City Research Online

City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Robust Bounds for Forward Start Options

David Hobson[†]   Anthony Neuberger[‡]

First version: October 9, 2008,
This version: December 16, 2009

Abstract

We consider the problem of finding a model-free upper bound on the price of a forward-start straddle with payoff $|F_{T_2} - F_{T_1}|$. The bound depends on the prices of vanilla call and put options with maturities T_1 and T_2, but does not rely on any modelling assumptions concerning the dynamics of the underlying. The bound can be enforced by a super-replicating strategy involving puts, calls and a forward transaction.

We find an upper bound, and a model which is consistent with T_1 and T_2 vanilla option prices for which the model-based price of the straddle is equal to the upper bound. This proves that the bound is best possible. For lognormal marginals we show that the upper bound is at most 30% higher than the Black-Scholes price.

The problem can be recast as finding the solution to a Skorokhod embedding problem with non-trivial initial law so as to maximise $\mathbb{E}|B_\tau - B_0|$.

1   Introduction

In this article we consider the problem of pricing forward start options. More especially, if F_t is the forward price of a traded security and if T_1 and T_2 are maturities with $T_0 < T_1 < T_2$, where T_0 is the current time, then we wish to price a security paying $|F_{T_2} - F_{T_1}|$, ie a straddle with the strike set to be the prevailing value at T_1.

Our philosophy is that rather than pricing under a given (and inevitably mis-specified) model, we assume we know the call prices for maturities T_1 and T_2, and we use those prices to reduce the set of feasible price processes to those which are consistent with these calls under a martingale measure, and then we search over the feasible price processes to give the forward start straddle with the highest price. The pricing problem can also be expressed in a different way as a dual problem where we identify the highest model price with the cheapest super-replicating hedge. The resulting price is robust in the sense that it gives a model-free no-arbitrage bound. This bound can be enforced by using calls (with maturities T_1 and T_2), and the forward as hedging instruments. Similar ideas have been applied

---

[†]Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK
[‡]Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK

1

to other path-dependent options, including barrier options and the lookback option, by Hobson [11], Brown, Hobson and Rogers [4] and most recently Cox and Obłój [6].

Part of the interest in the forward start straddle is that the model which attains the maximum is such that, conditional on the price at T_1, the price at T_2 takes one of two values (at least in the atom-free case with nice densities). As this conditional distribution places mass at two-points it can be thought of as a distribution with minimal kurtosis. In this weak sense at least, a long position in a forward start option (suitably hedged using conventional options) is akin to a short position in the kurtosis of the underlying asset.

The main result, expressed in financial language, is the following.

Theorem 1 *Suppose that call prices are given for a pair of maturities $T_1 < T_2$ (for a continuum of strikes on each date) and that these prices are consistent with no-arbitrage[1]. Consider the price of a forward start straddle[2] on the forward price of the asset. Then there exists a model-independent[3] upper bound on the price of this derivative; this bound can be enforced through the purchase of a portfolio of call options and a single forward transaction. Moreover, there is a model which is consistent with the observed vanilla prices for which the (appropriately discounted) payoff of the forward start straddle is equal to the bound; hence the bound is a least model-free upper bound.*

The model-free upper bound on the price of the forward start option with payoff $|F_{T_2} - F_{T_1}|$ is increasing in the final maturity T_2. However, the bound on the price of a forward start option is not necessarily decreasing in the starting maturity T_1, and there are examples where the price of a forward start straddle with payoff $|F_{T_2} - F_{T_1}|$ exceeds that of a vanilla at-the-money straddle with payoff $|F_{T_2} - F_{T_0}|$, where T_0 is the current time.

The lack of monotonicity in the starting maturity of the price of a forward start straddle is one of the surprising results of this study.

As noted by Breeden and Litzenberger [3], knowledge of European call option prices (for the continuum of strikes) is equivalent to knowledge of the marginal distributions of the price process under a risk-neutral measure. Hence we will assume that we know the laws of $X \sim F_{T_1}$ and $Y \sim F_{T_2}$ and that they are given

---

[1]We say that a set of traded securities is consistent with no arbitrage if there is no portfolio of traded instruments (which in our case are the puts and calls with maturities T_1 and T_2, and the riskless bond) which can be combined with a simple semi-static hedging strategy in the forward (buy and hold over $(0, T_1]$ or $(0, T_2]$ or $(T_1, T_2]$) such that the initial cost is zero but the final payoff is non-negative almost surely, and positive with positive probability. The fact that a butterfly spread has a non-negative payoff means that it must have a non-negative price, else there is an arbitrage. Hence option prices for a fixed maturity must be convex in the strike. Similarly, option prices must be increasing in maturity. For further discussion, see, for example, Davis and Hobson [7].

[2]We generally work with the straddle, but from the identity $|y - x| = 2(y - x)^+ - (y - x)$ it is clear that the results can be reformulated in terms of puts or calls.

[3]If the call prices with maturities T_1 and T_2 are consistent with no-arbitrage then we can ask for what range of prices can we introduce the forward start straddle, such that the augmented set of traded securities is consistent with no-arbitrage. The limits of this range give model independent upper and lower bounds on the prices of the forward start straddle. This concept of a model-independent bound is implicit in Hobson [11] and explicit in Föllmer and Schied [8], Cox and Obłój [6] and Hobson [10].

by μ and ν respectively. By the martingale property we have $\mathbb{E}[Y|X] = X$ so that μ and ν have the same mean. Typically, we will use a shift of coordinate system and assume the mean to be zero. However, in the sections on the financial context where the marginals are derived from positive prices and the associated measures lie on $\mathbb{R}^+$, this is not appropriate and we will assume that the measures have equal but positive means.

Define $H(\mu, \nu) := \sup \mathbb{E}|Y - X|$, where the supremum is taken over pairs of random variables (X, Y) with the appropriate marginals, and satisfying the martingale condition $\mathbb{E}[Y|X] = X$. The problem of calculating H can be recast as a Skorokhod embedding problem for Brownian motion B (Skorokhod [17], see also Obłój [14] for a thorough survey). The Skorokhod embedding problem (SEP) for Brownian motion null at zero is, given a centred probability measure ν, to find a uniformly integrable[4] stopping time τ such that $B_\tau \sim \nu$. Our problem is a variant on this in the sense that instead of $B_0 \equiv 0$ we have $B_0 \sim \mu$. The problem becomes to find the solution of a SEP with given initial and terminal laws with the additional optimality property that $\mathbb{E}[|B_\tau - B_0|]$ is maximised. Since, in general, there is no unique solution to the SEP, adding an optimality criterion has proved to be a useful way of characterising solutions with particular properties (eg Azéma and Yor [1], Perkins [15], Jacka [13] and Vallois [18], and, for the problem with non-trivial initial law, Hobson and Pedersen [12]). The connection between the forward start option and the SEP is made precise by identifying X with F_{T_1} and B_0, Y with F_{T_2} and B_τ, and noting that the martingale property of the forward price means that it is a time-change of Brownian motion.

The first and most immediate question is to determine when the problem is feasible, in the sense that given centred probability measures μ and ν, when does there exist a martingale with initial distribution μ and terminal distribution ν. By an application of Jensen's inequality it can be seen that a necessary condition for such a martingale to exist is that $\mu \preceq \nu$ in the sense of convex order — by construction of solutions of the SEP this can also be seen to be sufficient.

We want to study H in the feasible case $\mu \preceq \nu$.

Proposition 2 *Suppose μ, ν, χ are centred probability measures, and that $\mu \prec \nu \prec \chi$ in the sense of convex order. Then $H(\mu, \nu) \leq H(\mu, \chi)$. However, it is not necessarily the case that $H(\mu, \chi) \geq H(\nu, \chi)$.*

This counter-intuitive result (see Lemma 4 and Example 5 below) is indicative of some of the subtleties of the problem. Nonetheless it turns out that optimal solutions always exist, and they always have a particular simple form whereby conditional on X, Y takes one of two values (in the non-atomic case at least). In the SEP setting, τ is the first exit time of B from an interval which depends on B_0 alone.

A special case of the main theorem, Theorem 19, is the following:

Theorem 3 *Suppose μ and ν are centred probability measures with bounded support, and suppose $\mu \preceq \nu$ and that μ has no atoms. Then, there exist increasing*

---

[4]A stopping time τ for Brownian motion is said to be uniformly integrable if the family $(B_{t \wedge \tau})_{t \geq 0}$ is uniformly integrable.

functions f and g with $f(x) \leq x \leq g(x)$, such that if $X \sim \mu$ and if conditional on $X = x$, $Y \in \{f(x), g(x)\}$ respects the martingale properties[5], then $Y \sim \nu$. Moreover, f, g can be chosen such that the joint law maximises $\mathbb{E}[|Y - X|]$ amongst pairs of random variables satisfying $X \sim \mu$, $Y \sim \nu$ and $\mathbb{E}[Y|X] = X$, and then

$$H(\mu, \nu) = \int \mu(dx) \frac{(g(x) - x)(x - f(x))}{(g(x) - f(x))}.$$

One unfortunate feature of the solution is that it is non-constructive, in the sense that given general measures μ and ν we are not able to give explicit formulae for f and g. (However, there are some simple examples where exact formulae can be given, and it is always possible to reverse engineer solutions by fixing μ, f and g, subject to some consistency conditions, and deducing the appropriate law for ν.) This is reminiscent of the situation for the barrier solution of the SEP due to Root [16].

The idea behind the proof is to write down a Lagrangian formulation of the problem, and to derive relationships between the multipliers, which ultimately give the characteristics of the optimal solution. The optimal multipliers are related to a particular convex function, but it is possible to derive a bound from any convex function. Hence, even in cases where it is difficult to determine H precisely, it is straightforward to give families of simple and concrete bounds. Inequalities derived in this fashion, see especially Example 6.2, may be of independent interest.

The remainder of the paper is structured as follows. In the next section we describe the set-up and introduce notation. In Section 3 we consider the non-monotonicity of H in the initial law. In Sections 4, 5 and we 6 we introduce the Lagrangian approach and give examples. These sections provide intuition and motivation for the later analysis. In Sections 7 and 8 we describe and prove the main result, which follows by taking limits over discrete approximations to the initial and terminal distributions. Sections 9 and 10 give further examples and financial interpretation.

2 Notation

In this article we will use three different probabilistic set-ups. The financial set-up involves the forward price process $(F_t)_{0 \leq t \leq T_2}$. The Brownian set-up describes a Brownian motion $(B_t)_{t \geq 0}$, with B_0 non-trivial, and Brownian stopping times τ. The random variable set-up consists of a pair of random variables X and Y. In each case there is an implicit probability triple and filtration $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ (which may change between the three set-ups, although we use the same symbol $\mathbb{P}$ in each case).

The idea is that we identify F_{T_1} with B_0 and X and F_{T_2} with B_τ and Y. The relationship between F and B is based on the fact that any martingale (and the forward price is a martingale under a risk neutral measure) can be written as a time-change of Brownian motion. Thus we can identify price process with law μ at T_1 and law ν at T_2 with solutions of the Skorokhod Embedding Problem for a Brownian

---

[5] Thus $\mathbb{P}(Y = f(x)|X = x) = (g(x) - x)/(g(x) - f(x)) = 1 - \mathbb{P}(Y = g(x)|X = x)$

motion started with law μ and stopped to have law ν. Moreover, since the payoff of the forward-start straddle depends only on the law of F_{T_1} and F_{T_2} (respectively B_0 and B_τ) we can reduce the problem further to an analysis of constructions of pairs of random variables X and Y such that $\mathbb{E}[Y|X] = X$. Hence, when we speak about a feasible model we will typically be referring to a pair (X, Y) with $X \sim \mu$ and $Y \sim \nu$, but this is connected via a solution of the SEP to $B_0 \sim \mu$ and $B_\tau \sim \nu$ and thence via a time-change to a financial model with $F_{T_1} \sim B_0$ and $F_{T_2} \sim B_\tau$.

The fact that $(F_t)_{t \geq 0}$ is a true martingale corresponds to τ being uniformly integrable, and in turn to the fact that (X, Y) satisfy the martingale condition $\mathbb{E}[Y|X] = X$. Finally, all these conditions relate to μ and ν having the same mean. Clearly, by a shift this mean can be taken to be zero, so that μ and ν are centred.

Let $\mathcal{M}$ be the set of feasible models, which, as discussed above, can be identified with the set of pairs of random variables with the correct marginals, and satisfying the martingale property:

$$\mathcal{M} = \mathcal{M}(\mu, \nu) = \{(X, Y) : X \sim \mu, Y \sim \nu, \mathbb{E}[Y|X] = X\}.$$

There is a simple condition which determines whether $\mathcal{M}$ is non-empty, namely that $\mu \preceq \nu$ in the sense of convex order (and we will use $\prec$ and $\preceq$ only in this sense), or equivalently $U_\mu(x) \leq U_\nu(x)$ uniformly in x, where for a centred probability measure χ, $U_\chi(x) = \mathbb{E}[|X - x| : X \sim \chi]$ is the potential. Note there is a one-to-one correspondence between centred probability measures and potential functions U with the properties U convex, $|U| \geq x$ and $\lim_{x \to \pm\infty} U(x) - |x| = 0$, see Chacon [5]. Then, provided $\mathcal{M}$ is non-empty we define

$$H(\mu, \nu) = \sup_{(X,Y) \in \mathcal{M}(\mu,\nu)} \mathbb{E}[|Y - X|],$$

and our primary concern is with identifying this object H.

In the exposition we will sometimes need to consider an iterated version of the problem, so for centred probability measures $\mu \preceq \nu \preceq \eta$ let

$$\mathcal{M}(\mu, \nu, \eta) = \{(X, Y, Z) : X \sim \mu, Y \sim \nu, Z \sim \eta, \mathbb{E}[Y|X] = X, \mathbb{E}[Z|Y] = Y\}.$$

Further, for a centred measure χ, let χ^m be the measure which is the law of Brownian motion started with law χ and run until the process hits the set $2^{-m}\mathbb{Z} = \{2^{-m}k; k \in \mathbb{Z}\}$, so $\chi^m = \mathcal{L}(B_\tau; B_0 \sim \chi, \tau = \inf\{u : B_u \in 2^{-m}\mathbb{Z}\})$. Equivalently, χ^m is the measure with potential U^m where $U^m \equiv U$ on $2^{-m}\mathbb{Z}$ and is obtained by linear interpolation over intervals $(k2^{-m}, (k+1)2^{-m})$ elsewhere.

Denote by δ_x the point mass at x. Finally, given increasing functions f, g with $f(x) < x < g(x)$, set $\tau_{f,g,x}$ to be the first time that Brownian motion leaves the interval $(f(x), g(x))$ where $B_0 = x$ (we write $\tau_{f,g,x} = \inf\{u : B_u \notin (f(x), g(x))|B_0 = x\}$) and let $\hat{\nu}(f, g, \mu) = \mathcal{L}(B_{\tau_{f,g,B_0}}; B_0 \sim \mu)$.

3 Monotonicity properties, and lack thereof, for H.

We begin with some simple properties of H, an example which shows that H also has some counter-intuitive properties and a useful lemma.

Lemma 4 (i) *For $\mu \prec \nu$, $H(\mu, \nu) \leq H(\delta_0, \mu) + H(\delta_0, \nu)$, and hence for centred probability measures we have $H(\mu, \nu) < \infty$.*

(ii) *If $\mu \preceq \nu \preceq \chi$ then $H(\mu, \nu) \leq H(\mu, \chi)$. Hence if $\nu_n \downarrow \nu$ then $H(\mu, \nu) \leq \liminf H(\mu, \nu_n)$.*

Proof: (i) Irrespective of the martingale condition, $\mathbb{E}[|Y - X|] \leq \mathbb{E}[|Y|] + \mathbb{E}[|X|] = H(\delta_0, \mu) + H(\delta_0, \nu)$. Note that $\mathbb{E}[|Y|] = U_{\mathcal{L}(Y)}(0)$ is independent of the joint law of X and Y.

(ii) If $\mu \prec \nu \prec \chi$ then there are constructions of random variables (X, Y, Z) such that $\mathbb{E}[Z|Y] = Y$ and $\mathbb{E}[Y|X] = X$ with the appropriate marginals. If we take a supremum over these constructions only then we find

$$H(\mu, \chi) \geq \sup_{\mathcal{M}(\mu, \nu, \chi)} \mathbb{E}[|Z - X|].$$

By the conditional version of Jensen's inequality $\mathbb{E}[|Z - X|] \geq \mathbb{E}[|Y - X|]$ and

$$\sup_{\mathcal{M}(\mu, \nu, \chi)} \mathbb{E}[|Z - X|] \geq \sup_{\mathcal{M}(\mu, \nu, \chi)} \mathbb{E}[|Y - X|] = \sup_{\mathcal{M}(\mu, \nu)} \mathbb{E}[|Y - X|] = H(\mu, \nu)$$

$\square$

Example 5 Given $\mu \preceq \nu \preceq \chi$ we have that $H(\mu, \nu) \leq H(\mu, \chi)$ and it would be nice to be able to conclude also that $H(\mu, \chi) \geq H(\nu, \chi)$. (Then we would have that H was monotonic in its first argument, which would facilitate approximating μ with a sequence μ_n.) However this is not the case, and we can have either $H(\mu, \chi) < H(\nu, \chi)$ or $H(\mu, \chi) > H(\nu, \chi)$.

The more usual and expected case is that $H(\mu, \chi) > H(\nu, \chi)$. For a simple example take $\nu \equiv \chi \neq \mu$. For an example in the less expected direction take $\mu = \delta_0$, ν to be the uniform measure on the two-point set $\{\pm 1\}$, and χ to place mass $1/2n$ at $\pm n$ and mass $1 - 1/n$ at the origin. Then $H(\delta_0, \chi) = 1$. However, if $Y \sim \nu$ and if we set $Z = nY$ with probability $1/n$ and $Z = 0$ otherwise, then $Z \sim \chi$, $\mathbb{E}[Z|Y] = Y$ and $H(\nu, \chi) \geq 2(n-1)/n$. (In fact it is easy to see that there is equality in this last expression.) Provided $n > 2$ we have $H(\delta_0, \chi) < H(\nu, \chi)$. $\square$

Recall the definition of ν^m and observe that $\nu \preceq \nu^m$.

Lemma 6 *Suppose $\mu \preceq \nu$. Then $H(\mu, \nu) \leq H(\mu, \nu^m) \leq H(\mu, \mu^m) + H(\mu^m, \nu^m) \leq H(\mu^m, \nu^m) + 2^{-m}$*

Proof: The only inequality which is not immediate is the middle one. Let $\sigma^m = \inf\{u \geq 0 : B_u \in 2^{-m}\mathbb{Z}; B_0 \sim \mu\}$ and let τ^m be an embedding of ν^m based on initial law μ. Then, necessarily, $\tau^m \geq \sigma^m$, and if $\theta_t(\omega)$ is the shift operator $\theta_t(\omega) = \sigma^m(\omega) + t$, if $\tilde{B}_t = B_{\theta_t}$ and if $\tilde{\tau} = \tau^m - \sigma^m$ then $\tilde{\tau}$ is an embedding of ν^m for the Brownian motion $\tilde{B}$ started with initial law $\tilde{B}_0 \sim \mu^m$. By the triangle inequality, $\mathbb{E}|B_{\tau^m} - B_0| \leq \mathbb{E}|B_{\tau^m} - B_{\sigma^m}| + \mathbb{E}|B_{\sigma^m} - B_0|$. But $\mathbb{E}|B_{\tau^m} - B_{\sigma^m}| = \mathbb{E}|\tilde{B}_{\tilde{\tau}} - \tilde{B}_0| \leq H(\mu^m, \nu^m)$ and σ^m is the unique, uniformly integrable embedding of μ^m for Brownian motion started with law μ so that $\mathbb{E}|B_{\sigma^m} - B_0| = H(\mu, \mu^m)$. $\square$

4 Upper bounds and a financial interpretation

Suppose we can find α, β and γ such that for all x and y we have $L(x, y) \leq 0$ where

$$L(x, y) = |y - x| - \alpha(x) - \beta(y) - \gamma(x)(x - y). \tag{1}$$

If so then for all elements of the sample space,

$$|Y - X| \leq \alpha(X) + \beta(Y) + \gamma(X)(X - Y),$$

and, taking expectations and using the martingale property,

$$H(\mu, \nu) \leq \int \alpha(x)\mu(dx) + \int \beta(y)\nu(dy).$$

The following simple example is a first illustration of the method and gives a sample result. For $k > 0$, $0 \leq k(|b| - 1/k)^2/2$ and so, with $b = Y - X$, and using $(Y - X)^2 = Y^2 - X^2 + 2X(X - Y)$, we have

$$|Y - X| \leq \frac{k}{2}\left(Y^2 - X^2 + 2X(X - Y)\right) + \frac{1}{2k}.$$

Hence, (recall we are assuming $\mathbb{E}[Y|X] = X$) for any k, $\mathbb{E}|Y - X| \leq (kA^2/2) + 1/2k$ where $A^2 = \mathbb{E}[Y^2] - \mathbb{E}[X^2] = \mathbb{E}[(Y - X)^2]$. Minimising over k we find $k = A^{-1}$, and so

$$\mathbb{E}|Y - X| \leq A \equiv \sqrt{\mathbb{E}[Y^2] - \mathbb{E}[X^2]}.$$

We can get this result directly (ie without writing down the pathwise inequality) just from Jensen's inequality:

$$\mathbb{E}|Y - X| \leq (\mathbb{E}[(Y - X)^2])^{1/2} = A$$

but one advantage of the method based on inequalities of the form $L(x, y) \leq 0$ is that the various terms can be meaningfully identified in the financial context as static hedging portfolios. Thus $\alpha(X)$ is a portfolio of options with maturity T_1, $\beta(Y)$ is a portfolio of options with maturity T_2, and $\gamma(X)(X - Y)$ is the gains from trade on the forward market from a strategy of going short $\gamma(X)$ forwards over the period $[T_1, T_2]$. It is also possible to identify when the bound is tight. In this case we must have $|Y - X| = 1/k = A$, so that $Y = X \pm 1/k$.

Lemma 7 *Consider the problem of hedging a forward-start straddle on the forward price F_t with payoff $|F_{T_1} - F_{T_2}|$. Suppose that α, β and γ are such that (1) holds, and that α and β are twice differentiable[6]. Then there is a super-replicating strategy involving puts and calls on F_t which costs $\int \alpha(x)\mu(dx) + \int \beta(y)\nu(dy)$ where μ is the law of F_{T_1} and ν is the law of F_{T_2}.*

Proof: By the arguments of Breeden and Litzenberger [3] it is possible to recreate a (sufficiently regular) payoff of $\Gamma(F_T)$ as a portfolio of put options with strike K:

$$\Gamma(F_T) = \Gamma(0) + F_T\Gamma'(0) + \int_0^\infty \Gamma''(k)(k - F_T)^+ dk.$$

---

[6]We need to give a suitable interpretation to α'' and β'', so that a weaker sufficient condition is that α and β are each the difference between a pair of convex functions. In fact for the optimal construction β is always convex.

Note that $(k - F_T)^+$ is the payoff of a put with strike k and maturity T.

Consider the strategy of purchasing a portfolio of puts with maturity T_1 which recreates a payoff $\alpha(F_{T_1})$ and a portfolio of puts with maturity T_2 which recreates a payoff $\beta(F_{T_2})$.

In addition, if at T_1 the price $F_{T_1} = x$, go short $\gamma(x)$ units of the forward over the period $[T_1, T_2]$.

The final value of this portfolio is $\alpha(F_{T_1}) + \beta(F_{T_2}) + \gamma(F_{T_1})(F_{T_1} - F_{T_2})$ which super-replicates $|F_{T_2} - F_{T_1}|$. Since the forward transaction is costless, the cost of the super-replicating strategy is as claimed. $\qquad\square$

Remark 8 (i) We should emphasise that our forward-start straddle is written on the forward price which we denote F_t. If the forward-start straddle is written on a traded security S_t which in a constant interest rate world has drift r, then we can set $F_t = e^{-rt}S_t$ and then $e^{-rT_2}|S_{T_2} - S_{T_1}| = |F_{T_2} - \lambda F_{T_1}|$ with $\lambda = e^{-r(T_2 - T_1)}$ a deterministic factor. Some of the ideas of this paper can be extended immediately to this situation, for example we replace (1) with $L(x, y) = |y - \lambda x| - \alpha(x) - \beta(y) - \gamma(x)(x - y)$, but other elements of the story cannot be generalised so easily. The case $\lambda = 1$ is already quite intricate, so we do not consider $\lambda \neq 1$.

(ii) In this article we concentrate on model-independent upper bounds for the prices of forward-start options. This naturally leads to the question as to the existence and form of lower bounds. Preliminary analysis has shown that the corresponding extremal process involves a trichotomy of potential values of Y for each possible value of X (in the nice case where μ and ν have densities). However, the form of the solution is much more involved, and we will not attempt to describe the solution here.

5   The Lagrangian Approach

We take a Lagrangian approach, which has proved useful in several papers on problems of this type, see, for example, Brown et al [4]. To motivate the analysis and explain the methods we begin the exposition by assuming that we are in the nice case where the functions we work with are differentiable, and the measures have densities. In particular, we suppose $\mu(dx) = \eta(x)dx$ and $\nu(dy) = \xi(y)dy$, and let the joint density of (X, Y) be $\rho(x, y)dxdy$. The problem is to maximise

$$\int \int |y - x|\rho(x, y)dxdy,$$

subject to the marginal and martingale conditions

$$\int \rho(x, y)dy - \eta(x) = 0, \quad \int \rho(x, y)dx - \xi(y) = 0, \quad \int \rho(x, y)(x - y)dy = 0.$$

If the constraints have Lagrange multipliers $\alpha(x), \beta(y), \gamma(x)$, then the problem becomes to maximise over ρ

$$\left[\int \alpha(x)\eta(x)dx + \int \beta(x)\xi(y)dy + \int \int \rho(x, y)L(x, y)dx\, dy \right]$$

where $L(x, y)$ is as given in (1). For the maximum to be finite we must have $L(x, y) \leq 0$, and the issue is to choose α, β and γ to make this true, in such a way that $\int \alpha(x)\eta(x)dx + \int \beta(x)\xi(y)dy$ is minimised.

From the dependence of L on y, for each x we expect there to be equality $L(x, y) = 0$ at two points $y = f(x)$ and $y = g(x)$ with $f(x) < x < g(x)$. At these points the y-derivative of L is zero. Hence $\beta'(g(x)) = \gamma(x) + 1$ and $\beta'(f(x)) = \gamma(x) - 1$.
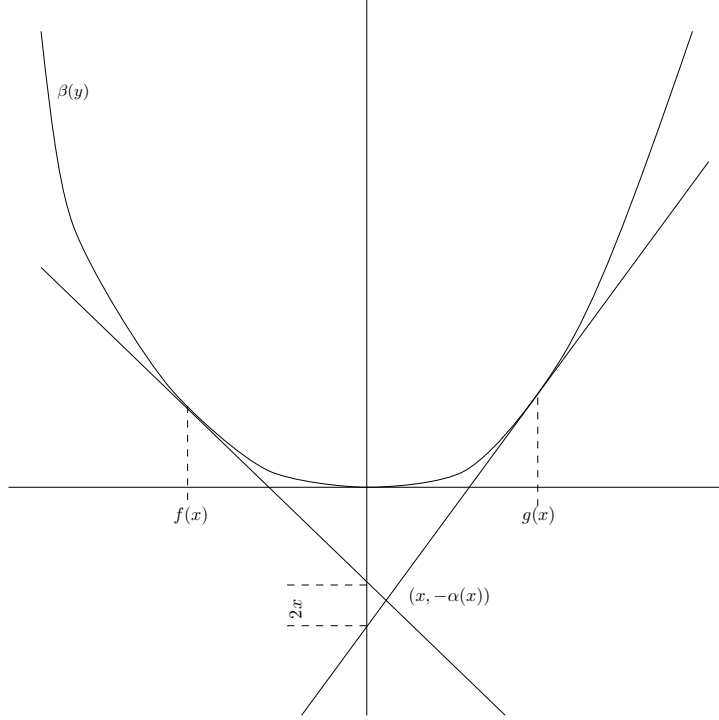

Figure 1: The relationship between the various quantities which can be derived from β. The points $f(x)$ and $g(x)$ are two values in the horizontal direction such that the difference in the slope of β at these two points is 2, and such that these tangents intersect at a point with horizontal coordinate x. The height of the intersection point is $-\alpha(x)$, and $\gamma(x)$ is such that the slopes of the tangents are $\gamma(x) \pm 1$.

The key insight is that the best way is to find suitable β is via its convex dual. The construction begins with a convex function $G(x)$ normalised such that $G(0) = 0 = G'(0-)$. Set ϕ to be the increasing function given (where G' is well defined) by $\phi(x) = G'(x)$, and define β via

$$\beta(y) = \sup_x \{xy - G(x)\}; \tag{2}$$

note that $\phi = (\beta')^{-1}$. Then $G(x) = x\phi(x) - \beta(\phi(x))$ and

$$g(x) = \phi(\gamma(x) + 1) \qquad\qquad f(x) = \phi(\gamma(x) - 1). \tag{3}$$

From the definition of L and g we have

$$0 \;=\; L(x, g(x))$$

9

$$
\begin{aligned}
&= \quad g(x) - x - \alpha(x) - \beta(g(x)) - \gamma(x)(x - g(x)) \\
&= \quad (1 + \gamma(x))\phi(1 + \gamma(x)) - \beta(\phi(1 + \gamma(x))) - \alpha(x) - x(1 + \gamma(x)) \\
&= \quad G(\gamma(x) + 1) - x - \alpha(x) - x\gamma(x). \tag{4}
\end{aligned}
$$

Similarly,

$$
0 = L(x, f(x)) = G(\gamma(x) - 1) + x - \alpha(x) - x\gamma(x) \tag{5}
$$

and subtracting these last two expressions we obtain

$$
2x = G(\gamma(x) + 1) - G(\gamma(x) - 1). \tag{6}
$$

If we then define H and γ via

$$
H(z) = (G(z + 1) - G(z - 1))/2; \qquad \gamma = H^{-1}. \tag{7}
$$

then (6) holds. Note that H is increasing so that γ is well defined and increasing, and since β' and ϕ are also increasing we have that g and f are increasing.

Finally, adding (4) and (5) we find

$$
\alpha(x) = -x\gamma(x) + \frac{1}{2}\left[G(\gamma(x) + 1) + G(\gamma(x) - 1)\right]. \tag{8}
$$

An alternative expression involving β is

$$
\begin{aligned}
\alpha(x) &= \quad (g(x) - x)\beta'(g(x)) - \beta(g(x)) \tag{9} \\
&= \quad (f(x) - x)\beta'(f(x)) - \beta(f(x)). \tag{10}
\end{aligned}
$$

For a given convex G this completes the construction of a trio (α, β, γ) for which $L(x, y)$ given by (1) satisfies $L \leq 0$.

In determining (α, β, γ) it is convenient to assume that G is continuously differentiable and strictly convex. However, this is by no means necessary, and the only issues are in choosing the appropriate inverse γ to H in (7), which then enters the definition of α. The easiest way to determine the correct form for the quantities α and γ (and f and g) is via the graphical representation in Figure 1.

Theorem 9 *Let μ and ν be a pair of centred probability measures which are increasing in convex order. Let G be convex with $G(0) = 0 = G'(0-)$, and define $\beta \equiv \beta_G$ and $\alpha \equiv \alpha_G$ via (2) and (8), where $\gamma \equiv \gamma_G$ is defined in (7). Then, for all $(X, Y) \in \mathcal{M}(\mu, \nu)$*

$$
\mathbb{E}[|Y - X|] \leq \int \alpha(x)\mu(dx) + \int \beta(y)\nu(dy). \tag{11}
$$

Proof: We simply need to show that $L(x, y) \leq 0$. To see this, for $y > x$ (the case $y < x$ is similar)

$$
\begin{aligned}
&L(x, y) \\
&= \quad (y - x) + x\gamma(x) - \frac{1}{2}\left[G(\gamma(x) + 1) + G(\gamma(x) - 1)\right] - \beta(y) + \gamma(x)(y - x) \\
&= \quad y(1 + \gamma(x)) - G(\gamma(x) + 1) - \beta(y)
\end{aligned}
$$

But $\beta(y) \geq zy - G(z)$ for all z including $z = 1 + \gamma(x)$. Hence $L(x, y) \leq 0$.

Thus, given G we have a bound on $\mathbb{E}[|Y - X|]$ of the form

$$\mathbb{E}|Y - X| \leq \int \alpha(x)\mu(dx) + \int \beta(y)\nu(dy)$$

where $\alpha = \alpha_G$ and $\beta = \beta_G$.

Remark 10 The normalisation of G such that $G(0) = G'(0-) = 0$ is convenient, but not important. If instead we set $\tilde{G}(x) = G(x - c) + d$, then we find $\tilde{\beta}(y) = \beta(y) + cy - d$, $\tilde{\gamma}(x) = \gamma(x) + c$ and $\tilde{\alpha}(x) = \alpha(x) - cx + d$, so that the bound in (11) is unchanged. In finance terms, any super-replicating strategy that involves options positions at times T_1 and T_2 and a forward position over $[T_1, T_2]$ can be trivially modified by adding a long position at T_1, a short position at T_2 and an offsetting forward position.

Remark 11 The construction begins with G and this is the primary object used to calculate g and f from (3). Combining these with μ we can deduce the law $\hat{\nu} = \hat{\nu}_{f,g,\mu}$ of Y:

$$\hat{\nu}((-\infty, y]) = \mathbb{P}(Y \leq y) = \mu((-\infty, g^{-1}(y)]) + \int_{g^{-1}(y)}^{f^{-1}(y)} \frac{g(x) - x}{g(x) - f(x)}\mu(dx) \quad (12)$$

Alternatively, given the convex duality between β and G, we can also start with β as the primitive object. In this way we can choose convex functions G and initial laws μ so that there is equality in (11) and hence optimality, for a certain law $\hat{\nu}$. In this sense it is easy to produce examples for which the bound is tight. However, the real aim is to start with laws μ and ν and to construct G and the bound. This will prove to be much harder.

Remark 12 Recall the definitions of f and g, which for the purposes of this remark we assume to be differentiable. Then they satisfy a certain consistency condition. From (3) we have

$$\beta'(g(x)) - 1 = \gamma(x) = \beta'(f(x)) + 1 \quad (13)$$

and so

$$\beta''(g(x))g'(x) = f'(x)\beta''(f(x)) \quad (14)$$

Also, from (9) and (10) we obtain

$$\beta(g(x)) - g(x)\beta'(g(x)) - (\beta(f(x)) - f(x))\beta'(f(x)) + x(\beta'(g(x)) - \beta'(f(x))) = 0. \quad (15)$$

Using (13), differentiating, and then using (14) we get

$$\frac{2}{g(x) - f(x)} = \beta''(g(x))g'(x) = f'(x)\beta''(f(x)) \quad (16)$$

so that in the appropriate domains

$$\beta''(y) = \frac{2}{f'(f^{-1}(y))(g(f^{-1}(y)) - y)} \quad (17)$$

and

$$\beta''(y) = \frac{2}{g'(g^{-1}(y))(y - f(g^{-1}(y)))}. \qquad (18)$$

Alternatively we can rewrite (15) as

$$0 = \int_{f(x)}^{g(x)} (y - x)\beta''(y)dy \qquad (19)$$

which after substituting with (17) and some manipulations yields

$$0 = \int_{g^{-1}(y)}^{f^{-1}(y)} \frac{(g(z) - f^{-1}(y))}{g(z) - f(z)} dz. \qquad (20)$$

Substituting using the first inequality in (16), then changing variable and integrating, and finally using (13) we obtain

$$\int_{g^{-1}(f(x))}^{x} \frac{2}{g(z) - f(z)} dz = \int_{f(x)}^{g(x)} \beta''(y)dy = 2, \qquad (21)$$

so that

$$1 = \int_{g^{-1}(y)}^{f^{-1}(y)} \frac{1}{g(z) - f(z)} dz, \qquad (22)$$

and then (20) is equivalent to

$$\int_{g^{-1}(y)}^{f^{-1}(y)} \frac{g(z)}{g(z) - f(z)} dz = f^{-1}(y) \qquad (23)$$

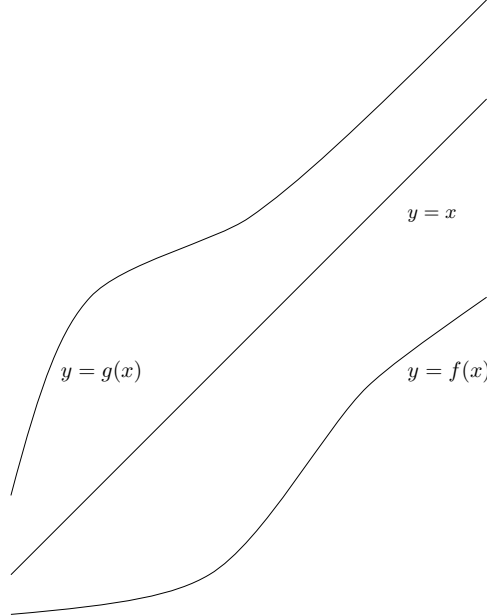with a related expression interchanging the roles of f and g.

Figure 2: A representation of functions f and g.

In particular, given f and g we can define a candidate convex function β via (17) and (18), but when the construction of Section 5 is applied to this candidate

β we will only recover the original f and g if (23) holds. Equations (22) and (23) play the role of global consistency conditions on the functions f, g which determines whether they are associated with optimal constructions. Note that it is a non-local condition in that it relates f and g over whole intervals and not at isolated points. We will use this consistency condition to select the optimal solution (f, g) from the many which lead to embeddings.

Remark 13 The bound is attained if $Y \in \{f(X), g(X)\}$, or equivalently if the stopping rule τ is of the form $\tau = \tau(f, g, B_0) = \inf\{u : B_u \in \{f(B_0), g(B_0)\}\}$. In that case we have an alternative representation of the bound[7] as

$$2\mathbb{E}\left[\frac{(g(X) - X)(X - f(X))}{(g(X) - f(X))}\right], \qquad (24)$$

at least in the case where μ and ν have densities and f, g and their inverses are continuous and differentiable. The expression (24) follows directly from the fact that $\mathbb{P}(Y = f(x)|B_0 = x) = (g(x) - x)/(g(x) - f(x))$. This expression can also be derived via calculus from (12) using the definitions of α and β.

6 Examples

6.1 Example: Quadratic functions $G(x) = x^2/2k$.

In this case $\beta(y) = ky^2/2$, $H(z) = z/k$, $\gamma(z) = zk$, $\alpha(x) = (1 - k^2x^2)/2k$. We immediately recover the result in the opening remarks of Section 4: $\mathbb{E}[|Y - X|] \leq k\mathbb{E}[Y^2 - X^2]/2 + 1/2k$. This result can be optimised by appropriate choice of k.

We have $g(x) = x + k^{-1}$ and $f(x) = x + k^{-1}$. If $k = 1$ and $\mu \sim U[-1, 1]$, then recalling that $\hat{\nu}(f, g, \mu) = \mathcal{L}(B_\tau : B_0 \sim \mu, \tau = \inf\{u > 0 : B_u \in \{f(B_0), g(B_0)\}\})$, we have that $\hat{\nu} \sim U[-2, 2]$ and $H(\mu, \hat{\nu}) = 1$.

6.2 Example: Entropy

For this example it is natural to assume that X and Y are non-negative random variables, scaled to have unit mean.

Take $G(x) = Ae^{x/\xi}$. Then

$$\beta(y) = \xi(y \ln y - y \ln(A/\xi) - y)$$

$$\gamma(w) = \xi \ln(w/A \sinh(1/\xi))$$

and

$$\alpha(x) = -\xi(x \ln x) + \xi x \ln(A \sinh(1/\xi)) + x \coth(1/\xi)$$

The bound is

$$\mathbb{E}[|Y - X|] \leq \mathbb{E}[\beta(Y) + \alpha(X)] = \xi\mathbb{E}[Y \ln Y - X \ln X] + J(\xi)$$

---

[7]Observe that $f(x) \leq x \leq g(x)$ and so $(g(x) - x)(x - f(x))/(g(x) - f(x)) \leq \min\{g(x) - x, x - f(x)\}$. Hence in (24) it is appropriate to use the convention that $0 \times 0/0 = 0$.

where
$$J(\xi) = \xi \ln \xi - \xi + \xi \ln(\sinh(1/\xi)) + \coth(1/\xi).$$

Note that J is a decreasing convex function on $\mathbb{R}^+$, with $J(0) = 2$. (The fact that $J(0) = 2$ corresponds to the trivial bound $\mathbb{E}|Y - X| \leq \mathbb{E}|Y| + \mathbb{E}|X| = 2$.)

Let $\tilde{J}$ be the convex dual to J, so that

$$\tilde{J}(z) = \inf_{\xi > 0}(\xi z + J(\xi)).$$

Then

Proposition 14 *Let X and Y be positive random variables each with unit mean and such that $\mathbb{E}[Y|X] = X$. Suppose that $\mathbb{E}[Y \ln Y - X \ln X] \leq \Delta$. Then*

$$\mathbb{E}|Y - X| \leq \tilde{J}(\Delta).$$

The bound is tight, in the sense that for each $\Delta > 0$ there exists a pair (X, Y) with $\mathbb{E}[Y \ln Y - X \ln X] = \Delta$ for which $\mathbb{E}|Y - X| = \tilde{J}(\Delta)$.

Corollary 15 *We have $J(\xi) \leq \min\{1/(2\xi), 2\}$, and then $\tilde{J}(z) \leq \sqrt{2z} \wedge 2$. It follows that $\mathbb{E}|Y - X| \leq \sqrt{2\Delta} \wedge 2$.*

Corollary 16 *If X and Y satisfy the hypotheses of Proposition 14 but have mean c then $\mathbb{E}|Y - X| \leq c\tilde{J}(\Delta/c)$.*

Note that, unlike in the quadratic example, the pre-multiple A plays no role in the final bound. Note further that as for the quadratic example, α takes the same functional form as β, so we get this very nice inequality involving the entropies of the two distributions. This makes the resulting inequality particularly attractive, but is a special feature of these examples.

We discuss the financial implications of this bound in Section 10 below. For this example we have that $g(x) = xe^{1/\xi}/(\xi \sinh(1/\xi))$ and $f(x) = xe^{-1/\xi}/(\xi \sinh(1/\xi))$. Both these functions are linear which makes it particularly simple to construct examples where the bound is attained. If X has an exponential distribution, then the construction yields Y which is a mixture of two exponentials.

6.3   Example: Multiplicity of Embeddings

Suppose $\mu \sim U[-1, 1]$ and $\nu \sim U[-2, 2]$. We know from Example 6.1 that $H(\mu, \nu) = 1$, and that for the optimal construction $f(x) + 1 = x = g(x) - 1$. Our goal in this example is to show that this is not the only pair (f, g) for which $\hat{\nu}(f, g, \mu) = \nu$.

Fix $a \in (-1, 1)$ and suppose we have increasing functions $f : [-1, 1] \mapsto [-2, a]$ and $g : [-1, 1] \mapsto [a, 2]$. For $\hat{\nu}(f, g, \mu)$ to equal ν we must have

$$g'(z) = \frac{2(z - f(z))}{g(z) - f(z)}, \qquad f'(z) = \frac{2(g(z) - z)}{g(z) - f(z)}.$$

Define $v_a(z) = (4 - 4az + a^2)^{1/2}$. Then, recall $g(-1) = a = f(1)$ and $g(1) = 2 = -f(-1)$, $g(z) = z + a/2 + v_a(z)/2$ and $f(z) = z + a/2 - v_a(z)/2$. For each $a \in (-1, 1)$ this construction defines an embedding of ν. However at most one of these

14

constructions can be associated with the embedding which maximises $\mathbb{E}|Y - X|$, and this will be the one for which f and g satisfy the global consistency condition.

We can define a candidate β from (19) and then $\beta''(y) = w_a(y)$ where $w_a(y) = (1 - ay + a^2)^{1/2}$. However, if we consider (20) for $y = a$ we get

$$\int_{g^{-1}(a)}^{f^{-1}(a)} \frac{g(x)}{(g(x) - f(x))} dx = \int_{-1}^{1} \frac{x + a/2 + v_a(x)/2}{v_a(x)} dx = 1 + \frac{2}{3}a$$

This is equal to $f^{-1}(a) = 1$ if and only if $a = 0$, so that out of the many pairs (f, g) which embed (μ, ν) only the pair defined from $a = 0$ is consistent with a construction based upon a convex function β.

7   Constructing bounds given the marginals

In the previous section we derived upper bounds on $\mathbb{E}[|Y - X|]$ by considering families of functions derived from a convex function G. There is a one-to-one relationship between G and β, and so from either it is possible to deduce expressions for α and γ, and thence, at least in the regular case where G and β are smooth and strictly convex, we can obtain expressions for the monotonic functions f and g. Finally, conditional on the law μ for X we can find a bound for $\mathbb{E}[|Y - X|]$.

The construction gives a bound for any feasible law ν of Y but the bound is attained only for a particular law $\nu = \hat{\nu}(f, g, \mu)$.

The issue is to reverse this construction, and given μ and ν to find G or β for which we can construct a best bound. Alternatively, given μ and ν we want to minimise the right-hand-side of (11) over G and more especially to prove this gives the lowest possible upper bound on $H(\mu, \nu)$. A related problem is to find functions $f(x) < x < g(x)$, such that a construction of the form $Y \in \{f(X), g(X)\}$ is optimal for the problem. This is complicated by the fact that it is not sufficient simply to find f, g such that if $X \sim \mu$ and both $\mathbb{E}[Y|X] = X$ and $Y \in \{f(X), g(X)\}$ then $Y \sim \nu$.

Lemma 17 *Suppose f, g are strictly increasing, continuous and differentiable and f, g solve (22) and (20). Then if β is given by the solution of (16), G is the convex dual of β and $\alpha = \alpha_G$ then*

$$H(\mu, \nu) = \int \alpha(x)\mu(dx) + \int \beta(y)\nu(dy).$$

where $\nu = \hat{\nu}(f, g, \mu)$.

Proof: Given f and g satisfying (22) and (20) we can define β via (16), or the equivalent expressions (17) or (18). Given that (17) and (18) define β'' in two different ways it should be checked that these two definitions do not lead to an inconsistency or self-contradiction. In fact differentiation of (22) shows that (17) follows from (18) and vice-versa.

It follows from (20) and (17) that

$$0 = \int_{f(x)}^{g(x)} (y - x)\beta''(y) dy.$$

Integrating the right-hand-side it follows that (15) holds and we can define α via either (9) or (10). From the equivalence of these two representations (and $\beta''(g(x)g'(x)) = \beta''(f(x)f'(x))$) we deduce as in (21) that $2 = \beta'(g(x)) - \beta'(f(x))$. Let $\gamma(x) = \beta'(g(x)) - 1 = \beta'(f(x)) + 1$, then we have a triple α, β, γ. Moreover, since the tangents to β at $f(x)$ and $g(x)$ intersect at $(x, -\alpha(x))$ it is clear that when we define $\phi(\gamma(x) + 1)$ and $\phi(\gamma(x) - 1)$ we recover g and f respectively. By hypothesis, $\tau(f, g, \mu)$ embeds ν so that $H(\mu, \nu) = H(\mu, \hat{\nu}(f, g, \mu)) \geq \int \alpha(x)\mu(dx) + \int \beta(y)\nu(dy)$, the inequality following from the fact that $H(\mu, \nu)$ is a supremum over all embeddings. The reverse inequality follows from Theorem 9. □

The lemma provides a partial result, but it still remains to show that it is possible to find f, g which solve (22) and (20) and the embedding condition (12). It seems very difficult to exhibit f, g which solve this problem. Instead we will approximate ν with a discrete distribution for which we can prove that an appropriate function β exists, and derive the required result by taking limits.

8 Optimal upper bounds

The goal of this section is to find the value of $H(\mu, \nu)$ for arbitrary measures on $\mathbb{R}^+$, by finding an upper bound, and by showing the bound is attained. The approach is to begin with a point mass μ and a discrete measure ν, and to progress to the full problem via a series of extensions.

8.1 The discrete case: preliminary results

Suppose that $X \sim \mu$ and $Y \sim \nu$ (with $\mu \preceq \nu$) are discrete, centred random variables with finite support. Denote the atoms by $\mu_i = \mu(\{x_i\})$ and $\nu_j = \nu(\{y_j\})$, where the points $\{x_i\}$ and $\{y_i\}$ are ordered such that $x_1 < x_2 < \cdots < x_m$ and $y_1 < y_2 < \cdots < y_n$. The problem is to find $\sup_{(X,Y) \in \mathcal{M}(\mu,\nu)} \mathbb{E}|Y - X|$. In this simple setting this can be written as a finite linear programme:

$$\max_{\rho_{ij}} \left(\sum_{i,j} \rho_{ij} |y_j - x_i| \right) \tag{25}$$

subject to the constraints

$$\sum_j \rho_{ij} = \mu_i, \quad \sum_i \rho_{ij} = \nu_j, \quad \sum_j \rho_{ij}(x_i - y_j) = 0, \quad \rho_{ij} \geq 0. \tag{26}$$

The associated dual problem is to find

$$\min_{\alpha_i, \beta_j \gamma_i} \left(\sum_i \alpha_i \mu_i + \sum_j \beta_j \nu_j \right),$$

where $\alpha_i, \beta_j, \gamma_i$ are chosen to satisfy $L(x_i, y_j) \leq 0$ for all i and j, where, in turn,

$$L(x_i, y_j) = |x_i - y_j| - \alpha_i - \beta_j - \gamma_i(x_i - y_j).$$

By the complementary slackness condition we have that for an optimum $\rho_{ij}L(x_i, y_j) = 0$.

Given the constants β_j we can define a function $\beta(y)$ via $\beta(y_j) = \beta_j$ and by linear interpolation between these points, with $\beta(y) = \infty$ outside $[y_1, y_n]$. Then also we can define $L(x_i, y) = |x_i - y| - \alpha_i - \beta(y) - \gamma_i(x_i - y)$.

The primal problem is feasible and therefore has a solution and the values of the primal and dual problems are equal, (for this fundamental result see Gale, Kuhn and Tucker [9], or for a recent treatment Vanderbei [19]).

Lemma 18 *The solution of the linear programme is such that $\beta(y)$ is convex in y and γ_i is increasing in i. Further, if $y_j > x_i$ and $\rho_{ij} > 0$ then $\rho_{kl} = 0$ for all (k, l) for which $(k < i, l > j)$ and if $y_j < x_i$ and $\rho_{ij} > 0$ then $\rho_{kl} = 0$ for all (k, l) for which $(k > i, l < j)$.*

Proof: Suppose $\beta(y)$ is not convex. Then for some $j \in \{2, \ldots, n-1\}$

$$\beta_j > \beta_{j+1} \frac{(y_j - y_{j-1})}{(y_{j+1} - y_{j-1})} + \beta_{j-1} \frac{(y_{j+1} - y_j)}{(y_{j+1} - y_{j-1})}.$$

Fix i and suppose first that $x_i \leq y_j$. Then from the fact that $L(x_i, y_k) \leq 0$ for $k = j \pm 1$ we obtain

$$\begin{aligned}
\beta_{j+1} &\geq \alpha_i - (1 + \gamma_i)x_i + (1 + \gamma_i)y_{j+1} \\
\beta_{j-1} &\geq \alpha_i - (1 + \gamma_i)x_i + (1 + \gamma_i)y_{j-1}
\end{aligned}$$

and we conclude that $\beta_j > \alpha_i - (1 + \gamma_i)x_i + (1 + \gamma_i)y_j$. Hence $L(x_i, y_j) < 0$ and $\rho_{ij} = 0$.

A similar argument (but replacing $(\gamma_i + 1)$ with $(\gamma_i - 1)$) applies if $x_i > y_j$ and then $\rho_{ij} = 0$ for all i. Hence $\sum_i \rho_{ij} = 0$, a contradiction.

Now consider the monotonicity of γ. We want to show that if $x_k > x_i$ then $\gamma_k > \gamma_i$. We consider two cases depending on whether $(x_k, -\alpha_k)$ lies above or below the tangent to β with slope $\gamma_i + 1$.

Suppose $(x_k, -\alpha_k)$ lies strictly below the line $y = -\alpha_i + (1 + \gamma_i)(x - x_i)$. This condition can be rewritten as $(\alpha_i + x_i(1 + \gamma_i)) < (\alpha_k + x_k(1 + \gamma_i))$.

We know there exists $y > x_k$ for which $L(x_k, y) = 0$. (In particular, there exists $\rho_{kj} > 0$ for which the associated $y_j > x_k$ and then $L(x_k, y_j) = 0$.) Then

$$\begin{aligned}
0 = L(x_k, y) &= (y - x_k)(1 + \gamma_k) - \alpha_k - \beta(y) \\
&= [(y - x_i)(1 + \gamma_i) - \alpha_i - \beta(y)] + [(y - x_k)(\gamma_k - \gamma_i)] \\
&\quad + [(\alpha_i + x_i(1 + \gamma_i)) - (\alpha_k + x_k(1 + \gamma_i))].
\end{aligned}$$

The first and last of the square-bracketed terms are negative since $L(x_i, y) \leq 0$ and by the hypothesis that $(x_k, -\alpha_k)$ lies below the tangent, and hence $\gamma_k > \gamma_i$.

If $(x_k, -\alpha_k)$ lies at or above the line $y = \alpha_i + (1 + \gamma_i)(x - x_i)$, then it must lie strictly above the tangent to β with slope $\gamma_i - 1$, so we must have $(\alpha_i + x_i(\gamma_i - 1)) > (\alpha_k + x_k(\gamma_i - 1))$. Then by a similar argument to before we find for some $y < x_k$ that $0 = L(x_k, y) > (x_k - y)(\gamma_i - \gamma_k)$.

Finally, suppose $\rho_{ij} > 0$ for $y_j > x_i$ and $\rho_{kl} > 0$ for $(k > i, l < j)$. We want to obtain a contradiction. By definition,

$$0 \geq L(x_k, y_j) + L(x_i, y_l)$$

17

$$= ((y_j - x_k)(1 + \gamma_k) - \beta_j - \alpha_k) + ((y_l - x_i)(1 + \gamma_i) - \beta_l - \alpha_i)$$
$$= L(x_i, y_j) + L(x_k, y_l) + (\gamma_k - \gamma_i)(y_j - y_l)$$
$$= (\gamma_k - \gamma_i)(y_j - y_l) > 0$$

The reverse case for $y_j < x_i$ is similar. $\qquad\qquad\square$

In the discrete case, β is piecewise linear or equivalently β'' is a purely atomic measure. For this reason $\phi \equiv (\beta')^{-1}$ is not uniquely defined and the same applies to f and g. For this reason we need an alternative parameterisation. The same issue can arise whenever μ has atoms, and in these cases it is convenient to introduce some independent randomisation.

Define $F_X(x) = \mu((-\infty, x])$ and let U be a uniform random variable on $[0, 1]$. Then $F_X^{-1}(U) \sim \mu$, and our approach for considering the case where X is not a continuous random variable is to condition on U rather than X, and to define a trio of increasing functions, $p < q < r$ with domain $[0, 1]$. In particular, we suppose $B_0 \equiv X = q(U)$ (so that $q \equiv F_X^{-1}$), and we try to find $p : [0, 1] \mapsto \mathbb{R}$ and $r : [0, 1] \mapsto \mathbb{R}$ such that if

$$\tau_{p,q,r} = \inf\{t : B_t \notin (p(U), r(U)) | B_0 = q(U)\} \tag{27}$$

then $Y \equiv B_\tau \sim \nu$. The relationships between f, g and p, r are that $f \equiv p \circ q^{-1}$ and $g \equiv r \circ q^{-1}$.

The embedding condition (recall (12)) becomes

$$\nu((-\infty, y]) = \int_0^1 du I_{\{r(u) \le y\}} + \int_0^1 du I_{\{p(u) \le y < r(u)\}} \frac{r(u) - q(u)}{r(u) - p(u)} \tag{28}$$

Note that he embedding condition is easiest to express in terms of the functions p, q and r, whereas it is more natural to describe the 'global consistency condition' as conditions on f and g.

8.2 The discrete case: determining p, q and r for the case of constant X

Suppose that $\mu = \delta_x$, the unit mass at x. If $\nu = \delta_x$ then we take $p(u) = q(u) = r(u) = x$. Otherwise, suppose that Y has law mean x and takes values $y_{k_1} < \ldots < y_{k_m} < x \le y_{j_1} < \ldots < y_{j_n}$ with probabilities $\nu_{k_1}, \ldots, \nu_{k_m}$ and $\nu_{j_1}, \ldots \nu_{j_n}$.

The aim is to construct increasing functions $p(u) < x < r(u)$ such that if $U \sim U[0, 1]$ and $\tau(u) = \inf\{s \ge 0 : B_s \in \{p(u), r(u)\}\}$ then $B_{\tau(U)} \sim \nu$. (The resulting construction is the analogue of Skorokhod's original solution of the SEP, Skorokhod [17]).

The construction proceeds by induction: clearly if $m = 1 = n$, then we take $p \equiv p(u) = y_{k_1}$ and $r \equiv r(u) = y_{j_1}$ and the martingale condition forces $\nu_{j_1} = (x - y_{k_1})/(y_{j_1} - y_{k_1})$. Note that in this case

$$\mathbb{E}|Y - X| = 2\frac{(r - x)(x - p)}{(r - p)} = 2\int_0^1 \left[\frac{(r(u) - x)(x - p(u))}{r(u) - p(u)} \right] du.$$

So suppose $m + n > 2$. Let $u_1 = (y_{j_1} - y_{k_1}) \min\{\nu_{j_1}/(x - y_{k_1}), \nu_{k_1}/(y_{j_1} - x)\}$. For $u \le u_1$ we set $p(u) = y_{k_1}$ and $r(u) = y_{j_1}$. Without loss of generality, suppose

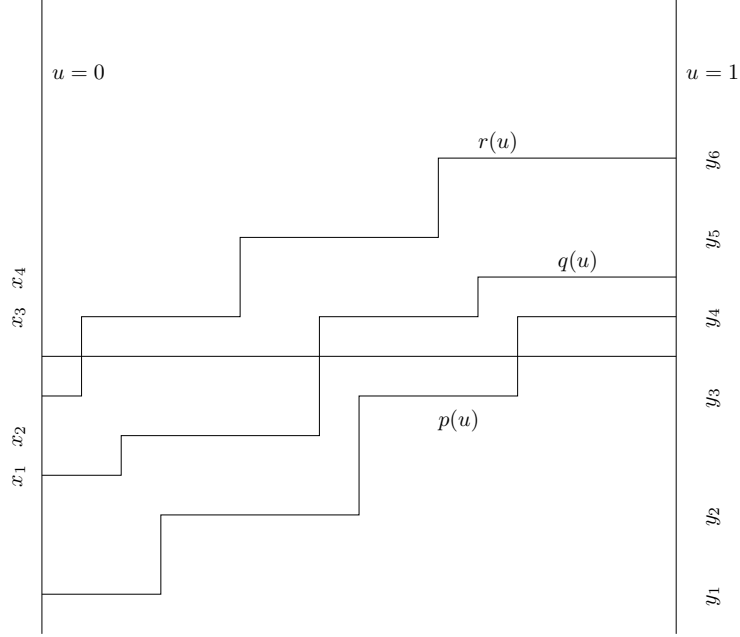Figure 3: The functions p, q, r in the discrete case. In this example the measure ν places mass on $\{y_1 < \cdots < y_6\}$ and μ places mass on $\{x_1 < \cdots < x_4\}$.

$\nu_{j_1}/(x - y_{k_1}) \leq \nu_{k_1}/(y_{j_1} - x)$. (This will necessarily be the case if $y_{j_1} = x$.) Then $\mathbb{P}(Y = y_{j_1}, U \leq u_1) = u_1(x - y_{k_1})/(y_{j_1} - y_{k_1}) = \nu_{j_1}$ and $\mathbb{P}(Y = y_{k_1}, U \leq u_1) = u_1(y_{j_1} - x)/(y_{j_1} - y_{k_1}) \leq \nu_{k_1}$. Conditional on $U \leq u_1$ we have embedded the mass at y_{j_1} and some of the mass at y_{k_1}, and so conditional on $U > u_1$ we must have that Y does not take the value y_{j_1}. Since, U conditioned on $U > u_1$ is again a uniform random variable we can use the inductive hypothesis to complete the construction.

In this way we construct increasing functions p and r with $p(0) = y_{k_1}, p(1) = y_{k_m}, r(0) = y_{j_1}, r(1) = y_{j_n}$. It also follows that $\mathbb{E}[|Y - X|] = u_1\mathbb{E}[|Y - X||U \leq u_1] + \mathbb{E}[|Y - X|; U > u_1]$, and applying the inductive hypothesis to the latter we again get

$$\mathbb{E}|Y - X| = 2\int_0^1 \left[\frac{(r(u) - x)(x - p(u))}{r(u) - p(u)} \right] du.$$

8.3 The discrete case: determining p, q and r for the case of general X

The extension to random variables X taking finitely many values is straightforward — if $X = q(U)$ then conditioning on the value $X = x$ is equivalent to conditioning on $q^{-1}(x-) < U \leq q^{-1}(x+)$ — and then the solutions for individual x can be pasted together. The results of Lemma 18 concerning where the joint measure ρ places mass are sufficient to ensure that r and p from this concatenation of solutions are increasing.

For discrete measure we have $X = q(U)$ and $Y \in \{p(U), r(U)\}$, and by con-

struction $Y \sim \nu$. For the optimal p, r we have

$$H(\mu, \nu) = 2\mathbb{E}\left[\frac{(r(U) - q(U))(q(U) - p(U))}{r(U) - p(U)}\right].$$

In particular, if

$$\mu = \sum_{i=1}^{m} \mu_i \delta_{x_i}, \qquad\qquad \nu = \sum_{j=1}^{n} \nu_j \delta_{y_j},$$

then let $(\eta_j^{(i)})_{1 \le j \le n}$ be the distribution on $\{y_j\}_{1 \le j \le n}$ given by $\eta_j^{(i)} = \rho_{ij}/\mu_i$. For each i we use the solution of Section 8.2 to produce functions $p_i(u) < x_i < r_i(u)$, such that, if $B_0^{(i)} \sim \delta_{x_i}$, $U_i \sim U[0,1]$ and $\tau_i(u) = \inf\{s \ge 0 : B_s^{(i)} \notin (p_i(u), r_i(u))\}$ then $B_{\tau(U_i)}^{(i)} \sim \eta^{(i)}$.

Now, with $q = F_X^{-1}$ and $U \sim U[0,1]$, we define p and r via

$$(p(u), r(u)) = \left(p_i\left(\frac{u - \sum_{l<i}\mu_l}{\mu_i}\right), r_i\left(\frac{u - \sum_{l<i}\mu_l}{\mu_i}\right)\right) \qquad \sum_{l<i}\mu_l < u \le \sum_{l\le i}\mu_l.$$

Then, the condition from Lemma 18 that $y_j < x_i$ and $\rho_{ij} > 0$ implies $\rho_{i+1,l} = 0$ for $l < j$ ensures that p and r so defined are increasing.

8.4 General bounded measures by approximation

The idea to cover general centred measures is to approximate μ and ν with finite measures μ^m and ν^m. For these discrete problems we find the associated increasing p^m, q^m, r^m. We have to show that these sequences converge and that the limits p, q, r are associated with a construction which embeds ν and is optimal.

Suppose that X and Y have bounded support, and suppose μ^m and ν^m are the approximations for μ and ν with support $2^{-m}\mathbb{Z}$ (recall Section 2 where η^m is defined as an approximation of η from above), and let p^m, q^m, r^m be the associated increasing functions, the construction of which is as described in Section 8.3. For each fixed m the pair (X^m, Y^m) attains $H(\mu^m, \nu^m)$.

By Helley's Selection Theorem (eg Billingsley [2]) there exists a subsequence down which each of p^m, q^m, r^m and their inverses converge to p, q, r, p^{-1}, q^{-1} and r^{-1}, at least at points of continuity of the limit functions. Write P^m for the inverse to p^m with similar expressions for q^m, r^m, p, q and r.

We have that

$$\nu^m((-\infty, y]) = \int_0^1 du I_{\{r^m(u) \le y\}} + \int_0^1 du I_{\{p^m(u) \le y < r^m(u)\}} \frac{r^m(u) - q^m(u)}{r^m(u) - p^m(u)} \quad (29)$$

and that $\nu^m((-\infty, y]) \to \nu((-\infty, y])$ at least at continuity points of ν. Moreover,

$$\int_0^1 du I_{\{u \le R^m(y)\}} \to \int_0^1 du I_{\{u \le R(y)\}}$$

since $R^m(y) \to R(y)$ and

$$\int_0^1 du I_{\{p^m(u) \le y < r^m(u)\}} \frac{(r^m(u) - q^m(u))}{r^m(u) - p^m(u)} \to \int_0^1 du I_{\{p(u) \le y < r(u)\}} \frac{(r(u) - q(u))}{r(u) - p(u)}$$

20

since the limit function has only countably many discontinuities, and except at these discontinuities the integrand converges (and is bounded). It follows that (28) is satisfied, and $\tau = \tau_{p,q,r}$ embeds ν.

It remains to show that this construction is optimal. Even if it is not we have the bound

$$H(\mu,\nu) \geq \mathbb{E}[|B_\tau - B_0|] = 2\mathbb{E}\left[\frac{(r(U) - q(U))(q(U) - p(U))}{r(U) - p(U)}\right]. \tag{30}$$

On the other hand, by Lemma 4, $H(\mu,\nu) \leq \liminf H(\mu^m, \nu^m)$. But

$$\begin{aligned}
H(\mu^m, \nu^m) &= 2\mathbb{E}\left[\frac{(r^m(U) - q^m(U))(q^m(U) - p^m(U))}{r^m(U) - p^m(U)}\right] \\
&\to 2\mathbb{E}\left[\frac{(r(U) - q(U))(q(U) - p(U))}{r(U) - p(U)}\right]
\end{aligned}$$

by bounded convergence, so there is equality throughout in (30).

8.5   Distributions on $\mathbb{R}^+$.

The final task is to extend the results of the previous section from bounded measures to measures on $\mathbb{R}^+$. Observe that results for centred distributions with bounded support extend by translation to any pair of distributions with bounded support and the same mean.

Suppose that μ and ν have support on $\mathbb{R}^+$, and that both have mean c.

From the put-call parity relation we have $|Y - X| = 2(X - Y)^+ + (Y - X)$ and so

$$\hat{H}(\mu,\nu) := \sup_{(X,Y)\in\mathcal{M}} \mathbb{E}[(X - Y)^+]$$

satisfies $\hat{H}(\mu,\nu) = H(\mu,\nu)/2$.

For each n set $\tilde{X}^{(n)} = X \wedge n$ and $\tilde{Y}^{(n)} = Y \wedge \lambda(n)$ where $\lambda(n)$ is chosen so that $\mathbb{E}[\tilde{X}^{(n)}] = \mathbb{E}[\tilde{Y}^{(n)}]$. It can be shown that $\lambda(n) \geq n$. Then for any joint distribution of (X,Y) we have

$$\mathbb{E}(X - Y)^+ = \lim_{n\uparrow\infty} \mathbb{E}[(X - Y)^+ I_{\{X<n\}}] \tag{31}$$

and

$$\begin{aligned}
\mathbb{E}[(X - Y)^+ I_{\{X<n\}}] &= \mathbb{E}[(\tilde{X}^{(n)} - \tilde{Y}^{(n)})^+ I_{\{\tilde{X}^{(n)}<n\}}] \\
&\leq \mathbb{E}[(\tilde{X}^{(n)} - \tilde{Y}^{(n)})^+] \\
&\leq \hat{H}(\mu^{(n)}, \nu^{(n)}),
\end{aligned} \tag{32}$$

where I_A denotes the indicator function of the set A and $\mu^{(n)}$ and $\nu^{(n)}$ denote the laws of $\tilde{X}^{(n)}$ and $\tilde{Y}^{(n)}$. Both $\tilde{X}^{(n)}$ and $\tilde{Y}^{(n)}$ are bounded random variables and so by the results of the previous section,

$$\hat{H}(\mu^{(n)}, \nu^{(n)}) = \int_0^1 \frac{(r^{(n)}(u) - q^{(n)}(u)(q^{(n)}(u) - p^{(n)}(u))}{(r^{(n)}(u) - p^{(n)}(u))} du$$

for appropriate functions $0 \leq p^{(n)}(u) \leq q^{(n)}(u) \leq r^{(n)}(u)$. Then

$$\frac{(r^{(n)}(u) - q^{(n)}(u)(q^{(n)}(u) - p^{(n)}(u)))}{(r^{(n)}(u) - p^{(n)}(u))} \leq q^{(n)}(u) - p^{(n)}(u) \leq q^{(n)}(u) \leq q(u).$$

This last inequality follows by construction, since q and $q^{(n)}$ are inverse distribution functions of X and $\tilde{X}^{(n)}$ respectively.

Down a subsequence if necessary we have that $p^{(n)}$ and $r^{(n)}$ converge to p and r say, and then by dominated convergence

$$\hat{H}(\mu^{(n)}, \nu^{(n)}) \to \int_0^1 \frac{(r(u) - q(u)(q(u) - p(u))}{(r(u) - p(u))} du. \tag{33}$$

Combining (31), (32) and (33) we conclude that the right hand side of (33) is an upper bound for $\hat{H}(\mu, \nu)$. Moreover, by the same limiting arguments as before, p, q and r define a feasible construction of a random variable Y and hence

$$H(\mu, \nu) = 2 \int_0^1 \frac{(r(u) - q(u)(q(u) - p(u))}{(r(u) - p(u))} du$$

as required.

We have proved:

Theorem 19 *Suppose that μ and ν are probability measures on $\mathbb{R}^+$ each with mean c, and suppose that $\mu \preceq \nu$ in convex order. Let U be a uniform random variable on $[0, 1]$.*

There exist increasing functions p, q, r such that $X = q(U)$ and $Y \in \{p(U), r(U)\}$ with $\mathbb{E}[Y|U] = \mathbb{E}[X|U]$ satisfy $X \sim \mu$, $Y \sim \nu$ and $\mathbb{E}[Y|X] = X$, and the pair (X, Y) is such that $H(\mu, \nu) = \mathbb{E}[|Y - X|]$. Moreover,

$$H(\mu, \nu) = 2 \int_0^1 \frac{(r(u) - q(u))(q(u) - p(u))}{r(u) - p(u)} du.$$

If μ has no atoms then there exist increasing f and g such that $X \sim \mu$, $Y \in \{f(X), g(X)\}$ with $\mathbb{E}[Y|X] = X$ satisfy $Y \sim \nu$ and the pair (X, Y) is such that $H(\mu, \nu) = \mathbb{E}[|Y - X|]$ Moreover,

$$H(\mu, \nu) = 2 \int \frac{(g(x) - x)(x - f(x))}{g(x) - f(x)} \mu(dx).$$

Further, if f and g are strictly increasing, continuous and differentiable, then if β is given by (17)-(18), and then α is defined from β as in Section 5, then an alternative expression is

$$H(\mu, \nu) = \int \beta(y)\nu(dy) + \int \alpha(x)\mu(dx)$$

Proof: The first part of the theorem was proved at the beginning of this section, and the second part follows immediately from the identification $f \equiv p \circ q^{-1}$ and $g = r \circ q^{-1}$, note that the hypothesis is sufficient to ensure that q^{-1} is continuous and well defined. The final part follows from Lemma 17. $\qquad \square$

9 Numerical examples

In this section we present results from two numerical examples. In the first case we consider a pair of (continuous) uniform random variables and in the second case we consider a pair of normal random variables. The first step in each case is to approximate the initial and target random variables with discrete random variables. The problem of determining the joint law which maximises the expected value of $\mathbb{E}|Y - X|$ can then be reduced to a finite linear programme of the form (25)-(26). The results of these programmes are presented in Figures 4 and 5 in the form of the associated functions f and g. In particular, it is implicit in these figures that the linear programme has found a solution where either $Y = f(X)$ or $Y = g(X)$ for increasing functions f and g as required by the analysis of Section 4.
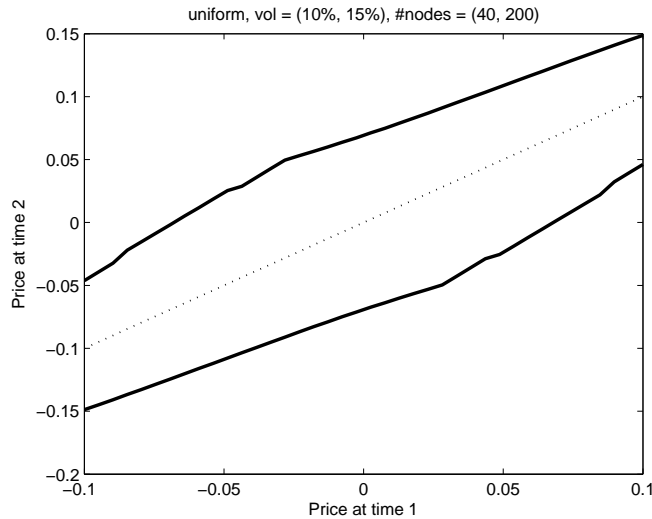

Figure 4: The functions $y = f(x)$ and $y = g(x)$ for a numerical example in which $X \sim U[-0.1, 0.1]$ and $Y \sim U[-0.15, 0.15]$. In fact both random variables are approximated by discrete uniform random variables on 40 and 200 points respectively. The linear programme finds an optimum which places mass at a 'cloud' of points on a grid; these points have been smoothed to improve the clarity of the figure.

10 Efficiency of the bound in the lognormal case

Suppose that X and Y have lognormal distributions. In particular suppose that μ and ν are the laws of

$$ce^{\sigma\sqrt{T_1}G_1 - \sigma^2 T_1/2} \quad \text{and} \quad ce^{\sigma\sqrt{T_2}G_2 - \sigma^2 T_2/2}$$

respectively, for a pair of standard Gaussian random variables G_1 and G_2.

A candidate martingale model for which the prices satisfy $F_{T_1} \sim \mu$ and $F_{T_2} \sim \nu$ is the Black-Scholes model

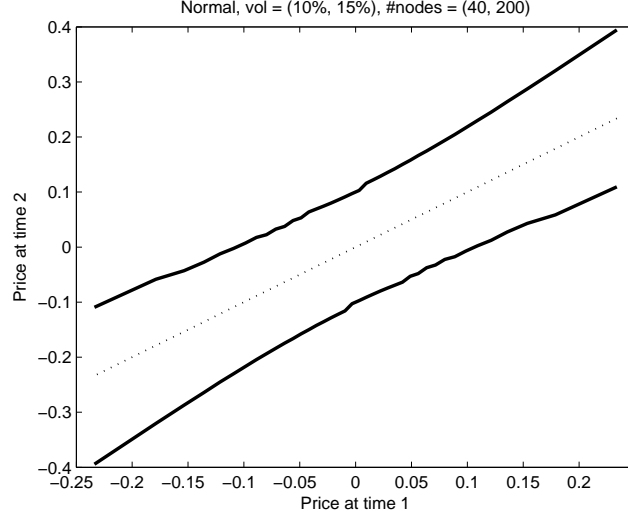$$dF_t = \sigma F_t dB_t \qquad F_0 = c.$$

23

Figure 5: The functions $y = f(x)$ and $y = g(x)$ for a numerical example in which $X \sim N(0, \sigma_X^2)$ and $Y \sim N(0, \sigma_Y^2)$ where $\sigma_X = 0.1$ and $\sigma_Y = 0.15$. Both random variables are approximated by discrete random variables such that X is approximated with a distribution consisting of 40 atoms, and Y is approximated by a random variable with 200 atoms.

Under that (complete market) model the price $E \equiv E(V)$ for the forward start straddle is given by

$$
\begin{aligned}
E(V) = \mathbb{E}[|F_{T_2} - F_{T_1}|] &= \mathbb{E}[F_{T_1}]\mathbb{E}[|F_{T_2}/F_{T_1} - 1|] \\
&= c\mathbb{E}[|e^{\sigma\sqrt{T_2 - T_1}G - \sigma^2(T_2 - T_1)/2} - 1|] \\
&= 2c\mathbb{P}(-\sqrt{V}/2 \le G \le \sqrt{V}/2)
\end{aligned}
$$

where $V = \sigma^2(T_2 - T_1)$ and G is standard Gaussian. When V is small this is approximately $E(V) \doteq c\sqrt{V}\sqrt{2/\pi}$.

Now consider the upper bound on the price of the option across all models which are consistent with the marginal distributions and the martingale property. We are going to use the entropy criterion to give a bound on $H(\mu, \nu)$. Since the family of lognormal distributions is closed under multiplication, we might hope that the entropy bound is moderately tight. It can be shown that for lognormal distributions it always outperforms the bound based on quadratic functionals and the Cauchy-Schwarz inequality.

We have

$$
\mathbb{E}[Y \ln Y - X \ln X] = \frac{cV}{2};
$$

and then by Corollaries 15 and 16

$$
H(\mu, \nu) \le c\tilde{J}(V/2) \le c\sqrt{V}.
$$

If we compare (the bound on) the model-free upper bound with the Black-Scholes

24

model-based price we find that the ratio of the prices satisfies

$$1 \leq \frac{H(\mu, \nu)}{E(V)} \leq \frac{\tilde{J}(V/2)}{4\Phi(\sqrt{V}/2) - 2}.$$

The model-based price, the entropy-based upper bound and the ratio of these quantities are plotted in Figure 6. Note that the smallest upper bound H is a function of $\sigma^2 T_1$ and $V \equiv \sigma^2(T_2 - T_1)$ whereas both the Black-Scholes model-based price and the entropy-based bound depend on V alone.
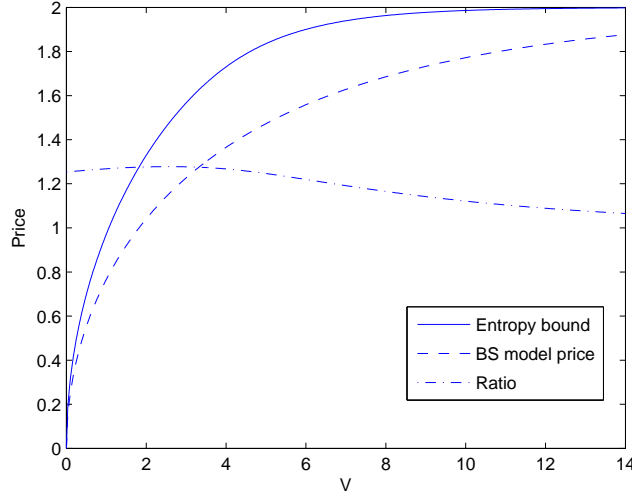

Figure 6: A plot of $E(V)$ (the Black-Scholes model-based price, dashed line) and $\tilde{J}(V/2)$ the entropy-based bound (solid line), scaled such that $c = 1$. The ratio between these prices is also shown; for small V the limit ratio is $\sqrt{\pi/2}$, for large V the limit is 1. We have $0 \leq E(V) \leq H(\mu, \nu) \leq \tilde{J}(V/2) \leq 2 \wedge \sqrt{V}$. It should be noted that most plausible parameter combinations are represented by low values of V, so the left-hand side of this figure is the most relevant.

Consider two agents (of different sexes) who wish to price a forward start straddle and who are in a market with vanilla call and put prices which are consistent with lognormal distributions for the asset price. The first agent assumes that prices follow a Black-Scholes model and charges $E(V)$. If he delta-hedges the straddle, and if the price realisation is consistent with the constant volatility model then he will hedge perfectly. The second agent makes no modelling assumptions. She charges a higher price for the straddle (but at worst 30% higher, and typically less) and uses the premium to purchase a portfolio of puts and calls and at T_1 makes an investment in the forward market. Under optimal portfolio choice, then whatever the realisation of the price process she will super-replicate.

The second agent charges more for the option, but not much more. Her hedging strategy is also much simpler (it is semi-static, which may be a significant advantage in a environment with transaction costs) and, most importantly, it is robust to model misspecification, and deviations from a Black-Scholes world.

References

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

[2] P. Billingsley. *Probability and Measure*. Wiley, New York, second edition, 1986.

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

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

[5] R.V. Chacon. Potential processes. *Trans. Amer. Math. Soc*, 226:39–58, 1977.

[6] A.M.G. Cox and J. Obłój. Model free pricing and hedging of double barrier options. 2008.

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

[8] H. Föllmer and A. Schied. *Stochastic Finance; An Introduction In Discrete Time*. de Gruyter, 2004.

[9] D. Gale, H.W. Kuhn, and A.W. Tucker. Linear programming and the theory of games. In T.C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 317–329. Wiley, New York, 1995.

[10] D. G. Hobson. The Skorokhod Embedding Problem and model-independent prices for options. 2009.

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

[12] D.G. Hobson and J.L. Pedersen. The minimum maximum of a continuous martingale with given initial and terminal laws. *Annals of Probability*, 30:978–999, 2002.

[13] S. D. Jacka. Optimal stopping and best constants for Doob-like inequalities. I. The case $p = 1$. *Ann. Probab.*, 19(4):1798–1821, 1991.

[14] J. Obłój. The Skorokhod embedding problem and its offspring. *Probability Surveys*, 1:321–392, 2004.

[15] 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.

[16] D. H. Root. The existence of certain stopping times on Brownian motion. *Ann. Math. Statist.*, 40:715–718, 1969.

[17] 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.

[18] P. Vallois. Quelques inégalités avec le temps local en zero du mouvement brownien. *Stochastic Process. Appl.*, 41(1):117–155, 1992.

[19] R.J. Vanderbei. *Linear Programming: Foundations and Extensions.* Operations Research and Management Science. Springer, 2008.