$ and $\psi(u) = -2\log(u)$. Thus $H^B(x, y) = \psi(y) - \psi(x) - \psi'(x)(y - x)$ and the strategy $(\psi, -\psi')$ replicates the payoff perfectly for any price realisation. The observation that H^B has one model-independent price was first made by Bondarenko in [3]. Similarly, $H^Q(x, y) = \psi(y) - \psi(x) - \psi'(x)(y - x)$, where $\psi(x) = x^2$. An alternative analysis of these two payoffs is given by Neuberger [25]. Both Bondarenko [3] and Neuberger [25] advocate the use of H^B due to the fact that its price is not sensitive to the price path, but only to the value of X_T.*

Remark 7.11. *As discussed at the start of this section, for the trivial partition the price of the variance swap is model-independent. We know that as the mesh size of the partition converges to zero then the model-independent bound converges to that in the continuous-time limit, as given, for example, in Corollary 7.6. We now want to investigate the rate of convergence. Our focus is on the kernel H^R.*

Suppose call prices are such that $\mu \in L^2$. Note that we must have $\mu \in L^2$ else $\mathbb{E}[V_{H^R}(X, P)] = \infty$ for all finite partitions and all consistent models.

Let $\tilde{X}$ and Λ be as given in (38) and let $m = f(0) = \tilde{X}_0$. Fix $N \in \mathbb{N}$ and $\lambda > 1$ and let $\gamma_n = m\lambda^n$ $(n = 0, 1, \ldots, N-1)$ and $\gamma_N = \infty$. Let $T_n = \Lambda^{-1}(\gamma_n)$. Then $T_0 = 0$ and $T_N = T$. Finally set $P^{(N)} = \{T_0, T_1, \ldots, T_N\}$. Then, with

$\tilde{S}_T = \sup_{u\le T}\tilde{X}_u$, *and* $D^{(N)} = V_{H^R}(\tilde{X},P^{(N)}) - V_{H^R}(\tilde{X},P_\infty)$,

$$\begin{aligned} D^{(N)} &= \sum_{n=1}^{N}\frac{(\gamma_n-\gamma_{n-1})^2}{\gamma_{n-1}^2}I_{\{\tilde{S}_T\ge\gamma_n\}} + \sum_{n=1}^{N}\frac{(\tilde{X}_T-\tilde{X}_{T_{n-1}})^2}{\tilde{X}_{T_{n-1}}^2}I_{\{\gamma_{n-1}\le\tilde{S}_T<\gamma_n\}} - \sum_{n=1}^{N}\frac{(\tilde{X}_T-\tilde{S}_T)^2}{\tilde{S}_T^2}I_{\{\gamma_{n-1}\le\tilde{S}_T<\gamma_n\}} \\ &= \sum_{n=1}^{N-1}\frac{(\gamma_n-\gamma_{n-1})^2}{\gamma_{n-1}^2}I_{\{\tilde{S}_T\ge\gamma_n\}} + \sum_{n=1}^{N-1}\frac{(\tilde{X}_T-\tilde{X}_{T_{n-1}})^2}{\tilde{X}_{T_{n-1}}^2}I_{\{\gamma_{n-1}\le\tilde{S}_T<\gamma_n,\tilde{X}_T\ge m\}} \\ &\quad + \frac{(\tilde{X}_T-\gamma_{N-1})^2}{\gamma_{N-1}^2}I_{\{\gamma_{N-1}\le\tilde{S}_T,\tilde{X}_T\ge m\}} + \sum_{n=1}^{N}\left(\frac{(\tilde{X}_T-\tilde{X}_{T_{n-1}})^2}{\tilde{X}_{T_{n-1}}^2} - \frac{(\tilde{X}_T-\tilde{S}_T)^2}{\tilde{S}_T^2}\right)I_{\{\gamma_{n-1}\le\tilde{S}_T<\gamma_n,\tilde{X}_T<m\}} \end{aligned}$$

The first three terms are positive and the last is negative. The positive terms can be bounded by

$$(\lambda-1)^2\sum_{n=1}^{N-1}I_{\{\log(\tilde{S}_T/m)/\log\lambda\ge n\}} \le (\lambda-1)^2\frac{\log(\tilde{S}_T/m)}{\log\lambda},$$

$(\lambda-1)^2$ *and* $\tilde{X}_T^2\gamma_{N-1}^{-2} = \tilde{X}_T^2 m^{-2}\lambda^{-2(N-1)}$ *respectively.*

We have $\mathbb{E}[V_{H^R}(\tilde{X},P^{(N)})] \ge \int\psi_{\alpha(\mu)}(x)\mu(dx) = \mathbb{E}[V_{H^R}(\tilde{X},P_\infty)]$. *Conversely, under the assumption that* $\mu\in L^2$ *and hence* $\mathbb{E}[\tilde{X}_T^2] = am^2 < \infty$, $\mathbb{E}[\tilde{S}_T^2] \le 4am^2 < \infty$ *and* $\mathbb{E}[\log(\tilde{S}/m)] = b < \infty$, *we have*

$$\mathbb{E}[V_{H^R}(\tilde{X},P^{(N)})] - \mathbb{E}[V_{H^R}(\tilde{X},P_\infty)] \le \left(1+\frac{b}{\log\lambda}\right)(\lambda-1)^2 + a\lambda^{-2(N-1)}.$$

Now choose $\lambda = 1+(2N)^{-1}\log N$. *Then,* $\mathbb{E}[V_{H^R}(\tilde{X},P^{(N)})] - \mathbb{E}[V_{H^R}(\tilde{X},P_\infty)]$ *is bounded above by a term of order* $\log N/N$. *Further, for any* $N\ge 3$

$$\mathbb{E}[V_{H^R}(\tilde{X},P_\infty)] \le \inf\mathbb{E}[V_{H^R}(X,P^{(N)})] \le \mathbb{E}[V_{H^R}(\tilde{X},P^{(N)})] \le \mathbb{E}[V_{H^R}(\tilde{X},P_\infty)] + \left(2a+b+\frac{1}{2}\right)\frac{\log N}{N},$$

where the infimum is taken over all consistent models.

8 Non-zero interest rates

To date we have worked with forward prices. This has the implication that the dynamic part of a hedging strategy has zero cost. In this section we outline how our analysis can be extended to non-zero, but deterministic, interest rates.

Suppose that interest rates are deterministic. Let $D_t = D_t(T)$ be the discount factor over $[t,T]$ so that the asset price realisation ($s = (s_t)_{0\le t\le T}$) and the forward price realisation are related by $s(t) = D_t f(t)$. In the case of constant interest rates $D_t(T) = e^{-r(T-t)}$ so that $s(t) = e^{-r(T-t)}f(t)$.

Let P be a partition of $[0,T]$. For $k\in\{0,1,...,N-1\}$ write $s_k = s(t_k)$, $f_k = f(t_k)$ and $D_k = D_{t_k}(T)$. Set $D_{k,k+1} = D_{k+1}/D_k$. Note that if interest rates are non-negative then $D_{k,k+1}\ge 1$.

Let G be the kernel of a variation swap and write $G_k(x,y) = G(D_k x, D_k y)$. Then the payoff of the variance swap is given by

$$V_G(s,P) = \sum_{k=0}^{N-1}G(D_k f_k, D_{k+1}f_{k+1}) = \sum_{k=0}^{N-1}G_k(f_k, D_{k,k+1}f_{k+1}).$$

Proposition 8.1. *Suppose that there exists a variation swap kernel* H, *functions* η, ϵ, B *and a constant* $A\in\mathbb{R}$ *such that for all* $D>0$

$$G_k(x,yD) \ge AH(x,y) + \eta(y) - \eta(x) + \epsilon(x,k,D)(y-x) + B(k,D). \tag{39}$$

Without loss of generality we may take $\eta(f(0)) = 0$.

Suppose that there exists a semi-static sub-hedging strategy (ψ,Δ) *for the variation swap with kernel* H. *Then*

$$V_G(s,P) \ge (A\psi+\eta)(f(T)) + \sum_k[\epsilon(f_k,k,D_{k,k+1}) + \delta_{t_k}((f(t)_{t\le t_k})](f_{k+1}-f_k) + \sum_k B(k,D_{k,k+1}),$$

and there is a model-independent sub-hedge and price lower bound for V_G.

Proof. We have

$$\begin{aligned}
V_G(s,P) &= \sum_{k=0}^{N-1} G_k(f_k, D_{k,k+1}f_{k+1}) \\
&\geq \sum_k [AH(f_k, f_{k+1}) + \eta(f_{k+1}) - \eta(f_k) + \epsilon(f_k, k, D_{k,k+1})(f_{k+1} - f_k) + B(k, D_{k,k+1})] \\
&\geq A[\psi(f(T)) + \sum_k \delta_{t_k}((f(t)_{t\leq t_k})(f_{k+1} - f_k)] + \eta(f(T)) \\
&\qquad + \sum_k \epsilon(f_k, k, D_{k,k+1})(f_{k+1} - f_k) + \sum_k B(k, D_{k,k+1}).
\end{aligned}$$

□

Remark 8.2. *If we are content to assume that interest rates are non-negative then we only need (39) to hold for $D \geq 1$.*

Remark 8.3. *The price for the floating leg associated with the hedge is the price of the static vanilla portfolio with payoff $(A\psi + \eta)(f(T))$ plus the constant $\sum_{k=0}^{N-1} B(k, D_{k,k+1})$.*

Corollary 8.4. *Suppose H is an increasing variance kernel, and ψ is of Class $\mathcal{K}$. If (39) holds then we have a path-wise sub-hedge and a model independent bound on the price of V_G.*

In the setting of increasing or decreasing variance kernels the bound in (40) will be tight provided $(\psi, -\psi')$ is a tight semi-static hedge for $V_H(f,P)$ and there is equality in Equation (39).

Example 8.5. *Suppose $G(x,y) = H^R(x,y) = \frac{(y-x)^2}{x^2}$. Then $G_k(x,y) = G(x,y)$, so that $\epsilon(x,k,D)$ and $B(k,D)$ will not depend on k. Moreover,*

$$\begin{aligned}
G(x,yD) &= \frac{1}{x^2}(Dy - Dx + Dx - x)^2 \\
&= D^2\left(\frac{y-x}{x}\right)^2 + D\frac{(D-1)}{x}(y-x) + (D-1)^2
\end{aligned}$$

Suppose that interest rates are non-negative so that $D_{k,k+1} \geq 1$. Then (39) holds for $A = 1$, $\eta = 0$, $\epsilon(x,D) = D(D-1)/x$ and $B(D) = (D-1)^2$.

Note that there is an inequality in (39) for $A = 1$. If $D_{k,k+1}$ is independent of k (the natural example is to assume that interest rates are constant and the partition is uniform, in which case $d = \log D_{k,k+1} = rT/N$) then we can have equality by taking $A = e^{2rT/N}$. In that case we have an improved bound, but the improvement becomes negligible in the limit $N \uparrow \infty$.

Example 8.6. *Suppose $G(x,y) = H^L(x,y) = (\log(y) - \log(x))^2$. Then $G_k(x,y) = G(x,y)$ and $G(x,yD) = (\log D + \log y - \log x)^2 = H^L(x,y) + 2\log D(\log y - \log x) + (\log D)^2$.*

Suppose now that the partition is such that $D_{k,k+1}$ is independent of k, and set $d = \log D_{k,k+1}$. Then Equation (39) holds with equality for $A = 1$, $\eta(y) = 2d\log y$, $\epsilon = 0$ and $B(D) = d^2$.

Example 8.7. *Suppose $G(x,y) = H^B(x,y) = -2(\log y - \log x) - (y/x - 1))$. Then $G_k(x,y) = G(x,y)$ and*

$$\begin{aligned}
G(x,yD) &= -2(\log y - \log x + \log D) + 2D(y-x) + 2(D-1) \\
&= H^B(x,y) + 2(D-1)(y/x - 1) + H^B(1,D).
\end{aligned}$$

Then Equation (39) holds with equality for $A = 1$, $\eta(y) = 0$, $\epsilon(x,D) = 2(D-1)/x$, $B(D) = H^B(1,D)$.

We can consider the limit as the partition becomes dense, in which case the bounds for the variance swap become tight. For definiteness we will assume that we have a sequence of uniform partitions with mesh size tending to zero, and that interest rates are constant, though this can be weakened for the squared return and Bondarenko kernels.

Then, for each of the three examples above we have that $\sum_{k=0}^{N-1} B(k, D_{k,k+1}) = NB(e^{rT/N}) \to 0$. Further, in each case $\eta(y) \to 0$, and $A = 1$. Then in the limit the lower bound on the price of the variance swap based on the price realisation s is the same as the upper and lower bounds for the variance swap defined relative to the forward price f. Thus, for variance swaps based on frequent monitoring, the bounds we have calculated in earlier sections based on the forward price may also be used for undiscounted price processes.

8.1 Super-hedges and upper bounds

Corollary 8.8. *Suppose there exists H, η, ϵ, B, and A such that*

$$G_k(x, yD) \le AH(x, y) + \eta(y) - \eta(x) + \epsilon(x, k, D)(y - x) + B(k, D), \tag{40}$$

and suppose that there exists a semi-static super-hedging strategy (ψ, Δ) for the variation swap with kernel H. Then there is a corresponding model-independent super-hedge and price upper bound for V_G.

The analysis of the kernels H^R,H^L,H^B and upper bounds is similar to that in Examples 8.5—8.7 above. For the kernel H^B, the choices listed in Example 8.7 give equality in (40) and can be used equally for upper bounds. Provided that we have an upper bound for $D_{k,k+1}$, so that $D_{k,k+1} \le \bar{D}$ uniformly in k, for the kernel H^R we may take $A = \bar{D}^2$, $\eta = 0$, $\epsilon(x, D) = D(D-1)/x$ and $B(D) = (D-1)^2$. Finally, for H^L, provided interest rates are non-negative, we can write

$$G(x, yD) = H^L(x, y) + 2\log D(\log y - \log x) + (\log D)^2 \le H^L(x, y) + 2\frac{\log D}{x}(y - x) + (\log D)^2$$

so that (40) holds for $A = 1$, $\eta = 0$, $\epsilon(x, D) = 2(\log D)/x$ and $B(D) = (\log D)^2$. Note that, unlike for the lower bound in Example 8.6, for the upper bound we do not need to assume that $D_{k,k+1}$ is independent of k.

Remark 8.9. *In his analysis of lower bounds for the kernel H^L, Kahalé [22] does not need to assume the partition is uniform and that interest rates are constant (or more generally that $D_{k,k+1}$ is constant), and can allow for arbitrary finite partitions and deterministic interest rates. Our results complement his results nicely. Although we need the assumption that $D_{k,k+1}$ is constant to recover Kahalé's result in the setting of lower bounds and the kernel H^L, in all other cases of study (upper bounds for V_{H^L} and upper and lower bounds for V_{H^R} and V_{H^B}) our methods also allow for arbitrary partitions and non-constant but deterministic interest rates.*

9 Numerical Results

Given a continuum of call prices, it is possible to calculate the model independent bounds for the prices of variance swaps. When the implied terminal distribution of the asset price is simple, for example a uniform distribution, it is sometimes possible to calculate the monotone functions associated with the Perkins embedding explicitly (see Example 5.4) and to obtain a closed form integral expression for the model independent upper and lower bounds. For more realistic and complex target laws, the monotone functions and bounds can still be calculated numerically. The case when the terminal law is lognormally distributed is of particular practical interest.

A standard time frame for a volatility swap is 30 days or one month ($T = 1/12$), which is the time frame used for the widely quoted 'VIX index'. Figure 3 plots the upper and lower bounds for the prices of variance swaps based on the kernels H_R and H_L relative to the cost of $-2\log$ contracts (the Neuberger/Dupire price of the standard hedge or 'VIX price'). In the first picture prices are plotted against the volatility parameter of the lognormal (terminal) distribution centred at 1. More precisely, the bounds are plots of

$$\sigma \to \mathbb{E}[\psi_{\kappa,H}(X_{\sigma/\sqrt{12}})]/\mathbb{E}[-2\log X_{\sigma/\sqrt{12}}], \qquad \text{and} \qquad \sigma \to \mathbb{E}[\psi_{\ell,H}(X_{\sigma/\sqrt{12}})]/\mathbb{E}[-2\log X_{\sigma/\sqrt{12}}],$$

where $X_\sigma \equiv e^{\sigma N - \sigma^2/2}$ is the lognormal random variable with volatility parameter σ and $H = H^R$ or H^L. Here, $\psi_{K,H}$ is the function given in Definition 5.2 and κ is chosen according to Theorem 6.3 (with ℓ chosen similarly). In the second picture bounds are plotted against a parameter representing the radius of the uniform distribution.

In both cases the upper bound for the kernel H_L and the lower bound for the kernel H_R correspond to the decreasing function κ associated with the Perkins embedding, while the other two bounds are constructed with the increasing function ℓ associated with the reversed Perkins embedding.

Note that the price of a variance swap in the Black-Scholes model (as given by $\mathbb{E}[-2\log X_{\sigma\sqrt{T}}]$) is an increasing function of volatility. The upper and lower bounds are also increasing functions of volatility, and, as can be seen in the figure, they also become wider as volatility increases, when expressed as a ratio against the no-jump case. For reasonable values of volatility, and for both kernels, the impact of jumps is to affect the price by a factor of less than two, and for the kernel H^L the bounds are even tighter. The observation that the bounds for the kernel H_R are wider than those for the kernel H_L is partly explained by considering the leading term in the expansion of the hedging error (see Section 4.2). We have $J_R(x) \approx 2x^3/3$ whereas $J_L(x) \approx -x^3/3$ so that the magnitude of the leading error term for H_R is twice that of the leading error term for H_L. Note that for the optimal martingales the jumps are not local, so this approximation becomes less relevant as σ increases.

For the uniform case the relative bounds again become wider as the parameter of the uniform distribution increases. Again the bounds for the kernel H^R are wider than those for H^L.

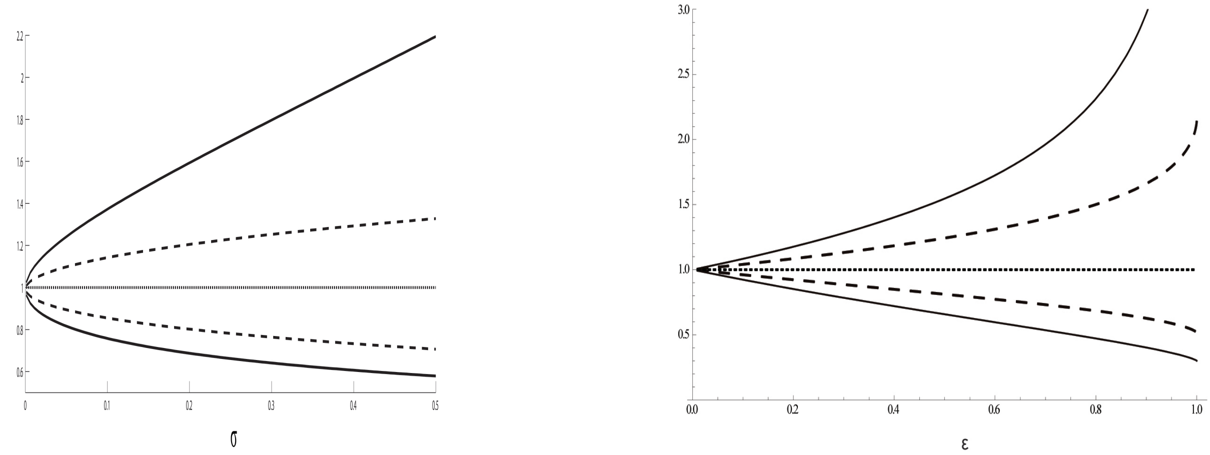

Figure 3: Model independent upper and lower bounds for the prices of variance swaps based on the kernels H_R (solid lines) and on H_L (dashed lines) relative to the price of $-2\log$ contracts (dotted line). There are two cases: on the left where the terminal distribution is lognormal with volatility between 0 and 0.5, (and $T = 1/12$), and on the right where the terminal distribution is uniform on $[1-\epsilon, 1+\epsilon]$, as ϵ ranges between 0 and 1. Here we work with variance swaps on forward prices.

10 Summary and concluding remarks

This article developed from an attempt to express the results of Kahalé [22] on no-arbitrage lower bounds for the prices of variance swaps in the framework of model-independent hedging, in which extremal models and prices are associated with extremal solutions of the Skorokhod embedding problem. Beginning with Hobson [17], the focus in this literature is on hedging, and on finding path-wise inequalities relating the payoff of the exotic, path-dependent derivative and the payoff of a static vanilla call portfolio combined with the gains from trade from an investment in the underlying security. In the context of variance swaps we find that the lower bound is associated with a martingale price process which can be expressed as a time-change of the Perkins solution of the Skorokhod embedding problem. This embedding has appeared previously in finance in the construction of model-independent bounds for the prices of barrier options (Brown et al [6]).

We approach the problem of finding hedging strategies in a more general setting than Kahalé [22] in that we consider a variety of kernels in the definition of the variance swap, whereas Kahalé [22] focuses on the kernel H^L. The ability to consider general kernels allows us to emphasise the dependence of the payoff on the presence and character of the jumps, and to show that the nature of this dependence is strongly influenced by the form of the kernel. Bondarenko [3] and Neuberger [25] argue that the finance industry should consider defining variance swaps using the kernel H^B as then they can be replicated perfectly, even in the presence of jumps, recall Example 7.10. The counterargument is that variance swaps provide value precisely because they are not redundant in this way. Sophisticated investors want to be able to take positions on the likely presence and direction of jumps. This is possible if the variance swap is defined using the kernel H^R or H^L, but not using H^B.

Kahalé [22] works directly with the undiscounted asset price, and does not give special attention to contracts written on the forward price. He introduces the class of V-convex functions which have the property that each such function gives a lower bound on the price of the variance swap, and an associated sub-hedge. He then proceeds to show that functions ψ of Class $\mathcal{L}$ (in our notation) are V-convex. In this way he can deduce a lower bound on the price of a variance swap. Further, for a particular choice of decreasing function he can show that this lower bound can be attained in the continuous time limit under a well-chosen stochastic model — hence the bound he attains must be a best bound.

In contrast, this paper splits the problem into two parts. First we consider the interest-free case. This simplifies the problem and reveals the link to the Perkins solution for the Skorokhod embedding problem. Theorem 6.3 relates the hedging of variance swaps in an interest-free environment to the optimality properties of the Perkins embedding proved in [19]. The differences between the two-step approach presented here and the approach in Kahalé [22] are summarised by the differences between condition (24) in this paper and the corresponding condition for V-convexity in Kahalé [22, Equation (3.1)]. The equation in Kahalé must be disentangled from the interest-rate terms in order to lead to a result about the extremality of hedges constructed from the Perkins embedding. On the other hand, there is a direct line of argument that links Equation (24) in this paper to Theorem 6.3.

There is no disadvantage to the two-step approach as there are simple inequalities which extend the zero-interest rate bounds to the general case. In the limit of a dense sequence of partitions the same bounds are optimal in both the undiscounted and forward price settings. This is further support for the point that, mathematically at least, the fundamental quantities are those derived in a zero-interest rate setting. Finally using inequalities such as (37) allows us to quantify the price difference between contracts written on the undiscounted and forward prices for discrete monitoring, in itself an interesting result.

A further contribution of this article is to provide a derivation of bounds on the prices of variance swaps without any recourse to probability. This involves construction of a class of hedges parameterised by monotone functions, and the choice of an optimal element in this class for a given set of call prices, together with Föllmer's non-probabilistic Itô calculus. Price trajectories for which the bound is path-wise tight have at most one jump, after which the trajectory is constant. Probability is only required to show that these trajectories correspond to a stochastic model for the price process. The relationship between the optimality of the cheapest hedge, derived in a purely non-probabilistic fashion, and the optimality of the Perkins embedding provides a pleasing completeness to the story.

References

[1] J. Azéma and M. Yor. Une solution simple au problème de Skorokhod. In *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*, volume 721 of *Lecture Notes in Math.*, pages 90–115. Springer, Berlin, 1979.

[2] A. Bick and W. Willinger. Dynamic spanning without probabilities. *Stochastic Processes and their Aplications*, 50:349–374, 1994.

[3] O. Bondarenko. Variance trading and market price of variance risk. Working paper, 2007.

[4] D.T. Breeden and R.H. Litzenberger. Prices of state-contingent claims implicit in option prices. *J. Business*, 51:621–651, 1978.

[5] M. Broadie and A. Jain. The effect of jumps and discrete sampling on volatility and variance swaps. *Int. J. of Th. and App. Finance*, 11(8):761–791, 2008.

[6] H. Brown, D.G. Hobson, and L.C.G Rogers. Robust hedging of barrier options. *Math. Finance*, 11(3):285–314, 2001.

[7] P. Carr and A. Corso. Covariance contracting for commodities. *Energy and Power Risk Management*, April:42–45, 2001.

[8] P. Carr and R. Lee. Variation and share-weighted variation swaps on time-changed Lévy processes. Preprint, 2010.

[9] P. Carr, R. Lee, and L. Wu. Variance swaps on time-changed Lévy processes. Preprint, 2010.

[10] A. Cox and J. Wang. Root's barrier: construction, optimality and applications to variance options. Preprint, 2011.

[11] M.H.A. Davis and D.G. Hobson. The range of traded option prices. *Mathematical Finance*, 17(1):1–14, 2007.

[12] K. Demeterfi, E. Derman, M. Kamal, and J. Zou. A guide to volatility and variance swaps. *The Journal of Derivatives*, 6(4):9–32, 1999.

[13] K. Demeterfi, E. Derman, M. Kamal, and J. Zou. More than you ever wanted to know about volatility swaps. *Goldman-Sachs Quantitive Strategies Research Notes*, 1999.

[14] L.E. Dubins and D. Gilat. On the distribution of maxima of martingales. *Proceedings of the American Mathematical Society*, 68:337–338, 1978.

[15] B. Dupire. Arbitrage pricing with stochastic volatility. *Société Générale, Options Division, Paris*, 1992.

[16] H. Föllmer. Calcul d'Itô sans probabilites. In *Séminaire de Probabilités, XV (Univ. Strasbourg, Strasbourg, 1981)*, volume 850 of *Lecture Notes in Math.*, pages 143–150. Springer, Berlin, 1981.

[17] D.G Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2:329–347, 1998.

[18] D.G. Hobson. The Skorokhod embedding problem and model independent bounds for option prices. In *Paris-Princeton Lecture Notes on Mathematical Finance*. Springer, 2010.

[19] D.G. Hobson and M. Klimmek. Maximising functionals of the joint law of the maximum and terminal value in the Skorokohd embedding problem. Preprint, 2010.

[20] D.G. Hobson and J. L. Pedersen. The minimum maximum of a continuous martingale with given initial and terminal laws. *Ann. Probab.*, 30(2):978–999, 2002.

[21] R. Jarrow, Y. Kchia, M. Larsson, and P. Protter. Discretely sampled variance and volatility swaps versus their continuous approximations. Preprint, 2011.

[22] N. Kahalé. Model-independent lower bound on variance swaps. Preprint, 2011.

[23] I. Martin. Simple variance swaps. Preprint, 2011.

[24] A. Neuberger. The Log Contract. *Journal of Portfolio Management*, 20(2):74–80, 1994.

[25] A. Neuberger. Realized skewness. Working paper, 2010.

[26] E. Perkins. The Cereteli-Davis solution to the H^1-embedding problem and an optimal embedding in Brownian motion. In *Seminar on Stochastic Processes, 1985 (Gainesville, Fla., 1985)*, pages 172–223. Birkhäuser Boston, Boston, MA, 1986.

[27] A. V. Skorokhod. *Studies in the theory of random processes*. Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.

[28] S. Zhu and G-H. Lian. A closed-form exact solution for pricing variance swaps with stochastic volatility. *Mathematical Finance*, 21(2):233–256, 2011.