Option Pricing Bounds and Statistical Uncertainty [1]

by

Per A. Mykland

TECHNICAL REPORT NO. 511

Department of Statistics
The University of Chicago
Chicago, Illinois 60637

September 2001

This version: September, 2003

[1] This research was supported in part by National Science Foundation grants DMS 99-71738 and 02-04639.

Option Pricing Bounds and Statistical Uncertainty [1]

by

Per A. Mykland
The University of Chicago

Contents

[1] This research was supported in part by National Science Foundation grant DMS 99-71738 and 02-04639.

1. Introduction.

1.1. From pricing bounds to trading strategies. In the presence of statistical uncertainty, what bounds can one set on derivatives prices? Specifically, suppose we find ourselves at a time $t = 0$, with the following situation:

A Past: Information has been collected up to and including time $t = 0$. For the purpose of this paper, this is mainly historical statistical/econometric information (we use the terms interchangeably). It could also, however, include cross-sectional implied quantities. Or well informed subjective quantifications.

The Present: We wish to value a derivative security, or portfolio of securities, whose final payoff is η. This could be for a purchase or sale, or just to value a book. In addition to other valuations of this instrument, we would like a bound on its value. If the derivative is a liability, we need an upper bound, which we call A. If it is an asset, the relevant quantity is a lower bound, call it B. We wish to attach probability $1-\alpha$, say 95 %, to such a bound.

The standard approach of options theory is to base prices on trading strategies. If we adopt this paradigm, bounds would also be based on such strategies. We suppose there are underlying market traded securities $S_t^{(1)}, ..., S_t^{(p)}$, as well as a money market bond $\beta_t = \exp\{\int_0^t r_u du\}$, that can be made use of. This lead to a consideration of

The Future: Consider the case of the upper bound A. We consider lower bounds later. A trading based approach would be the following. A would be the smallest value for which there would exist a portfolio A_t, self financing in in the underlying securities, so that $A_0 = A$ and $A_T \geq \eta$ with probability at least $1-\alpha$. We shall see important examples in Sections 3, 6, and 7, and give precise mathematical meaning to these concepts in Sections 4, and 5.

The bound A is what it would cost to liquidate the liability η through delta hedging. It is also relevant if one intends to actually follow another trading strategy than A_t. This is discussed in Section 2.3.

Our approach, therefore, is to find A by finding a trading strategy. How to do the latter is the problem we are trying to solve.

The question of finding such a bound might also come up without any statistical uncertainty. In fact, one can usefully distinguish between two cases, as follows. We let P be the actual probability distribution of the underlying processes. Now distinguish between

(1) the "probabilistic problem": P is fixed and known, but there is incompleteness or other barriers to perfect hedging. Mostly, this means that the "risk neutral probability" P^* is unknown; and

(2) the "statistical problem": P is not known.

This article is about problem (2). We discuss problem (1) in Section 1.2 below. The risk neutral measure P^* is originally defined and discussed in Harrison and Kreps (1979), Harrison and Pliskà (1981), Delbaen and Schachermayer (1994, 1995).

The approach that we follow is mainly based on prediction sets, as outlined in the next section. Section 3 give the original examples of such sets. A more theoretical framework is laid in Sections 4-5. Section 6 considers interest rates, and Section 7 the effect of market traded options. The incorporation of econometric or statistical conclusions is discussed in Sections 5.3 and 8.

This article is based in large part on Avellaneda, Levy and Paras (1995), Lyons (1995), and Mykland (2000, 2003a,b).

1.2. Related problems and related literature. There is a wealth of problems related to the one considered in this article, and which, in the interest of conciseness, we do not cover here. The following is a quick road map to a number of research areas. The papers cited are just a small subset of the work that exists in these areas.

First of all, a substantial area of study has been concerned with the "*probabilistic*" problem (1) above. P is known, but due to some form of incompleteness or other barrier to perfect hedging, there are either several P^*s, or one has to find methods of pricing which do not involve a risk-neutral measure. The situation (1) can arise due to transaction cost, possible jumps in the price of the underlying security, differential cost of borrowing and lending, and so on. Strategies in such circumstances include *super-hedging* (Cvitanić and Karatzas (1992, 1993), Cvitanić, Pham and Touzi (1998, 1999), El Karoui and Quenez (1995), Eberlein and Jacod (1997), Karatzas (1996), Karatzas and Kou (1996, 1998), and Kramkov (1996)), *mean variance hedging* (Föllmer and Schweizer (1991), Föllmer and Sondermann (1986), Schweizer (1990, 1991, 1992, 1993, 1994), and later also Delbaen and Schachermayer (1996), Delbaen, Monat, Schachermayer, Schweizer and Stricker (1997), Laurent and Pham (1999), and Pham, Rheinländer, and Schweizer (1998)), and *quantile style hedging* (see, in particular, Külldorff (1993), Spivak and Cvitanić (1998), and Föllmer and Leukert (1999, 2000)).

It should be noted that the P known and P unknown cases *can* overlap in the case of Bayesian statistical inference. Thus, if P is a statistical posterior distribution, quantile hedging can accomplish similar aims to those of this article.

Closely connected to super-hedging (whether for P known or unknown) is the study of *robustness*. There is a fine line between the two. In the latter case, one does not try to optimize a starting value, but instead one takes a reasonable strategy and sees when it will cover the final liability. Papers focusing on the latter include Bergman, Grundy and Wiener (1996), El Karoui, Jeanblanc-Picqué and Shreve (1998), and Hobson (1998).

There are also several other methods for considering bounds that reflect the riskiness of a position. Important work includes Lo (1987), Bergman (1995), Artzner, Delbaen, Eber and Heath (1999), Cvitanić and Karatzas (1999), Constantinides and Zariphopoulou (1999, 2001), Friedman (2000), Fritelli (2000), and Föllmer and Schied (2002).

Finally, there is *Value at Risk*, as discussed in another article in this volume.

2. Options hedging from prediction sets: Basic description.

2.1. Setup, and super-self financing strategies. The situation is described in the introduction. We have collected data. On the basis of these, we are looking for trading strategies in $S_t^{(1)}, ..., S_t^{(p)}, \beta_t$, where $0 \leq t \leq T$, that will super-replicate the payoff with probability at least $1 - \alpha$.

We way we will mostly go about this is to use the data to set a prediction set C, and then to super-replicate the payoff on C. A prime instance would be to create such sets for volatilities, cross-volatilities, or interest rates. If we are dealing with a single continuous security S, with random and time varying volatility σ_t at time t, we could write

$$dS_t = m_t S_t dt + \sigma_t S_t dB_t, \tag{2.1}$$

where B is a Brownian motion. The set C could then get the form

- Extremes based bounds (Avellaneda, Levy and Paras (1995), Lyons (1995)):

$$\sigma_- \ \leq \ \sigma_t \ \leq \ \sigma_+ \tag{2.2}$$

- Integral based bounds (Mykland (2000, 2003a,b,c)):

$$\Xi^- \leq \int_0^T \sigma_t^2 dt \leq \Xi^+. \tag{2.3}$$

There is a wide variety of possible prediction sets, in particular when also involving the interest rate, cf. Section 6.

It will be convenient to separate the two parts of the concept of super-replication, as we see in the following.

As usual, we call X_t^* the discounted process X_t. In other words, $X_t^* = \beta_t^{-1} X_t$, and vice versa. In certain explicitly defined cases, discounting may be done differently, for example by a zero coupon bond (cf. Section 6 in this paper, and El Karoui, Jeanblanc-Picqué and Shreve (1998)).

A process $V_t, 0 \leq t \leq T$, representing a dynamic portfolio of the underlying securities, is said to be a *super-self financing portfolio* provided there are processes H_t and D_t, so that, for all t, $0 \leq t \leq T$,

$$V_t \ = \ H_t + D_t, \quad 0 \leq t \leq T, \tag{2.4}$$

where D_t^* is a non-increasing process, and where H_t is self financing in the traded securities $S_t^{(1)}, ..., S_t^{(p)}$. In other words, one may extract dividend from a super-self financing portfolio, but one cannot add funds.

"Self financing" means, by numeraire invariance (see, for example, Section 6.B of Duffie (1996)), that H_t^* can be represented as a stochastic integral with respect to the $S_t^{(i)*}$'s, subject to regularity conditions to eliminate doubling strategies. There is some variation in how to implement this (see, e.g., Duffie (1996), Chapter 6.C (p. 103-105)). In our case, a "hard" credit restriction is used in Section 5.1, and a softer constraint is used in Section 4.2.

On the other hand, V_t is a *sub-self financing portfolio* if it admits the representation (2.4) with D_t^* as nondecreasing instead.

For portfolio V to super-replicate η on the set C, we would then require

(i) V is a super-self financing strategy

and

(ii) solvency: $V_T \geq \eta$ on C

If one can attach a probability, say, $1-\alpha$, to the realization of C, then $1-\alpha$ is the *prediction probability*, and C is a $1-\alpha$ *prediction set*. The probability can be based on statistical methods, and be either frequentist or Bayesian.

DEFINITION. Specifically, C is a $1-\alpha$ *prediction set*, provided $P(C \mid \mathcal{H}) \geq 1-\alpha$, $P-a.s.$. Here either *(i)* $P(\ \cdot\ \mid \mathcal{H})$ is a Bayesian posterior given the data at time zero, or *(ii)* in the frequentist case, P describes a class of models, and $\mathcal{H}$ represents an appropriate subset of the information available at time zero (the values of securities and other financial quantities, and possibly ancillary material). α can be any number in $[0,1)$. ∎

The above is deliberately vague. This is for reasons that will become clear in Sections 5.3 and 8, where the matter is pursued further.

For example, a prediction set will normally be random. Given the information at time zero, however, C is fixed, and we treat it as such until Section 5.3. Also, note that if we extend "Bayesian probability" to cover general belief, our definition of a prediction set does not necessarily imply an underlying statistical procedure.

The problem we are trying to solve is as follows. We have to cover a liability η at a non random time T. Because of our comparative lack of knowledge about the relevant set of probabilities, a full super-replication (that works with probability one for all P) would be prohibitively expensive, or undesirable for other reasons. Instead, we require that we can cover the payoff η with, at least, the same (Bayesian or frequentist) probability $1-\alpha$. Given the above, if the set C has probability $1-\alpha$, then also $V_T \geq \eta$ with probability at least $1-\alpha$, and hence this is a solution to our problem.

TECHNICAL POINT. All processes, unless otherwise indicated, will be taken to be *càdlàg*, *i.e.*, right continuous with left limits. In Sections 1-3, we have ignored what probabilities are used when defining stochastic integrals, or even when writing statements like "$V_T \geq \eta$", which tend to only be "almost sure". Also, the set C is based on volatilities which are only defined relative to a probability measure. And there is no mention of filtrations. Discussion of these matters is deferred until Sections 4 and 5. ∎

2.2. The bounds A and B. Having defined super-replication for a prediction set, we would now like the cheapest such super replication. This defines A.

DEFINITION. The *conservative ask price* (or offer price) at time 0 for a payoff η to be made at a time T is

$$A \;=\; \inf\{V_0 :\; (V_t) \text{ is a super-replication on } C \text{ of the liability } \eta\}. \tag{2.5}$$

∎

The definition is in analogy to that used by Cvitanić and Karatzas (1992, 1993), El Karoui and Quenez (1995), and Kramkov (1996). It is straightforward to see that A is a version of "value at risk" (see Chapter 14 (pp. 342-365) of Hull (1999)) that is based on dynamic trading. At the same time, A is coherent in the sense of Artzner, Delbaen, Eber and Heath (1999).

It would normally be the case that there is a super-replication A_t so that $A_0 = A$, and we argue this in Section 4.1. Note that in the following, V_t denotes the portfolio value of any super-replication, while A_t is the cheapest one, provided it exists.

Similarly, the *conservative bid price* can be defined as the supremum over all sub-replications of the payoff, in the obvious sense. For payoff η, one would get

$$B(\eta) \;=\; -A(-\eta), \tag{2.6}$$

in obvious notation, and subject to mathematical regularity conditions, it is enough to study ask prices. More generally, if one already has a portfolio of options, one may wish to charge $A(\text{portfolio } + \eta) - A(\text{portfolio})$ for the payoff η.

But is A the starting value of a trading strategy? And how does one find A?

Suppose that $\mathcal{P}^*$ is the set of all risk neutral probabilities that allocate probability one to the set C. And suppose that $\mathcal{P}^*$ is nonempty. If we set

$$\eta^* = \beta_T^{-1}\eta, \tag{2.7}$$

and if $P^* \in \mathcal{P}^*$, then $E^*(\eta^*)$ is a possible price that is consistent with the prediction set C. Hence a lower bound for A is

$$A' = \sup_{P^* \in \mathcal{P}^*} E^*(\eta^*). \tag{2.8}$$

It will turn out that in many cases, $A = A'$. But A' is also useful in a more primitive way. Suppose one can construct a super-replication V_t on C so that $V_0 \leq A'$. Then V_t can be taken as our super-replication A_t, and $A = V_0 = A'$.

We shall see two cases of this in Section 3.

2.3. The practical rôle of prediction set trading. How does one use this form of trading? If the prediction probability $1 - \alpha$ is set too high, the starting value may be too high given the market price of contingent claims.

There are, however, at least three other ways of using this technology. First of all, it is not necessarily the case that α need to be set all that small. A reasonable way of setting hedges might be to use a 60% or 70% prediction set, and then implement the resulting strategy. It should also be emphasized that an economic agent can use this approach without necessarily violating market equilibrium, *cf.* Heath and Ku (2001).

On the other hand, one can analyze a possible transaction by finding out what is the smallest α for which a conservative strategy exists with the proposed price as starting value. If this α is too small, the transaction might be better avoided.

A main way of using conservative trading, however, is as a backup device for other strategies, and this is what we shall discuss in the following.

We suppose that a financial institution sells a payoff η (to occur at time T), and that a trading strategy is established on the basis of whatever models, data, or other considerations that the trader or the institution wishes to make. We shall call this the "preferred" strategy, and refer to its current value as V_t.

On the other hand, we also suppose that we have established a conservative strategy, with current value A_t, where the relevant prediction interval has probability $1 - \alpha$. We also assume that a reserve is put in place in the amount of K units of account, where

$$K \ > \ A_0 - V_0 \ .$$

The overall strategy is then as follows. One uses the preferred strategy unless or until it eats up the excess reserve over the conservative one. If or when that happens, one switches to the conservative strategy. In other words, one uses the preferred strategy until

$$\tau \ = \ \inf\{ \ t \ : \ K = A_t^* - V_t^* \ \} \ \wedge \ T$$

where the superscript "$*$" refers, as before, to discounting with respect to whatever security the reserve is invested in. This will normally be a money market account or the discount bond Λ_t. The symbol $a \wedge b$ means $\min(a, b)$.

This trading strategy has the following desirable properties:

- If the prediction set is realized, the net present value of the maximum loss is

$$V_0 \ + \ K \ - \ \ \text{actual sales price of the contingent claim} \ .$$

- If the reserves allocated to the position are used up, continuing a different sort of hedge would often be an attractive alternative to liquidating the book.
- The trader or the institution does not normally have to use conservative strategies. Any strategy can be used, and the conservative strategy is just a backup.

The latter is particularly important because it does not require any interference with any institution's or trader's standard practice unless the reserve is used up. The trader can use what she (or Risk Management) thinks of as an appropriate model, and can even take a certain amount of directional bets. Until time τ.

The question of how to set the reserve K remains. From a regulatory point of view, it does not matter how this is done, and is more a reflection of the risk preferences of the trader or the institution. There will normally be a trade-off in that expected return goes up with reserve level K. To determine an appropriate reserve level one would have to look at the actual hedging strategy used. For market traded or otherwise liquid options one common strategy is to use implied volatility

(Beckers (1981), Bick and Reisman (1993)), and the level K can then be evaluated by empirical data. If a strategy is based on theoretical considerations, one can evaluate the distribution of the return for given K based on such a model.

3. Options hedging from prediction sets: The original cases. One should, perhaps, start here, rather than with the theoretical considerations in Section 2. Suppose that a stock follows (2.1) and pays no dividends, and that there is a risk free interest rate r_t. *Both σ_t and r_t can be stochastic and time varying.* We put ourselves in the context of European options, with payoff $f(S_T)$. For future comparison, note that when r and σ are constant, the Black-Scholes(1973)-Merton(1973) price of this option is $B(S_0, rT, \sigma\sqrt{T})$, where

$$B(S, \Xi, R) = \exp(-R)Ef(S\exp(R - \Xi/2 + \sqrt{\Xi}Z)), \tag{3.1}$$

and where Z is standard normal (see, for example, Ch. 6 of Duffie (1996)). In particular, for the call payoff $f(s) = (s - K)^+$,

$$B(S, R, \Xi) \;=\; S\Phi(d_1) - K\exp(-R)\Phi(d_2), \tag{3.2}$$

where

$$d_1 \;=\; (\log(S/K) + R + \Xi/2)\,/\sqrt{\Xi} \tag{3.3}$$

and $d_2 \;=\; d_1 - \sqrt{\Xi}$. This will have some importance in the future discussion.

3.1. Point-wise bounds. This goes back to Avellaneda, Levy, and Paras (1995) and Lyons (1995). See also Frey and Sin (1999) and Frey (2000). In the simplest form, one lets C be the set for which

$$\sigma_t \epsilon [\sigma_-, \sigma_+] \text{ for all } t \epsilon [0, T], \tag{3.4}$$

and we let r_t be non-random, but possibly time varying. More generally, one can consider bounds on the form

$$\sigma_-(S_t, t) \leq \sigma_t \leq \sigma_+(S_t, t) \tag{3.5}$$

A super-replicating strategy can now be constructed for European options based on the "Black-Scholes-Barenblatt" equation (cf. Barenblatt (1978)). The price process $V(S_t, t)$ is found by using the Black-Scholes partial differential equation, but the term containing the volatility takes on either the upper or lower limit in (3.5), depending on the sign of the second derivative $V_{SS}(s,t)$. In other words, V solves the equation

$$r(V - V_S S) = \tfrac{\partial V}{\partial t} + \tfrac{1}{2} S^2 \max_{(3.5)}(\sigma_t^2 V_{SS}), \tag{3.6}$$

with the usual boundary condition $V(S_T, T) = f(S_T)$.

The *rationale* for this is the following. By Itô's Lemma, and assuming that the actual realized σ_t satisfies (3.5), dV_t becomes:

$$\begin{aligned} dV(S_t,t) &= V_S dS_t + \tfrac{\partial V}{\partial t} dt + \frac{1}{2} V_{SS} S_t^2 \sigma_t^2 dt \\ &\leq V_S dS_t + \tfrac{\partial V}{\partial t} dt + \tfrac{1}{2} S^2 \max_{(3.5)}(\sigma_t^2 V_{SS}) dt \\ &= V_S dS_t + (V - V_S S_t)\beta_t^{-1} d\beta_t, \end{aligned} \tag{3.7}$$

in view of (3.6). Hence $V_t = V(S_t, t)$ is the value of a super-self financing portfolio, and it covers the option liability by the boundary condition.

To see the relationship to (2.4), note that the process D_t has the form

$$D_t^* = -\tfrac{1}{2} \int_0^t S_u^2 \left(\max_{(3.5)}(\sigma_t^2 V_{SS}) - \sigma_t^2 V_{SS} \right) du. \tag{3.8}$$

This is easily seen by considering (3.6)-(3.7) on the discounted scale.

The reason why V_0 can be taken to be A, is that the stated upper bound coincides with the price for one specific realization of σ_t that is inside the prediction region. Hence, also, V_t can be taken to be A_t.

Pointwise bounds have also been considered by Bergman, Grundy and Wiener (1996), El Karoui, Jeanblanc-Picqué and Shreve (1998), and Hobson (1998), but these papers have concentrated more on robustness than on finding the lowest price A.

3.2. Integral bounds. This goes back to Mykland (2000), and for the moment, we only consider convex payoffs f (as in puts and calls). The interest rate can be taken to be random, in which case f must also be increasing (as in calls). More general formulae are given in Sections 6.3-6.4. The prediction set C has the form

$$R_0 \geq \int_0^T r_u du \text{ and } \Xi_0 \geq \int_0^T \sigma_u^2 du. \tag{3.9}$$

Following Section 2.2, we show that $A = B(S_0, R_0, \Xi_0)$ and that a super-replication A_t exists.

Consider the instrument whose value at time t is

$$V_t \;=\; B(S_t, R_t, \Xi_t), \tag{3.10}$$

where

$$R_t = R_0 - \int_0^t r_u du \text{ and } \Xi_t = \Xi_0 - \int_0^t \sigma_u^2 du. \tag{3.11}$$

In equation (3.11), r_t and σ_t are the actual observed quantities. As mentioned above, they can be stochastic and random.

Our claim is that V_t is exactly self financing. Note that, from differentiating (3.1),

$$\frac{1}{2} B_{SS} S^2 = B_\Xi \text{ and } -B_R = B - B_S S. \tag{3.12}$$

Also, for calls and puts, the first of the two equations in (3.12) is the well known relationship between the "gamma" and the "vega" (cf., for example, Chapter 14 of Hull (1997)).

Hence, by Itô's Lemma, dV_t equals:

$$\begin{aligned} dB(S_t, \Xi_t, R_t) &= B_S dS_t + \frac{1}{2} B_{SS} S_t^2 \sigma_t^2 dt + B_\Xi d\Xi_t + B_R dR_t \\ &= B_S dS_t + (B - B_S S_t) r_t dt \\ &\quad + [\frac{1}{2} B_{SS} S_t^2 - B_\Xi] \sigma_t^2 dt \\ &\quad + [B - B_S S_t - B_R] r_t dt. \end{aligned} \tag{3.13}$$

In view of (3.12), the last two lines of (3.13) vanish, and hence there is a self financing hedging strategy for V_t in S_t and β_t. The "delta" (the number of stocks held) is $B_S'(S_t, R_t, \Xi_t)$.

Furthermore, since $B(S, \Xi, R)$ is increasing in Ξ and R, (3.9) yields that

$$\begin{aligned} V_T &= B(S_T, \Xi_T, R_T) \\ &\geq \lim_{\Xi \downarrow 0, R \downarrow 0} B(S_T, \Xi, R) \\ &= f(S_T) \end{aligned} \tag{3.14}$$

almost surely. In other words, one can both synthetically create the security V_t, and one can use this security to cover one's obligations. Note that if r_t is nonrandom (but can be time varying), there is no limit in R in (3.14), and so f does not need to be increasing.

The reason why V_0 can be taken to be A is the same as in section 3.1. Also, the stated upper bound coincides with the Black-Scholes (1973)-Merton (1973) price for constant coefficients $r = R_0/T$ and $\sigma^2 = \Xi_0/T$. This is one possible realization satisfying the constraint (3.9). Also, V_t can be taken to be A_t.

3.3. Comparison of approaches. The main feature of the two approaches described above is how similar they are. Apart from having all the features from Section 2, they also have in common that they work "independently of probability". This, of course, is not quite true, since stochastic integrals require the usual probabilistic setup with filtrations and distributions. It does mean, however, that one can think of the set of possible probabilities as being exceedingly large. A stab at an implementation of this is given in Section 5.1.

And then we should discuss the differences. To start on a one sided note, consider first the results in Table 1 for convex European payoffs.

TABLE 1

Comparative prediction sets for convex European options: r constant

device	prediction set	A_0 at time 0	delta at time t
Black-Scholes:	σ constant	$B(S_0, rT, \sigma^2 T)$	$\frac{\partial B}{\partial S}(S_t, r(T-t), \sigma^2(T-t))$
average based:	$\Xi^- \leq \int_0^T \sigma_u^2 du \leq \Xi^+$	$B(S_0, rT, \Xi^+)$	$\frac{\partial B}{\partial S}(S_t, r(T-t), \Xi^+ - \int_0^t \sigma_u^2 du)$
extremes based:	$\sigma_- \leq \sigma_t \leq \sigma_+$	$B(S_0, rT, (\sigma^+)^2 T)$	$\frac{\partial B}{\partial S}(S_t, r(T-t), \sigma_+^2(T-t))$

B is defined in (3.2)-(3.3) for call options, and more generally in (3.1). A_0 is the conservative price (2.5). Delta is the hedge ratio (the number of stocks held at time t to super-hedge the option).

To compare these approaches, note that the function $B(S, R, \Xi)$ is increasing in its last argument. It will therefore be the case that the ordering in Table 1 places the lowest value of A_0 at the top and the highest at the bottom. This is since $\sigma^2 T \leq \Xi^+ \leq \sigma_+^2 T$. The latter inequality stems from the fact that Ξ^+ is a prediction bound for an integral of a process, while σ_+^2 is the corresponding bound for the maximum of the same process.

But Table 1 is not the full story. The average based interval is clearly better than the extremes based one in that it provides a lower starting value A_0. This may not, however, be the case for options that are not of European type. For example, *caplets* (see Hull (1999), p. 538) on volatility would appear be better handled through extremes based intervals, though we have not investigated this issue. The problem is, perhaps, best understood in the interest rate context, when comparing caplets with European options on swaps ("swaptions", see Hull (1999), p. 543). See Carr, Geman and Madan (2001) and Heath and Ku (2001) for a discussion in terms of coherent measures of risk. To see the connection, note that the average based procedure, with starting value $A_0 = B(S_0, rT, \Xi^+)$, delivers an actual payoff $A_T = B(S_T, , 0, \Xi^+ - \int_0^T \sigma_u^2 du)$. Hence A_T not only dominates the required payoff $f(S_T)$ on C, but the actual A_T is a combination of option on the security S and swaption on the volatility, in both cases European.

Another issue when comparing the two approaches is how one sets the hedge in each case. In Section 3.2, one uses the actual σ_t (for the underlying security) to set the hedge. In Section 3.1, on the other hand, the hedge itself is based on the worst case non-observed volatility. In both cases, of course, the price is based on the worst case scenario.

There are advantages to both schemes. For the case of Section 3.2, the hedge ratio (delta) at time t for the average based set (2.3) is not strictly speaking observable, but only approximable to a high degree of accuracy. It is natural to approximate the integral of σ_t^2 by the observed quadratic variation of $\log S$.

Specifically, suppose at time t that one has recorded $\log S_{t_i}$ for $0 = t_0 < \ldots < t_k \leq t$. The observed quadratic variation is then

$$\hat{\Xi}_t = \sum_{i=1}^{k} (\log S_{t_i} - \log S_{t_{i-1}})^2 \tag{3.15}$$

This quantity converges in probability to $\int_0^t \sigma_u^2 du$, *cf.* Theorem I.4.47 (p.52) of Jacod and Shiryaev (1987). Note that the limit of (3.15) is often taken as the definition of the integrated volatility. and is then denoted by $[\log S, \log S]_t$. This is also called the (theoretical) quadratic variation of $\log S$. More generally, the quadratic covariation between processes X and Y is given by

$$[X, Y]_t = \text{limit in probability of } \sum_{i=1}^{k} (X_{t_i} - X_{t_{i-1}})(Y_{t_i} - Y_{t_{i-1}}) \tag{3.16}$$

as $\Delta t \to 0$.

The natural hedge ratio at time t for the average based procedure would therefore be

$$\tfrac{\partial B}{\partial S}(S_t, r(T-t), \Xi^+ - \hat{\Xi}_t). \tag{3.17}$$

If Δt is the average distance t/k, standard stochastic process results yield that, subject to regularity conditions, $\hat{\Xi}_t - \int_0^t \sigma_u^2 du = O_p(\Delta t^{1/2})$, see, for example, Jacod and Protter (1998), Zhang (2001) and Mykland and Zhang (2002b). This would also be the order of the hedging error relative to using the delta given in Table 1. How to adjust the prediction interval accordingly, remains to be investigated.

The method in Section 3.1 does not have any of these problems. Using unobserved worst case volatilities, however, can only be done with quite specific prediction sets. It is hard to see, for example, how to do this for the bounds in Section 3.2. Incidentally, as a corollary to the difference in hedging schemes, the approach in Section 2.1 will normally yield dividends *en route* (this is the D from (2.4)), while the only dividends in Section 3.2 will he paid at time of expiration.

3.4. An implementation with data. We here demonstrate by example how one can take data, create a prediction set, and then feed this into the hedging schemes above. We use the band from Section 3.2, and the data analysis of Jacquier, Polson and Rossi (1994), which analyses (among other series) the S&P 500 data recorded daily. The authors consider a stochastic volatility model that is linear on the log scale:

$$d\log(\sigma_t^2) = (a + b\log(\sigma_t^2))dt + cdW_t,$$

a.k.a., by exact discretization,

$$\log(\sigma_{t+1}^2) = (\alpha + \beta\log(\sigma_t^2)) + \gamma\epsilon_t,$$

where W is a standard Brownian motion and the ϵs are consequently i.i.d. standard normal. We shall in the following suppose that the effects of interest rate uncertainty are negligible. With some assumptions, their posterior distribution, as well as our corresponding options price, are given in Table 2. We follow the custom of stating the volatility *per annum* and on a square root scale.

TABLE 2

S&P 500: Posterior distribution of $\Xi = \int_0^T \sigma_t^2 dt$ for $T =$ one year

Conservative price A_0 corresponding to relevant coverage for at the money call option

posterior coverage	50%	80%	90%	95%	99%
upper end $\sqrt{\Xi}$ of posterior interval	.168	.187	.202	.217	.257
conservative price A_0	9.19	9.90	10.46	11.03	12.54

Posterior is conditional on $\log(\sigma_0^2)$ taking the value of the long run mean of $\log(\sigma^2)$. A_0 is based on prediction set (2.3) with $\Xi^- = 0$. A 5 % *p.a.* known interest rate is assumed. $S_0 = 100$.

In the above, we are bypassing the issue of conditioning on σ_0. Our excuse for this is that σ_0 appears to be approximately observable in the presence of high frequency data. Following Foster and Nelson (1996), Zhang (2001), and Mykland and Zhang (2002a), the error in observation is of the order $O_p(\Delta t^{1/4})$, where Δt is the average distance between observations. What modification has to be made to the prediction set in view of this error remains to be investigated. It may also be that it would be better to condition on some other quantity than σ_0.

The above does not consider the possibility of also hedging in market traded options, *cf.* Section 7.

4. Properties of trading strategies.

4.1. Super-self financing and supermartingale. The analysis in the preceding sections has been heuristic. In order to more easily derive results, it is useful to set up a somewhat more theoretical framework. In particular, we are missing a characterization of what probabilities can be applicable, both for the trading strategies, and for the candidate upper bound (2.8).

The discussion in this section will be somewhat more general than what is required for pure prediction sets. We also make use of this development in Section 7 on interpolation, and in Section 8 on (frequentist) confidence and (Bayesian) credible sets. Sharper results, that pertain directly to the pure prediction set problem, will be given in Section 5.

We consider a filtered space $(\Omega, \mathcal{F}, \mathcal{F}_t)_{0 \leq t \leq T}$. The processes $S_t^{(1)}, ..., S_t^{(p)}$, r_t and $\beta_t = \exp\{\int_0^t r_u du\}$ are taken to be adapted to this filtration. The $S^{(i)}$'s are taken to be continuous, though similar theory can most likely be developed in more general cases.

$\mathcal{P}$ is a set of probability distributions on $(\Omega, \mathcal{F})$.

Definition. A property will be said to hold $\mathcal{P} - a.s.$ if it holds $P - a.s.$ for all $P \in \mathcal{P}$.

"Super-self financing" now means that the decomposition (2.4) must be valid for all $P \in \mathcal{P}$, but note that H and D may depend on P. The stochastic integral is defined with respect to each P, *cf.* Section 4.2. ∎

To give the general form of the ask price A, we consider an appropriate set $\mathcal{P}^*$ of "risk neutral" probability distributions P^*.

Definition. Set

$$\mathcal{N} = \{C \subseteq \Omega : \forall P \in \mathcal{P} \; \exists E \epsilon \mathcal{F} : \; C \subseteq E \text{ and } P(E) = 0\}. \tag{4.1}$$

$\mathcal{P}^*$ is now defined as the set of probability measures P^* on $\mathcal{F}$ whose null sets include those in $\mathcal{N}$, and for which $S_t^{(1)*}, ..., S_t^{(p)*}$ are martingales. We also define $\mathcal{P}^e$ as the set of extremal elements in $\mathcal{P}^*$. P^e is extremal in $\mathcal{P}^*$ if $P^e \in \mathcal{P}^*$ and if, whenever $P^e = a_1 P_1^e + a_2 P_2^e$ for $a_1,\ a_2 > 0$ and $P_1^e,\ P_2^e \in \mathcal{P}^*$, it must be the case that $P^e = P_1^e = P_2^e$. ∎

Subject to regularity conditions, we shall show that there is a super-replicating strategy A_t with initial value A from (2.5).

First, however, a more basic result, which is useful for understanding super-self financing strategies.

THEOREM 4.1. *Subject to the regularity conditions stated below, (V_t) is a super-self financing strategy if and only if (V_t^*) is a càdlàg supermartingale for all $P^* \in \mathcal{P}^*$.*

The same, obviously, applies to the relationship between sub-self financing strategies and submartingales. We return to the regularity conditions below, but will for the moment focus on the impact of this result. Note that the minimum of two, or even a countable number, of supermartingales, remains a supermartingale. By Theorem 4.1, the same must then apply to super-self financing strategies.

COROLLARY 4.2. *Subject to the regularity conditions stated below, suppose that there exists a super-replication of η on Ω (the entire space). Then there is a super-replication A_t so that $A_0 = A$.*

The latter result will be important even when dealing with prediction sets, as we shall see in Section 5.

TECHNICAL CONDITIONS. The assumptions required for Theorem 4.1 and Corollary 4.2 are as follows. The system: $(\mathcal{F}_t)$ is right continuous; $\mathcal{F}_0$ is the smallest σ-field containing $\mathcal{N}$; the $S_t^{(i)}$ are $\mathcal{P} - a.s.$ continuous and adapted; the short rate process r_t is adapted, and integrable $\mathcal{P} - a.s.$; every $P \in \mathcal{P}$ has an equivalent martingale measure, that is to say that there is a $P^* \in \mathcal{P}^*$ that is equivalent to P. Define the following conditions. (E_1): "if X is a bounded random variable and there is a $P^* \in \mathcal{P}^*$ so that $E^*(X) > 0$, then there is a $P^e \in \mathcal{P}^e$ so that $E^e(X) > 0$". (E_2): "there is a real number K so that $\{V_T^* \geq -K\}^c \in \mathcal{N}$".

Theorem 4.1 now holds supposing that (V_t) is an adapted process, and assuming either

- condition (E_1) and that the terminal value of the process satisfies:

$$\sup_{P^*\epsilon\mathcal{P}^*} E^* V_T^{*-} \ < \ \infty; \text{ or}$$

- condition (E_2); or
- that (V_t) is continuous.

Corollary 4.2 holds under the same system assumptions, and provided either (E_1) and $\sup_{P^*\epsilon\mathcal{P}^*} E^*|\eta^*| \ < \ \infty$, or provided $\eta^* \geq -K \ \mathcal{P} - a.s.$ for some K. ∎

Note that under condition (E$_2$), Theorem 4.1 is a corollary to Theorem 2.1 (p. 461) of Kramkov (1996). This is because $\mathcal{P}^*$ includes the union of the equivalent martingale measures of the elements in $\mathcal{P}$. For reasons of symmetry, however, we have also sought to study the case where η^* is not bounded below, whence the condition (E$_1$). The need for symmetry arises from the desire to also study bid prices, cf. (2.6). For example, neither a short call not a short put are bounded below. See Section 4.2.

A requirement in the above results that does need some comment is the one involving extremal probabilities. Condition (E$_1$) is actually quite weak, as it is satisfied when $\mathcal{P}^*$ is the convex hull of its extremal points. Sufficient conditions for a result of this type are given in Theorems 15.2, 15.3 and 15.12 (p. 496-498) in Jacod (1979). For example, the first of these results gives the following as a special case (see Section 6). This will cover our examples.

PROPOSITION 4.3. *Assume the conditions of Theorem 4.1. Suppose that r_t is bounded below by a nonrandom constant (greater that $-\infty$). Suppose that $(\mathcal{F}_t)$ is the smallest right continuous filtration for which $(\beta_t, S_t^{(1)}, ..., S_t^{(p)})$ is adapted and so that $\mathcal{N} \subseteq \mathcal{F}_0$. Let $C \in \mathcal{F}_T$. Suppose that $\mathcal{P}^*$ equals the set of all probabilities P^* so that $(S_t^{(1)*}), ..., (S_t^{(p)*})$ are P^*-martingales, and so that $P^*(C) = 1$. Then Condition (E_1) is satisfied.*

EXAMPLE . To see how the above works, consider systems with only one stock ($p = 1$). We let (β_t, S_t) generate $(\mathcal{F}_t)$. A set $C \in \mathcal{F}_T$ will describe our restrictions. For example C can be the set given by (2.2) or (2.3). The fact that σ_t is only defined given a probability distribution is not a difficulty here: we consider Ps so that the set C has probability 1 (where quantities like σ_t are defined under P).

One can also work with other types of restrictions. For example, C can be the set of probabilities so that (3.9) is satisfied, and also $\Pi^- \leq [r,\sigma]_T \leq \Pi^+$, where the covariation "$[,]$" is defined in (3.16) in Section 3.3. Only the imagination is the limit here.

Hence, $\mathcal{P}$ is the set of all probability distributions P so that $S_0 = s_0$ (the actual value),

$$dS_t \;=\; \mu_t S_t dt + \sigma_t S_t dW_t, \tag{4.2}$$

with r_t integrable $P - a.s.$, and bounded below by a nonrandom constant, so that P(C) = 1, and so that

$$\exp\left\{-\int_0^t \lambda_u dW_u - \frac{1}{2}\int_0^t \lambda_u^2 du\right\} \qquad \text{is a } P\text{-martingale,} \tag{4.3}$$

where $\lambda_u = (\mu_u - r_u)/\sigma_u$. The condition (4.3) is what one needs for Girsanov's Theorem (see, for example, Karatzas and Shreve (1991), Theorem 3.5.1) to hold, which is what assures the required existence of equivalent martingale measure. Hence, in view of Proposition 4.3, Condition (E_1) is taken care of.

To gain more flexibility, one can let $(\mathcal{F}_t)$ be generated by more than one stock, and just let these stocks remain "anonymous". One can then still use condition (E_1). Alternatively, if the payoff is bounded below, one can use condition (E_2). ∎

4.2. Defining self-financing strategies. In essence, H_t being self-financing means that we can represent H_t^* by

$$H_t^* = H_0^* + \sum_{i=1}^{p} \int_0^t \theta_s^{(i)} dS_s^{(i)^*}. \tag{4.4}$$

This is in view of numeraire invariance (see, *e.g.*, Section 6.B of Duffie (1996)).

Fix $P \in \mathcal{P}$, and recall that the $S_t^{(i)*}$ are continuous. We shall take the stochastic integral to be defined when $\theta_t^{(1)}, \ldots, \theta_t^{(p)}$ is an element in $L^2_{\text{loc}}(P)$, which is the set of p-dimensional predictable processes so that $\int_0^t \theta_u^{(i)^2} d[S^{(i)^*}, S^{(i)^*}]_u$ is locally integrable P-a.s. The stochastic integral (4.4) is then defined by the process in Theorems I.4.31 and I.4.40 (p. 46–48) in Jacod and Shiryaev (1987).

A restriction is needed to be able to rule out doubling strategies. The two most popular ways of doing that are to insist either that H_t^* be in an L^2-space, or that it be bounded below (Harrison and Kreps (1979), Delbaen and Schachermayer (1995), Dybvig and Huang (1988), Karatzas (1996); see also Duffie (1996), Section 6.C). We shall here go with a criterion that encompasses both.

DEFINITION. A process H_t, $0 \leq t \leq T$, is *self-financing* with respect to $S_t^{(1)}, \ldots, S_t^{(p)}$ if H_t^* satisfies (4.4), and if $\{H_\lambda^{*-},\ 0 \leq \lambda \leq T,\ \lambda \text{ stopping time}\}$ is uniformly integrable under all $P^* \in \mathcal{P}^*$ that are equivalent to P. ∎

The reason for seeking to avoid the requirement that H_t^* be bounded below is that, to the extent possible, the same theory should apply equally to bid and ask prices. Since the bid price is normally given by (2.6), securities that are unbounded below will be a common phenomenon. For example, $B((S-K)^+) = -A(-(S-K)^+)$, and $-(S-K)^+$ is unbounded below.

It should be emphasized that our definition does, indeed, preclude doubling type strategies. The following is a direct consequence of optional stopping and Fatou's Lemma.

PROPOSITION 4.4. *Let $P \in \mathcal{P}$, and suppose that there is at least one $P^* \in \mathcal{P}^*$ that is equivalent to P. Suppose that H_t^* is self financing in the sense given above. Then, if there are stopping times λ and μ, $0 \leq \lambda \leq \mu \leq T$, so that $H_\mu^* \geq H_\lambda^*$, P-a.s., then $H_\mu^* = H_\lambda^*$, P-a.s.*

Note that Proposition 4.4 is, in a sense, an equivalence. If the conclusion holds for all H_t^*, it must in particular hold for those that Delbaen and Schachermayer (1995) term admissible. Hence, by Theorem 1.4 (p. 929) of their work, P^* exists.

4.3. Proofs for Section 4.1.

Proof of Theorem 4.1. The "only if" part of the result is obvious, so it remains to show the "if" part.

(a) Structure of the Doob-Meyer decomposition of (V_t^*). Fix $P^* \in \mathcal{P}^*$. Let

$$V_t^* = H_t^* + D_t^* \ , \quad D_0 = 0 \tag{4.5}$$

be the Doob-Meyer decomposition of V_t^* under this distribution. The decomposition is valid by, for example, Theorem 8.22 (p. 83) in Elliot (1982). Then $\{H_\lambda^{*-},\ 0 \leq \lambda \leq T,\ \lambda \text{ stopping time}\}$ is uniformly integrable under P^*. This is because $H_t^{*-} \leq V_t^{*-} \leq E^*(|\eta^*| \mid \bar{\mathcal{F}}_t)$, the latter inequality because $V_t^{*-} = (-V_t^*)^+$, which is a submartingale since V_t^* is a supermartingale. Hence uniform integrability follows by, say, Theorem I.1.42(b) (p. 11) of Jacod and Shiryaev (1987).

(b) Under condition (E_1), (V_t) *can be written* $V_t^* = V_t^{*c} + V_t^{*d}$, *where* (V_t^{*c}) *is a continuous supermartingale for all* $P^* \in \mathcal{P}^*$, *and* (V_t^{*d}) *is a nonincreasing process.* Consider the set C of $\omega \in \Omega$ so that $\Delta V_t^* \leq 0$ for all t, and so that $V_t^{*d} = \sum_{s \leq t} \Delta V_s^*$ is well defined. We want to show that the complement $C^c \in \mathcal{N}$. To this end, invoke Condition (E_1), which means that we only have to prove that $P^e(C) = 1$ for all $P^e \in \mathcal{P}^e$.

Fix, therefore, $P^e \in \mathcal{P}^e$, and let H_t^* and D_t^* be given by the Doob-Meyer decomposition (4.5) under this distribution. By Proposition 11.14 (p 345) in Jacod (1979), P^e is extremal in the set $M(\{S^{(1)*}, ..., S^{(p)*}\})$ (in Jacod's notation), and so it follows from Theorem 11.2 (p. 338) in the same work, that (H_t^*) can be represented as a stochastic integral over the $(S_t^{(i)*})$'s, whence (H_t^*) is continuous. $P^e(C) = 1$ follows.

To see that (V_t^{*c}) is a supermartingale for any given $P^* \in \mathcal{P}^*$, note that Condition (E_1) again means that we only have to prove this for all $P^e \in \mathcal{P}^e$. The latter, however, follows from the decomposition in the previous paragraph. (b) follows.

(c) (V_t^*) *is a super-replication of* η. Under condition (E_2), the result follows directly from Theorem 2.1 (p. 461) of Kramkov (1996). Under the other conditions stated, by (b) above, one can take (V_t^*) to be continuous without losing generality. Hence, by local boundedness, the result also in this case follows from the cited theorem of Kramkov's. ∎

Proof of Corollary 4.2. Let $(V_t^{(n)})$ be a super-replication satisfying $V_0^{(n)} \leq A + 1/n$. Set $V_t = \inf_n V_t^{(n)}$. (V_t) is a supermartingale for all $P^* \in \mathcal{P}^*$. By Proposition 1.3.14 (p. 16) in Karatzas and Shreve (1991), (V_{t+}^*) (taken as a limit through rationals) exists and is a *càdlàg* supermartingale except on a set in $\mathcal{N}$. Hence (V_{t+}^*) is a super-replication of η, with initial value no greater than A. The result follows from Theorem 4.1. ∎

Proof of Proposition 4.3. Suppose that $r_t \geq -c$ for some $c < \infty$. We use Theorem (15.2c) (p. 496) in Jacod (1979). This theorem requires the notation $Ss^1(X)$, which in is the set of probabilities under which the process X_t is indistinguishable from a submartingale so that $E \sup_{0 \leq s \leq t} |X_s| < \infty$ for all t (in our case, t is bounded, so things simplify). (cf. p. 353 and 356 of Jacod (1979).

Jacod's result (15.2c) studies, among other things, the set (in Jacod's notation) $S = \bigcap_{X \epsilon \mathcal{X}} Ss^1(X)$, and under conditions which are satisfied if we take $\mathcal{X}$ to consist of our processes $S_t^{(1)*}, ..., S_t^{(p)*}, -S_t^{(1)*}, ..., -S_t^{(p)*}, \beta_t e^{ct}, Y_t$. Here, $Y_t = 1$ for $t < T$, and I_C for $t = T$. (If necessary, $\beta_t e^{ct}$ can be localized to be bounded, which makes things messier but yields the same result). In other words, S is the set of probability distributions so that the $S_t^{(1)*}, ..., S_t^{(p)*}$ are martingales, r_t is bounded below by c, and the probability of C is one.

Theorem 15.2(c) now asserts a representation of all the elements in the set S in terms of its extremal points. In particular, any set that has probability zero for the extremal elements of S also has probability zero for all other elements of S.

However, $S = \widetilde{M}(\{S^{(1)*}, ..., S^{(p)*}\})$ (again in Jacod's notation, see p. 345 of that work) – this is the set of extremal probabilities among those making $S^{(1)*}, ..., S^{(p)*}$ a martingale. Hence, our Condition (E_1) is proved. ∎

5. Prediction sets: General Theory.

5.1. The Prediction Set Theorem. In the preceding section, we did not take a position on the set of possible probabilities. As mentioned at the beginning of Section 3.3, one can let this set be exceedingly large. Here is one stab at this, in the form of the set $\mathcal{Q}$.

ASSUMPTIONS (A). *(System assumptions). Our probability space is the set $\Omega = \mathbb{C}[0, T]^{p+1}$, and we let $(\beta_t, S_t^{(1)}, ..., S_t^{(p)})$ be the coordinate process, $\mathcal{B}$ is the Borel σ-field, and $(\mathcal{B}_t)$ is the corresponding Borel filtration. We let $\mathcal{Q}^*$ be the set of all distributions P^* on $\mathcal{B}$ so that*

(i) $(\log \beta_t)$ is absolutely continuous P^*-a.s., with derivative r_t bounded (above and below) by a non-random constant, P^*-a.s.;

(ii) the $S_t^{(i)*} = \beta_t^{-1} S_t^{(i)}$ are martingales under P^*;

(iii) $[\log S^{(i)*}, \log S^{(i)*}]_t$ is absolutely continuous P^*-a.s. for all i, with derivative (above and below) by a non-random constant, P^*-a.s. As before, "[,]" is the quadratic variation of the process, see our definition in (3.16) in Section 3.3;

(iv) $\beta_0 = 1$ *and* $S_0^{(i)} = s_0^{(i)}$ *for all* i.

We let $(\mathcal{F}_t)$ *be the smallest right continuous filtration containing* $(\mathcal{B}_{t+})$ *and all sets in* $\mathcal{N}$*, given by*

$$\mathcal{N} \;=\; \{F \subseteq \Omega : \; \forall P^* \in \mathcal{Q}^* \; \exists E \epsilon \mathcal{B} : \; F \subseteq E \text{ and } P^*(E) = 0\}. \tag{5.1}$$

and we let the information at time t *be given by* $\mathcal{F}_t$. *Finally, we let* $\mathcal{Q}$ *be all distributions on* $\mathcal{F}_T$ *that are equivalent (mutually absolutely continuous) to a distribution in* $\mathcal{Q}^*$. *If we need to emphasize the dependence of* $\mathcal{Q}$ *on* $s_0 = (s_0^{(1)}, ..., s_0^{(p)})$*, we write* $\mathcal{Q}_{s_0}$. ∎

REMARK . *An important fact is that* $\mathcal{F}_t$ *is analytic for all* t*, by Theorem III.10 (p. 42) in Dellacherie and Meyer (1978). Also, the filtration* $(\mathcal{F}_t)$ *is right continuous by construction.* $\mathcal{F}_0$ *is a non-informative (trivial)* σ*-field. The relationship of* $\mathcal{F}_0$ *to information from the past (before time zero) is established in Section 5.3.* ∎

The reason for considering this set $\mathcal{Q}$ as our world of possible probability distributions is the following. Stocks and other financial instruments are commonly assumed to follow processes of the form (2.1) or a multidimensional equivalent. The set $\mathcal{Q}$ now corresponds to all probability laws on this form, subject only to certain integrability requirements (for details, see, for example, the version of Girsanov's Theorem given in Karatzas and Shreve (1991), Theorem 3.5.1). Also, if these requirements fail, the S's do not have an equivalent martingale measure, and can therefore not normally model a traded security (see Delbaen and Schachermayer (1995) for precise statements). In other words, roughly speaking, the set $\mathcal{Q}$ covers all distributions of traded securities that have a form (2.1).

Typical forms of the prediction set C would be those discussed in Section 3. If there are several securities $S_t^{(i)}$, one can also set up prediction sets for the quadratic variations and covariations (volatilities and cross-volatilities, in other words). It should be noted that one has to exercise some care in how to formally define the set C corresponding to (2.1) – see the development in Sections 5.2-5.3 below.

The price A_0 is now as follows. A subset of $\mathcal{Q}^*$ is given by

$$\mathcal{P}^* \;=\; \{P^* \in \mathcal{Q}^* \;:\; P^*(C) = 1\}. \tag{5.2}$$

The price is then, from Theorem 5.1 below,

$$A_0 \;=\; \sup\{E^*(\eta^*) : P^* \epsilon \mathcal{P}^*\}, \tag{5.3}$$

where E^* is the expectation with respect to P^*, and

$$\eta^* = \exp\{-\int_0^T r_u du\}\eta. \tag{5.4}$$

It should be emphasized that though (5.2) only involves probabilities that give measure 1 to the set C, this is *only a computational device.* The prediction set C can have any real prediction probability $1-\alpha$, *cf.* statement (5.7) below. The point of Theorem 5.1 is to reduce the problem from $1-\alpha$ to 1, and hence to the discussion in Sections 3 and 4.

We assume the following structure for C.

DEFINITION. A set C in $\mathcal{F}_T$ is $\mathcal{Q}^*$-*closed* if, whenever P_n^* is a sequence in $\mathcal{Q}^*$ for which P_n^* converges weakly to P^* and so that $P_n^*(C) \to 1$, then $P^*(C) = 1$. Weak convergence is here relative to the usual supremum norm on $\mathbb{C}^{p+1} = \mathbb{C}^{p+1}[0,T]$, the coordinate space for $(\beta_., S_.^{(1)}, ..., S_.^{(p)})$. ∎

Obviously, C is $\mathcal{Q}^*$-closed if it is closed in the supremum norm, but the opposite need not be true. See Section 5.2 below.

The precise result is as follows. Note that $-K$ is a credit constraint; see below in this section.

THEOREM 5.1. *(Prediction Region Theorem). Let Assumptions (A) hold. Let C be a $\mathcal{Q}^*$-closed set, $C \in \mathcal{F}_T$. Suppose that $\mathcal{P}^*$ is non-empty. Let*

$$\eta \;=\; \theta(\beta_., S_.^{(1)}, ..., S_.^{(p)}), \tag{5.5}$$

where θ is continuous on Ω (with respect to the supremum norm) and bounded below by $-K\beta_T$, where K is a nonrandom constant ($K \geq 0$). We suppose that

$$\sup_{P^* \in \mathcal{P}^*} E^*|\eta^*| \;<\; \infty \tag{5.6}$$

Then there is a super-replication (A_t) of η on C, valid for all $Q \in \mathcal{Q}$, whose starting value is A_0 given by (5.3). Furthermore, $A_t \geq -K\beta_t$ for all t, $\mathcal{Q}$-a.s.

In particular,

$$Q(A_T \geq \eta) \;\geq\; Q(C) \text{ for all } Q \in \mathcal{Q} \;, \tag{5.7}$$

and this is, roughly, how a $1-\alpha$ prediction set can be converted into a trading strategy that is valid with at least the same probability. This works both in the frequentist and Bayesian cases, as described in Section 5.2. Note that both in Theorem 5.1 and in (5.7), Q refers to all probabilities in $\mathcal{Q}$, and not only the "risk neutral" ones in $\mathcal{Q}^*$.

The form of A_0 and the super-replicating strategy is discussed above in Section 3 and below in Sections 6 and 7 for European options.

The condition that θ be bounded below can be seen as a restriction on credit. Since K is arbitrary, this is not severe. Note that the credit limit is more naturally stated on the discounted scale: $\eta^* \geq -K$, and $A_t^* \geq K$. See also Section 4.2, where a softer bound is used.

The finiteness of credit has another implication. The portfolio (A_t), because it is bounded below, also solves another problem. Let I_C and $I_{\tilde{C}}$ be the indicator functions for C and its complement. A corollary to the statement in Theorem 5.1 is that (A_t) super-replicates the random variable $\eta' \;=\; \eta I_C - K\beta_T I_{\tilde{C}}$. And here we refer to the more classical definition: the super-replication is $\mathcal{Q}-a.s.$, on the entire probability space. This is for free: A_0 has not changed.

It follows that A_0 can be expressed as $\sup_{P^* \in \mathcal{Q}^*} E^*((\eta')^*)$, in obvious notation. Of course, this is a curiosity, since this expression depends on K while A_0 does not.

5.2. Prediction sets: A problem of definition. A main example of this theory is where one has prediction sets for the cumulative interest $-\log \beta_T = \int_0^T r_u du$ and for quadratic variations $[\log S^{(i)*}, \log S^{(j)*}]_T$. For the cumulative interest, the application is straightforward. For example, $\{R^- \leq -\log \beta_T \leq R^+\}$ is a well defined and closed set. For the quadratic (co-)variations, however, one runs into the problem that these are only defined relative to the probability distribution under which they live. In other words, if F is a region in $\mathbb{C}[0,T]^q$, and

$$C_Q \;=\; \{(-\log \beta_t, [\log S^{(i)*}, \log S^{(j)*}]_t, i \leq j)_{0 \leq t \leq T} \in F\}, \tag{5.8}$$

then, as the notation suggests, C_Q will depend on $Q \in \mathcal{Q}$. This is not permitted by Theorem 5.1. The trading strategy cannot be allowed to depend on an unknown $Q \in \mathcal{Q}$, and so neither can the set C. To resolve this problem, and to make the theory more directly operational, the following Proposition 5.2 shows that C_Q has a modification that is independent of Q, and that satisfies the conditions of Theorem 5.1.

PROPOSITION 5.2. *Let F be a set in $\mathbb{C}[0,T]^q$, where $q = \frac{1}{2}p(p-1)+1$. Let F be closed with respect to the supremum norm on $\mathbb{C}[0,T]^q$. Let C_Q be given by (5.8). Then there is a $\mathcal{Q}^*$-closed set C in $\mathcal{F}_T$ so that, for all $Q \in \mathcal{Q}$,*

$$Q\left(C \Delta C_Q\right) = 0, \tag{5.9}$$

where Δ refers to the symmetric difference between sets.

Only the existence of C matters, not its precise form. The reason for this is that relation (5.9) implies that C_{P^*} and C_Q can replace C in (5.2) and (5.7), respectively. For the two prediction sets on which our discussion is centered, (2.3) uses

$$F = \{(x_t)_{0 \le t \le T} \ \in \mathbb{C}[0,T], \text{ nondecreasing } : \ x_0 = 0 \text{ and } \Xi^- \ \le \ x_T \ \le \ \Xi^+\},$$

whereas (3.5) relies on

$$F = \{(x_t)_{0 \le t \le T} \ \in \mathbb{C}[0,T], \text{ nondecreasing } : \ x_0 = 0 \text{ and } \forall s,t \in [0,T],\ s \le t:\ \sigma_-^2(t-s) \le x_t - x_s \le \sigma_+^2(t-s)\}.$$

One can go all the way and jettison the set C altogether. Combining Theorem 5.1 and Proposition 5.2 immediately yields such a result:

THEOREM 5.3. *(Prediction Region Theorem, without Prediction Region). Let Assumptions (A) hold. Let F be a set in $\mathbb{C}[0,T]^q$, where $q = \frac{1}{2}p(p-1)+1$. Suppose that F is closed with respect to the supremum norm on $\mathbb{C}[0,T]^q$. Let C_Q be given by (5.8), for every $Q \in \mathcal{Q}$. Replace C by C_{P^*} in equation (5.2), and suppose that $\mathcal{P}^*$ is non-empty. Impose the same conditions on $\theta(\cdot)$ and $\eta = \theta(\beta_\cdot, S_\cdot^{(1)}, ..., S_\cdot^{(p)})$ as in Theorem 5.1. Then there exists a self financing portfolio (A_t), valid for all $Q \in \mathcal{Q}$, whose starting value is A_0 given by (5.3), and which satisfies (5.7). Furthermore, $A_t \ge -K\beta_t$ for all t, $\mathcal{Q}$-a.s.*

It is somewhat unsatisfying that there is no prediction region anymore, but, of course, C is there, underlying Theorem 5.3. The latter result, however, is easier to refer to in practice.

It should be emphasized that it is possible to extend the original space to include a volatility coordinate. Hence, if prediction sets are given on forms like (2.2) or (2.3), one *can* take the set to be given independently of probability. In fact, this is how Proposition 5.2 is proved.

In the case of European options, this may provide a "probability free" derivation of Theorem 5.1. Under the assumption that the volatility is defined independently of probability distribution, Föllmer (1979) and Bick and Willinger (1994) provide a non probabilistic derivation of Itô's formula, and this can be used to show Theorem 5.1 in the European case. Note, however, that this non probabilistic approach would have a harder time with exotic options, since there is (at this time) no corresponding martingale representation theorem, either for the known probability case (as in Jacod (1979)) or in the unknown probability case (as in Kramkov (1996) and Mykland (2000)). Also, the probability free approach exhibits a dependence on subsequences (see the discussion starting in the last paragraph on p. 350 of Bick and Willinger (1994)).

5.3. Prediction regions from historical data: A decoupled procedure. Until now, we have behaved as if the prediction sets or prediction limits were non random, fixed, and not based on data. This, of course, would not be the case with statistically obtained sets.

Consider the the situation where one has a method giving rise to a prediction set $\hat{C}$. For example, if $C(\Xi^-, \Xi^+)$ is the set from (2.3), then, a prediction set might look like $\hat{C} = C(\hat{\Xi}^-, \hat{\Xi}^+)$, where $\hat{\Xi}^-$ and $\hat{\Xi}^+$ are quantities that are determined (and observable) at time 0.

At this point, one runs into a certain number of difficulties. First of all, C, as given by (2.2) or (2.3), is not well defined, but this is solved through Proposition 5.2 and Theorem 5.3. In addition, there is a question of whether the prediction set(s), A_0, and the process (A_t) are measurable when also functions of data available at time zero. We return to this issue at the end of this section.

From an applied perspective, however, there is a considerably more crucial matter that comes up. It is the question of connecting the model for statistical inference with the model for trading.

What we advocate is the following two stage procedure: (1) find a prediction set C by statistical or other methods, and then (2) trade conservatively using the portfolio that has value A_t. When statistics is used, there are two probability models involved, one for each stage.

We have so far been explicit about the model for Stage (2). This is the nonparametric family $\mathcal{Q}$. For the purpose of inference – Stage (1) – the statistician may, however, wish to use a different family of probabilities. It could also be nonparametric, or it could be any number of parametric models. The choice might depend on the amount and quality of data, and on other information available.

Suppose that one considers an overall family Θ of probability distributions P. If one collects data on the time interval $[T_-, 0]$, and sets the prediction interval based on these data, the $P \in \Theta$ could be probabilities on $\mathbb{C}[T_-, T]^{p+1}$. More generally, we suppose that the P's are distributions on $\mathcal{S} \times \mathbb{C}[0, T]$, where $\mathcal{S}$ is a complete and separable metric space. This permits more general information to go into the setting of the prediction interval. We let $\mathcal{G}_0$ be the Borel σ-field on $\mathcal{S}$. As a matter of notation, we assume that $S_0 = (S_0^{(1)}, ..., S_0^{(p)})$ is $\mathcal{G}_0$-measurable. Also, we let P_ω be the regular conditional probability on $\mathbb{C}[0, T]^{p+1}$ given $\mathcal{G}_0$. (P_ω is well defined; see, for example p. 265 in Ash (1972)). A meaningful passage from inference to trading then requires the following.

NESTING CONDITION: For all $P \in \Theta$, and for all $\omega \in \mathcal{S}$, $P_\omega \in \mathcal{Q}_{S_0}$.

In other words, we do not allow the statistical model Θ to contradict the trading model $\mathcal{Q}$.

The inferential procedure might then consist of a mapping from the data to a random closed set $\hat{F}$. The prediction set is formed using (5.8), yielding

$$\hat{C}_Q = \{(-\log \beta_t, [\log S^{(i)*}, \log S^{(j)*}]_t, i \leq j)_{0 \leq t \leq T} \in \hat{F}\},$$

for each $Q \in \mathcal{Q}_{S_0}$. Then proceed via Proposition 5.2 and Theorem 5.1, or use Theorem 5.3 for a shortcut. In either case, obtain a conservative ask price and a trading strategy. Call these $\hat{A}_0$ and $\hat{A}_t$. For the moment, suspend disbelief about measurability.

To return to the definition of prediction set, it is now advantageous to think of this set as being $\hat{F}$. This is because there are more than one C_Q, and because C is only defined up to measure zero. The definition of a $1-\alpha$ prediction set can then be taken as a requirement that

$$P(\{(-\log\beta_t, [\log S^{(i)*}, \log S^{(j)*}]_t, i \le j)_{0\le t\le T} \in \hat{F}\} \mid \mathcal{H}) \ge 1-\alpha. \tag{5.10}$$

In the frequentist setting, (5.10) must hold for all $P \in \Theta$. $\mathcal{H}$ is a sub-σ-field of $\mathcal{G}_0$, and in the purely unconditional case, it is trivial. By (5.7), $P(\hat{A}_T \ge \eta \mid \mathcal{H}) \ge 1-\alpha$, again for all $P \in \Theta$.

In the Bayesian setting, $\mathcal{H} = \mathcal{G}_0$, and $P(\,\cdot\,|\mathcal{H})$ is a mixture of P_ω's with respect to the posterior distribution $\hat{\pi}$ at time zero. As mentioned after equation (5.7), the mixture would again be in $\mathcal{Q}_{S_0}$, subject to some regularity. Again, (5.7) would yield that $P(\hat{A}_T \ge \eta \mid \mathcal{H}) \ge 1-\alpha$, a.s.

It this discussion, we do not confront the questions that are raised by setting prediction sets by asymptotic methods. Such approximation is almost inevitable in the frequentist setting. For important contributions to the construction of prediction sets, see Barndorff-Nielsen and Cox (1996) and Smith (1999), and the references therein.

It may seem odd to argue for an approach that uses different models for inference and trading, even if the first is nested in the other. We call this the *decoupled prediction approach.* A main reason for doing this is that we have taken inspiration from the cases studied in Sections 3, 6 and 7. One can consider alternatives, however, cf. Section 8 below.

To round off this discussion, we return to the question of the of measurability. There are (at least) three functions of the data where measurability is in question: *(i)* the prediction set $\hat{F}$, *(ii)* the prediction probabilities (5.10) and *(iii)* the starting value $(\hat{A}_0)$.

We here only consider *(ii)* and *(iii)*, since the first question is heavily dependent on Θ and $\mathcal{S}$. In fact, we shall take the measurability of $\hat{F}$ for granted.

Let $\boldsymbol{F}$ be the collection of closed subsets F of $\mathbb{C}[0,T]^q$. We can now consider the following two maps:

$$\boldsymbol{F} \times \mathcal{S} \to \mathbb{R} \;:\; (F,\omega) \to P_\omega(\{(-\log\beta_t, [\log S^{(i)*}, \log S^{(j)*}]_t, i \le j)_{0\le t\le T} \in F) \tag{5.11}$$

and

$$\boldsymbol{F} \times \mathbb{R}^{p+1} \to \mathbb{R} \;:\; (F,x) \to A_0 = \; A_0^F(x). \tag{5.12}$$

Oh, yes, and we need a σ-field on $\boldsymbol{F}$. How can we otherwise do measurability? Make the detour via convergence; $F_n \to F$ if $\limsup F_n = \liminf F_n = F$, which is the same as saying that the indicator functions I_{F_n} converge to I_F point-wise. On $\boldsymbol{F}$, this convergence is metrizable (see the Proof of Proposition 5.4 for one such metric). Hence $\boldsymbol{F}$ has a Borel σ-field. This is our σ-field.

PROPOSITION 5.4. *Let Assumptions (A) hold. Impose the same conditions on* $\theta(\cdot)$ *and* $\eta = \theta(\beta_., S^{(1)}_., ..., S^{(p)}_.)$ *as in Theorem 5.1. Then the maps (5.11) and (5.12) are measurable.*

If we now *assume* that the map $\mathcal{S} \to \boldsymbol{F}$, $\omega \to \hat{F}$, is measurable, then standard considerations yield the measurability of $\mathcal{S} \to \mathbb{R}$, $\omega \to P_\omega(\{(-\log\beta_t, [\log S^{(i)*}, \log S^{(j)*}]_t, i \leq j)_{0\leq t\leq T} \in \hat{F})$ and $\mathcal{S}\times\mathbb{R}^{p+1} \to \mathbb{R}$, $(\omega, x) \to \hat{A}_0 = A_0^{\hat{F}}$. Hence problem *(iii)* is solved, and the resolution of *(ii)* follows since (5.11) equals the expected value of $P_\omega(\{(-\log\beta_t, [\log S^{(i)*}, \log S^{(j)*}]_t, i \leq j)_{0\leq t\leq T} \in \hat{F})$, given $\mathcal{H}$, both in the Bayesian and frequentist cases.

5.4. Proofs for Section 5.

Proof of Theorem 5.1. Assume the conditions of Theorem 5.1. Let $m \geq K$, and define $\theta^{(m)}$ by

$$\theta^{(m)}(\beta_., S^{(1)}_., ..., S^{(p)}_.) = \theta(\beta_., S^{(1)}_., ..., S^{(p)}_.)I_C(\beta_., S^{(1)}_., ..., S^{(p)}_.) - m\beta_T I_{\widetilde{C}}(\beta_., S^{(1)}_., ..., S^{(p)}_.),$$

where $\widetilde{C}$ is the complement of C.

On the other hand, for given probability $P^* \in \mathcal{Q}^*$, define σ_u^{ij} by

$$[\log S^{(i)*}, \log S^{(j)*}]_t = \int_0^t \sigma_u^{ij} du.$$

Also, for c as a positive integer, or $c = +\infty$, set

$$\mathcal{Q}_c^* = \{P^* \in \mathcal{Q}^* \ : \ \sup_t |r_t| + \sum_i \sigma_t^{ii} \leq c\} \ .$$

Let $\mathcal{P}_c^*$ be the set of all distributions in $\mathcal{Q}_c^*$ that vanish outside C. Under Assumptions (A), there is a $c_0 < +\infty$ so that $\mathcal{P}_c^*$ is nonempty for $c \geq c_0$. Also, consider the set $\mathcal{Q}_c^*(t)$ of distributions on $\mathbb{C}[t,T]^{p+1}$ satisfying the same requirements as those above, but instead of (iv) (in Assumption A) that, for all $u \in [0,t]$, $\beta_u = 1$ and $S_u^{(i)} = 1$ for all i.

(1) First, let $c_0 \leq c < +\infty$. Below, we shall make substantial use of the fact that the space $\mathcal{Q}_c^*(t)$ is compact in the weak topology. To see this, invoke Propositons VI.3.35, VI.3.36 and Theorem VI.4.13 (pp. 318 and 322) of Jacod and Shiryaev (1987)).

Consider the functional $\mathbb{C}[0,t]^{p+1} \times \mathcal{Q}_c^*(t) \to \mathbb{R}$ given by

$$\theta_t^{(m)}(b_\cdot, s_\cdot^{(1)}, ..., s_\cdot^{(p)}, P^*) \;=\; E^* b_t \beta_T^{-1} \theta^{(m)}(b_\cdot\beta_\cdot, s_\cdot^{(1)} S_\cdot^{(1)}, ..., s_\cdot^{(p)} S_\cdot^{(p)}).$$

Also, set, for $m \geq K$,

$$\theta_t^{(m)} = (b_\cdot, s_\cdot^{(1)}, ..., s_\cdot^{(p)}) \;=\; \sup_{P^* \in \mathcal{Q}_c^*(t)} \theta_t^{(m)}(b_\cdot, s_\cdot^{(1)}, ..., s_\cdot^{(p)}, P^*).$$

The supremum is $\mathcal{F}_t$-measurable since this σ-field is analytic (see Remark 5.1), and since the space $\mathcal{Q}_c^*(t)$ is compact in the weak topology. The result then follows from Theorems III.9 and III.13 (pp. 42-43) in Dellacherie and Meyer (1978); see also the treatment in Pollard (1984), pp. 196-197.

Since, again, the space $\mathcal{Q}_c^*(t)$ is compact in the weak topology, it follows that the supremum is a bounded. By convergence, $A_t^{(m)*} = \beta_t^{-1}\theta_t^{(m)}(\beta_\cdot, S_\cdot^{(1)}, ..., S_\cdot^{(p)})$ is an $(\mathcal{F}_t)$-supermartingale for all $P^* \in \mathcal{Q}_c^*$. Also, in consequence, $(A_t^{(m)*})$ can be taken to be *càdlàg*, since $(\mathcal{F}_t)$ is right continuous. This is by the construction in Proposition I.3.14 (p. 16-17) in Karatzas and Shreve (1991). Set $A_t^{(m)} = \beta_t A_t^{(m)*}$ (the *càdlàg* version).

(2) Consider the special case where $\eta = -K\beta_T$, and call $\widetilde{A}_t^{(m)*}$ the resulting supermartingale. Note that $\widetilde{A}_t^{(m)*} \leq -K$ on the entire space, and set

$$\tau \;=\; \inf\{t : \widetilde{A}_t^{(m)*} < -K\}.$$

τ is an $\mathcal{F}_t$ stopping time by Example I.2.5 (p. 6) of Karatzas and Shreve (1991).

By definition, $A_t^{(m)*} \geq \widetilde{A}_t^{(m*)}$ everywhere. Since both are supermartingales, we can consider a modified version of $A_t^{(m)*}$ so that it takes new value

$$A_t^{(m)} \;=\; \lim_{u \uparrow \tau} A_u^{(m)} \;\text{ for } \tau \leq t \leq T$$

. In view of Proposition I.3.14 (again) in Karatzas and Shreve (1991), this does not interfere with the supermartingale property of $A_t^{(m)*}$.

Now observe two particularly pertinent facts: (i) The redefinition of $A^{(m)}$ does not affect the initial value, since $\mathcal{P}_c^*$ is nonempty, and (ii) $A_t^{(m)} = A_t^{(K)}$ for all t, since $m \geq K$.

(3) On the basis of this, one can conclude that

$$A_0^{(K)} = \sup_{P^* \in \mathcal{P}_c^*} E^*(\eta^*), \tag{5.13}$$

as follows. By the weak compactness of $\mathcal{Q}_c^*$, there is a P_m^* be such that for given $(b_0, s_0^{(1)}, ..., s_0^{(p)})$, $\theta_0^{(m)}(b., s.^{(1)}, ..., s.^{(p)}) \leq \theta_0^{(m)}(b., s.^{(1)}, ..., s.^{(p)}, P_m^*) + m^{-1}$.

Also, there is a subsequence $P_{m_k}^*$ that converges weakly to some P^*.

Recall that m is fixed, and is greater than K. It is then true that, for $m_k \geq m$, and with $\widetilde{C}$ denoting the complement of C,

$$\begin{aligned} A_0^{(K)*} = A_0^{(m)*} &= \theta_0^{(m)}(b_0, s_0^{(1)}, ..., s_0^{(p)}) \\ &\leq \theta_0^{(m_k)}(b_0, s_0^{(1)}, ..., s_0^{(p)}, P_{m_k}^*) + m_k^{-1} \\ &\leq E_{m_k}^* \beta_T^{-1}\theta(\beta., S.^{(1)}, ..., S.^{(p)}) + P_{m_k}^*(\widetilde{C})(K - m_k) + m_k^{-1} \\ &\leq E_{m_k}^* \beta_T^{-1}\theta(\beta., S.^{(1)}, ..., S.^{(p)}) + P_{m_k}^*(\widetilde{C})(K - m) + m_k^{-1} \\ &\leq E^* \beta_T^{-1}\theta(\beta., S.^{(1)}, ..., S.^{(p)}) + \limsup_{m \to +\infty} P_{m_k}^*(\widetilde{C})(K - m) + o(1) \end{aligned} \tag{5.14}$$

as $k \to \infty$. The first term on the right hand side of (5.14) is bounded by the weak compactness of $\mathcal{Q}_c^*$. The left hand side is a fixed, finite, number. Hence $\limsup P_{m_k}^*(\widetilde{C}) = 0$. By the $\mathcal{Q}^*$-closedness of C, it follows that $P^*(C) = 1$.

Hence, (5.14) yields that the right hand side in (5.13) is an upper bound for $A_0^{(K)*} = A_0^{(m)*}$. Since this is also trivially a lower bound, (5.13) follows.

(4) Now make $A_t^{(m)}$ dependent on c, by writing $A_t^{(m,c)}$. For all $Q^* \in \mathcal{Q}^*$, the $A_t^{(m,c)*}$ are all Q^*-supermartingales, bounded below by $-m$. $A_t^{(m,c)*}$ is nondecreasing in c. Let $A_t^{(m,\infty)}$ denote the limit as $c \to +\infty$. By Fatou's Lemma, for $Q^* \in \mathcal{Q}^*$, and for $s \leq t$,

$$E^*(A_t^{(m,\infty)*} | \mathcal{F}_s) \leq \liminf_{c \to +\infty} E^*(A_t^{(m,c)*} | \mathcal{F}_s) = \liminf_{c \to +\infty} A_s^{(m,c)*} = A_s^{(m,\infty)*}.$$

Hence $A_t^{(m,\infty)*}$ is a supermartingale for all $m \geq K$. Also, by construction, $A_T^{(m,\infty)*} \geq \eta^*$. By the results of Kramkov (1996) or Mykland (2000), $A_{t+}^{(m,\infty)}$ is, therefore, a super-replication of η.

For the case of $t = 0$, (5.13) yields that

$$A_0^{(m,\infty)} = \sup_{P^* \in \mathcal{P}^*} E^*(\eta^*), \tag{5.15}$$

where the non obvious inequality ($\geq$) follows from the monotone convergence, and assumption (5.7). – Since one can choose $m = K$, Theorem 5.1 is proved. ∎

∎

Proof of Proposition 5.2. Extend the space $\mathbb{C}^{p+1}$ to $\mathbb{C}^{p+q}$. Consider the set $\widetilde{\mathcal{Q}}$ of probabilities Q on $\mathbb{C}^{p+q}$ for which the projection onto $\mathbb{C}^{p+1}$ is in $\mathcal{Q}^*$ and so that $([\log S^{(i)*}, \log S^{(j)*}]_t, i \leq j)$ are indistinguishable from $(x_t^{(k)}, k = p{+}2, ..., p{+}q)$. Now consider the set $F' = \{\omega : (-\log\beta, x^{(p+2)}, ..., x^{(p+q)}) \in F\}$. Note that F' is in the completion of $\mathcal{F}_t \otimes \{\mathbb{C}^{q-1},\ \emptyset\}$ with respect to $\widetilde{\mathcal{Q}}$. Hence, there is a C in $\mathcal{F}_T$ so that $P^*(C \Delta F') = 0$ for all $\mathcal{P}^* \in \mathcal{Q}^*$. This is our C.

To show that C is $\mathcal{Q}^*$-closed, suppose that a sequence (in $\mathcal{Q}^*$) $P_n^* \to P^*$ weakly. Construct the corresponding measures $\widetilde{P}_n^*$ and $\widetilde{P}^*$ in $\widetilde{\mathcal{Q}}$. By corollary VI.6.7 (p. 342) in Jacod and Shiryaev (1987), $\widetilde{P}_n^* \to \widetilde{P}^*$ weakly. Hence, since F and hence F' is closed, if $\widetilde{P}_n^*(F') \to 1$, then $\widetilde{P}^*(F') = 1$. The same property must then also hold for C. ∎

Proof of Proposition 5.4 Let d be the uniform metric on C^q, i.e., $d(x,y) = \sum_{i=1,\ldots,q} \sup_{t\in[0,T]} |x_t^i - y_t^i|$. Let $\{z_n\}$ be a countable dense set in C^q with respect to this metric. It is then easy to see that

$$\rho(F,G) = \sum_{n\in\mathbb{N}} \tfrac{1}{2^n}(|d(z_n,F) - d(z_n,G)| \wedge 1)$$

is a metric on $\boldsymbol{F}$ whose associated convergence is the pointwise one.

We now consider the functions $f_m(F,x) = (1 - md(x,F))^+$. These are continuous as maps $\boldsymbol{F} \times \mathbb{C}[0,T]^q \to \mathbb{R}$. From this, the indicator function $I_F(x) = \inf_{m\in\mathbb{N}} f(x)$ is upper semicontinuous, and hence measurable. The result for (5.11) then follows from Exercise 1.5.5 (p.43) in Strook and Varadhan (1979). The development for (5.12) is similar. ∎

6. Prediction sets: The effect of interest rates, and general formulae for European options.

6.1 Interest rates: market structure, and types of prediction sets. When evaluating options on equity, interest rates are normally seen by practitioners as a second order concern. In the following, however, we shall see how to incorporate such uncertainty if one so wishes. It should be emphasized that the following does not discuss interest rate derivatives as such. We suppose that intervals are set on integral form, in the style of (2.3). One could then consider the incorporation of interest rate uncertainty in several ways.

One possibility would be to use a separate interval for the interest rate:

$$R^- \leq \int_0^T r_u du \leq R^+. \tag{6.1}$$

In combination with (2.3), this gives $A = B(S_0, R^+, \Xi^+)$, for convex increasing payoff $f(S_T)$ *cf.* Section 3.2.

For more general European payoffs f, set

$$h(s) \;=\; \sup_{R^- \leq R \leq R^+} \exp\{-R\} f(\exp\{R\} s). \tag{6.2}$$

The bound A then becomes the bound for hedging payoff $h(S_T)$ under interval (2.3). This is seen by the same methods as those used to prove Theorem 6.2 below. Note that when f is convex or concave, then so is h, and so in this case $A \;=\; B(S_0, 0, \Xi^{\pm}; h)$. Here B is as in (3.1), but based on h instead of f. The $\pm$ on Ξ depends on whether f is convex $(+)$ or concave $(-)$. A more general formula is given by (6.13) in Section 6.3.

This value of A_0, however, comes with an important qualification. It is the value one gets by only hedging in the stock S and the money market bond β. But usually one would also have access to longer term bonds. In this case, the value of A would be flawed since it does not respect put-call parity (see p. 167 in Hull (1997). To remedy the situation, we now also introduce the zero coupon treasury bond Λ_t. This bond matures with the value one dollar at the time T which is also the expiration date of the European option.

If such a zero coupon bond exists, and if one decides to trade in it as part of the super-replicating strategy, the price A_0 will be different. We emphasize that there are two if's here. For example, Λ could exist, but have such high transaction cost that one would not want to use it. Or maybe one would encounter legal or practical constraints on its use. - These problems would normally not occur for zero coupon bonds, but can easily be associated with other candidates for "underlying securities". Market traded call and put options, for example, can often exist while being too expensive to use for dynamic hedging. There may also be substantial room for judgment.

We emphasize, therefore, that the price A_0 depends not only on one's prediction region, but also on the market structure. Both in terms of what exists and in terms of what one chooses to trade in. To reflect the ambiguity of the situation, we shall in the following describe Λ as *available* if it is traded and if it is practicable to hedge in it.

If we assume that Λ is, indeed, available, then as one would expect from Section 3, different prediction regions give different values of A_0. If one combines (2.3) and (6.1), the form of A_0, is somewhat unpleasant. We give the details in Section 6.4. Also, one suffers from the problem of setting a two dimensional prediction region, which will require prediction probabilities in each dimension that will be higher than $1 - \alpha$.

A better approach is the following. This elegant way of dealing with uncertain interest was first encountered by this author in the work of El Karoui, Jeanblanc-Picqué and Shreve (1998). Consider the stock price discounted (or rather, blown up) by the zero coupon bond:

$$S_t^{(*)} = S_t/\Lambda_t. \tag{6.3}$$

In other words, $S_t^{(*)}$ is the price of the forward contract that delivers S_T at time T. Suppose that the process $S^{(*)}$ has volatility σ_t^*, and that we now have prediction bounds similar to (2.3), in the form

$$\Xi^{*-} \leq \int_0^T \sigma_t^{*2} dt \leq \Xi^{*+}. \tag{6.4}$$

We shall see in Section 6.3 that the second interval gives rise to a nice form for the conservative price A_0. For convex European options such as puts and calls, $A_0 = B(S_0, -\log \Lambda_0, \Xi^{*+})$. The main gain from using this approach, however, is that it involves a scalar prediction interval. There is only one quantity to keep track of. And no multiple comparison type problems.

The situation for the call option is summarized in Table 3. The value A_0 depends on two issues: is the zero coupon bond available, and which prediction region should one use?

TABLE 3

Comparative prediction sets: r nonconstant

Convex European options, including calls

Λ_t available?	A_0 from (2.3) and (6.1)	A_0 from (6.4)
no	$B(S_0, R^+, \Xi^+)$	not available
yes	see Section 6.4	$B(S_0, -\log\Lambda_0, \Xi^{*+})$

B is defined in (3.2)-(3.3) for call options, and more generally in (3.1).

Table 3 follows directly from the development in Section 6.3. The hedge ratio corresponding to (6.4) is given in (6.12) below.

6.2. The effect of interest rates: the case of the Ornstein-Uhlenbeck model. We here discuss a particularly simple instance of incorporating interest rate uncertainty into the interval (6.4). In the following, we suppose that interest rates follow a linear model (introduced in the interest rate context by Vasicek (1977)),

$$dr_t = a_r(b_r - r_t)dt + c_r dV_t, \tag{6.5}$$

where V is a Brownian motion independent of B in (2.1).

The choice of interest rate model highlights the beneficial effects of the "decoupled" prediction procedure (Section 5.3): this model would be undesirable for hedging purposes as it implies that any government bond can be hedged in any other government bond, but on the other hand it may not be so bad for statistical purposes. Incidentally, the other main conceptual criticism of this model is that rates can go negative. Again, this is something that is less bothersome for a statistical analysis than for a hedging operation. This issue may, however, have become obsolete with the recent apparent occurrence of negative rates in Japan (see, *e.g.*, "Below zero" (*The Economist*, Nov. 14, 1998, p.81)).

Suppose that the time T to maturity of the discount bond Λ is sufficiently short that there is no risk adjustment, in other words, $\Lambda_0 = E\exp\{-\int_0^T r_t dt\}$. One can then parametrize the quantities of interest as follows: there are constants ν and γ so that

$$\int_0^T r_t dt \text{ has distribution } N(\nu, \gamma^2). \tag{6.6}$$

It follows that

$$\log \Lambda_0 = -\nu + \frac{1}{2}\gamma^2. \tag{6.7}$$

In this case, if we suppose that the stock follows (2.1), then

$$\int_0^T \sigma_u^{*2} du = \int_0^T \sigma_u^2 du + \gamma^2. \tag{6.8}$$

Prediction intervals can now be adjusted from (2.3) to (6.4) by incorporating the estimation uncertainty in γ^2. – Nonlinear interest rate models, such as the one from Cox, Ingersoll and Ross (1985), require, obviously, a more elaborate scheme.

6.3. General European options. We here focus on the single prediction set (6.4). The situation of constant interest rate (Table 1 in Section 3.3) is a special case of this, where the prediction set reduces to (2.3).

THEOREM 6.1. *Under the Assumptions (A), and with prediction set (6.4), if one hedges liability $\eta = g(S_T)$ in S_t and Λ_t, the quantity A has the form*

$$A_0 = \sup_\tau \widetilde{E}\Lambda_0 f(\frac{1}{\Lambda_0}\widetilde{S}_\tau), \tag{6.9}$$

where the supremum is over all stopping times τ that take values in $[\Xi^{*-}, \Xi^{*+}]$, and where $\widetilde{P}$ is a probability distribution on $\mathbb{C}[0,T]$ so that

$$d\widetilde{S}_t = \widetilde{S}_t d\widetilde{W}_t, \text{ with } \widetilde{S}_0 = s_0, \tag{6.10}$$

where s_0 is the actual observed value of S_0.

If one compares this with the results concerning nonconstant interest below in Section 6.4, the above would seem to be more elegant, and it typically yields lower values for A_0. It is also easier to implement since $\widetilde{S}$ is a martingale.

Now consider the case of convex or concave options. The martingale property of $\widetilde{S}$ yields that the A_0 in (6.9) has the value

$$A_0 \;=\; B(S_0, -\log \Lambda_0, \Xi^{*\pm}) \; . \tag{6.11}$$

As in Section 6.1, $\pm$ depends on whether f is convex of concave.

It is shown in Section 6.5 that the delta hedge ratio for convex g is

$$\tfrac{\partial B}{\partial S}(S_t, -\log \Lambda_t, \Xi^{*+} - \int_0^t \sigma_u^{*2} du). \tag{6.12}$$

In practice, one has to make an adjustment similar to that at the end of Section 3.3.

As a consequence of Theorem 6.1, we can also state the form of the value A when hedging only in stock and the money market bond. If h is defined as in (6.2), one gets similarly to (6.9) that

$$A_0 \;=\; \sup_\tau \widetilde{E} h(\widetilde{S}_\tau), \tag{6.13}$$

6.4. General European options: The case of two intervals and a zero coupon bond. Now assume that we have a prediction set consisting of the two intervals (2.3) and (6.1). We can now incorporate the uncertainty due to interest rates as follows. First form the auxiliary function

$$h(s, \lambda; f) \;=\; \sup_{R^- \leq R \leq R^+} \exp\{-R\}[f(\exp\{R\}s) - \lambda] + \lambda\Lambda_0 \qquad (6.14).$$

Our result is now that the price for the dynamic hedge equals the price for the best static hedge, and that it has the form of the price of an American option.

THEOREM 6.2. *Under the assumptions above, if one hedges in S_t and Λ_t, the quantity A has the form*

$$A_0(f) \;=\; \inf_\lambda \sup_\tau \widetilde{E} h(\widetilde{S}_\tau, \lambda; f) \tag{6.15}$$

where $\widetilde{P}$ is the probability distribution for which

$$d\widetilde{S}_t = \widetilde{S}_t d\widetilde{W}_t, \ \widetilde{S}_0 = S_0 \tag{6.16}$$

and τ is any stopping time between Ξ^- and Ξ^+.

As above, if f is convex or concave, then so is the h in (6.14). In other words, since convex functions of martingales are submartingales, and concave ones are supermartingales (see, for example, Karatzas and Shreve (1991), Proposition I.3.6 (p. 13)), the result in Theorem 6.2 simplifies in those cases:

$$\begin{aligned} f \text{ convex: } A_0 &= \inf_\lambda \widetilde{E} h(\widetilde{S}_{\Xi^+}, \lambda; f), \text{ and} \\ f \text{ concave: } A_0 &= \inf_\lambda \widetilde{E} h(\widetilde{S}_{\Xi^-}, \lambda; f), \end{aligned} \tag{6.17}$$

both of which expressions are analytically computable.

We emphasize that what was originally cumulative volatilities (Ξ^-, Ξ^+) have now become measures of time when computing (6.15). This is because of the Dambis (1965)/Dubins-Schwartz (1965) time change, which leads to time being measured on the volatility scale.

REMARK 6.1. *Note that in Theorem 6.2, the optimization involving R and λ can be summarized by replacing (6.15) with $A(f) = \sup_\tau \widetilde{E} g(\widetilde{S}_\tau; f)$, where $g(s; f)$ is the supremum of $Eh(s, \lambda; f)$ over (random variables) $R \in [R^-, R^+]$, subject to $E(\exp\{-R\}) = \Lambda_0$. R becomes a function of s, which in the case of convex f will take values R^- and R^+. This type of development is further pursued in Section 7 below.*

REMARK 6.2. *Bid prices are formed similarly. In Theorem 6.2,*

$$B(f) = \sup_\lambda \inf_\tau \widetilde{E} h(\widetilde{S}_\tau, \lambda; f).$$

This is as in equation (2.6).

The expression for $A(f)$ for the call option, $f(s) = (s - K)^+$, is the following. If v_0 solves

$$\Phi(d_2(S_0, v_0, \Xi^+)) = \frac{\exp(-R^-) - \Lambda_0}{\exp(-R^-) - \exp(-R^+)},$$

where Φ is the cumulative normal distribution and in the same notation as in (3.2)–(3.3), then one can start a super-replicating strategy with the price at time zero given in the following:

$$
\begin{aligned}
v_0 \geq R^+ &: C(S_0, R^+, \Xi^+) \\
R^+ > v_0 > R^- &: C(S_0, v_0, \Xi^+) + K\left(\exp(-v_0) - \exp(-R^+)\right)\Phi\left(d_2(S_0, v_0, \Xi^+)\right) \\
v_0 \leq R^- &: C(S_0, R^-, \Xi^+) + K\left(\exp(-R^-) - \Lambda_0\right)
\end{aligned}
$$

6.5. Proofs for Section 6.

Proof of Theorem 6.1. The A_t be a self financing trading strategy in S_t and Λ_t that covers payoff $g(S_T)$. In other words,

$$dA_t \;=\; \theta_t^{(0)} d\Lambda_t + \theta_t^{(1)} dS_t \text{ and } A_t \;=\; \theta_t^{(0)} \Lambda_t + \theta_t^{(1)} S_t$$

If $S_t^{(*)} = \Lambda_t^{-1} S_t$, and similarly for $A_t^{(*)}$, this is the same as asserting that

$$dA_t^{(*)} \;=\; \theta_t^{(1)} dS_t^{(*)}.$$

This is by numeraire invariance and/or Itô's formula. In other words, for a fixed probability P, under suitable regularity conditions, the price of payoff $g(S_T)$ is $A_0 = \Lambda_0 A_0^{(*)} = \Lambda_0 E^{(*)} A_T^{(*)} = \Lambda_0 E^{(*)} g(S_T^{(*)})$, where $P^{(*)}$ is a probability distribution equivalent to P under which $S^{(*)}$ is a martingale.

It follows that Theorem 5.1 can be applied as if $r = 0$ and one wishes to hedge in security $S_t^{(*)}$. Hence, it follows that

$$A_0 = \sup_{P^* \in \mathcal{P}^*} \Lambda_0 E^{(*)} g(S_T^{(*)})$$

By using the Dambis (1965)/Dubins-Schwarz (1965) time change, the result follows. ∎

Derivation of the hedging strategy (6.12). As discussed in Section 3.2, the function $B(S, R, \Xi)$ defined in (3.1), satisfies two partial differential equations, *viz.*, $\frac{1}{2} B_{SS} S^2 = B_\Xi$ and $-B_R = B - B_S S$. It follows that $B_{RR} = B_R - B_{SR} S$ and $B_{RS} = B_{SS} S$.

Now suppose that Ξ_t is a process with no quadratic variation. We then get the following from Itô's Lemma:

$$\begin{aligned} dB(S_t, \Xi_t, -\log\Lambda_t) &= B_S dS_t - B_R \frac{1}{\Lambda_t} d\Lambda_t \\ &\quad + B_\Xi (d < \log S^* >_t + d\Xi_t) \end{aligned} \tag{6.18}$$

If one looks at the right hand side of (6.18), the first line is the self financing component in the trading strategy. One should hold $B_S(S_t, \Xi_t, -\log\Lambda_t)$ units of stock, and $B_R(S_t, \Xi_t, -\log\Lambda_t)/\Lambda_t$ units of the zero coupon bond Λ. In order for this strategy to not require additional input during the life of the option, one needs the second line in (6.18) to be nonpositive. In the case of a convex or concave payoff, one just uses $d\Xi_t = -d < \log S^* >_t$, with Ξ_0 as Ξ^{*+} or Ξ^{*-}, as the case may be. ∎

Proof of Theorem 6.2. By Theorem 5.1,

$$A_0 = \sup_{P^* \in \mathcal{P}^*} E_{P^*} \exp\{-\int_0^T r_u du\} f(\exp\{\int_0^T r_u du\} S_T^*).$$

For a given $P^* \in \mathcal{P}^*$, define $P^{(1)}$, also in $\mathcal{P}^*$, by letting $v > 1$, $\sigma_t^{\text{new}} = \sigma_{vt}$ for $vt \leq T$ and zero thereafter until T. whereas we let $r_t^{\text{new}} = 0$ until T/v, and thereafter let $r_t^{\text{new}} = r_{(vt-T)/(v-1)}$. On the other hand, define $P^{(2)}$, also in $\mathcal{P}^*$, by letting $r_t^{(2)} = 0$ for $t < T/v$, and $r_t^{(2)} = Rv/T(1-v)$, where R maximizes the right hand side of (6.14) given $s = S_T^*$ and subject to $E\exp\{-R\} = \Lambda_0$.

Obviously,

$$\begin{aligned} E_{\mathcal{P}^*} \exp\{-\int_0^T r_u du\} f(\exp\{\int_0^T r_u du\} S_T^*) &= E_{\mathcal{P}^{(1)}} \exp\{-\int_0^T r_u du\} f(\exp\{\int_0^T r_u du\} S_T^*) \\ &\leq E_{\mathcal{P}^{(2)}} \exp\{-\int_0^T r_u du\} f(\exp\{\int_0^T r_u du\} S_T^*) \\ &\leq \inf_\lambda E_{\mathcal{P}^{(2)}} h(S_T^*, \lambda; f) \end{aligned}$$

by a standard constrained optimization argument.

By using the Dambis (1965)/Dubins-Schwarz (1965) time change (see, *e.g.*, Karatzas and Shreve (1991), p. 173-179), (6.15)-(6.16) follows. ∎

7. Prediction sets and the interpolation of options. .

7.1. Motivation.

A major problem with a methodology that involves intervals for prices is that these can, in many circumstances, be too wide to be useful. There is scope, however, for narrowing these intervals by hedging in auxiliary securities, such as market traded derivatives. The purpose of this section is to show that this can be implemented for European options. A general framework is briefly described in Section 7.2. In order to give a concise illustration, we show how to interpolate call options in Section 7.3. As we shall see, this interpolation substantially lowers the upper interval level A from (2.8).

Similar work with different models has been carried out by Bergman (1995), and we return to the connection at the end of Section 7.3. Our reduction of the option value to an optimal stopping problem, both in Theorem 7.1 and above in Theorem 6.1, mirrors the development in Frey (2000). Frey's paper uses the bounds of Avellaneda, Levy and Paras (cf. Assumption 3 (p. 166) in his paper; the stopping result is Theorem 2.4 (p. 167)). In this context, Frey (2000) goes farther than the present paper in that it also considers certain types of non-European options. See also Frey and Sin (1999).

7.2. Interpolating European payoffs.

We first describe the generic case where restrictions on the volatility and interest rates are given by

$$\Xi^- \leq \int_0^T \sigma_t^2 dt \leq \Xi^+ \text{ and } R^- \leq \int_0^T r_u du \leq R^+. \tag{7.1}$$

We suppose that there market contains a zero coupon bond, there are p market traded derivatives $V_t^{(i)}$ $(i = 1, ..., p)$ whose payoffs are $f_i(S_T)$ at time T. Again, it is the case that the price for the dynamic hedge equals the best price for a static hedge in the auxiliary securities, with a dynamic one in S_t only:

THEOREM 7.1. *Under the assumptions above, if one hedges in S_t, Λ_t, and the $V_t^{(i)}$ ($i = 1, ..., p$), the quantity A has the form*

$$A(f; f_1, ..., f_p) = \inf_{\lambda_1 ..., \lambda_p} A(f - \lambda_1 f_1 - ... \lambda_p f_p) + \sum_{i=1}^{p} \lambda_i V_0^{(i)}, \tag{7.2}$$

where $A(f - \lambda_1 f_1 - ... \lambda_p f_p)$ is as given by (6.15)-(1.16).

A special case which falls under the above is one where one has a prediction interval for the volatility of the future S^* on S. Set $S_t^* = S_t/\Lambda_t$, and replace equation (2.1) by $dS_t^* = \mu_t S_t^* dt + \sigma_t S_t^* dW_t^*$. S^* is then the value of S in numeraire Λ, and the interest rate is zero in this numeraire. By numeraire invariance, one can now treat the problem in this unit of account. If one has an interval or the form (6.4), this is therefore the same as the problem posed in the form (7.1), with $R^- = R^+ = 0$. There is no mathematical difference, but (6.4) is an interval for the volatility of the future S^* rather than the actual stock price S. This is similar to what happens in Theorem 6.1.

Still with numeraire Λ, The Black-Sholes price is $B(S_0, \Xi, -\log \Lambda_0; f)/\Lambda_0 = B(S_0^*, \Xi, 0; f)$. In this case, h (from (6.14)) equals f. Theorems 7.1-7.2, Algorithm 7.1, and Corollary 7.3 go through unchanged. For example, equation (6.15) becomes (after reconversion to dollars) $A(f) = \Lambda_0 \sup_\tau \widetilde{E} f(\widetilde{S}_\tau)$, where the initial value in (6.16) is $\widetilde{S}_0 = S_0^* = S_0/\Lambda_0$.

7.3. The case of European calls.

To simplify our discussion, we shall in the following assume that the short term interest rate r is known, so that $R^+ = R^- = rT$. This case also covers the case of the bound (6.4). We focus here on the volatility only since this seems to be the foremost concern as far as uncertainty is concerned. In other words, our prediction interval is

$$\Xi^+ \geq \int_0^T \sigma_u^2 du \geq \Xi^-. \tag{7.3}$$

Consider, therefore, the case where one wishes to hedge an option with payoff $f_0(S_T)$, where f_0 is (non strictly) convex. We suppose that there are, in fact, market traded call options $V_t^{(1)}$ and $V_t^{(2)}$ with strike prices K_1 and K_2. We suppose that $K_1 < K_2$, and set $f_i(s) = (s - K_i)^+$.

From Theorem 7.1, the price A at time 0 for payoff $f_0(S_T)$ is

$$A(f_0; f_1, f_2) = \inf_{\lambda_1,\lambda_2} \sup_\tau \widetilde{E}\ (h - \lambda_1 h_1 - \lambda_2 h_2)(\widetilde{S}_\tau) + \sum_{i=1}^{2} \lambda_i V_0^{(i)}, \tag{7.4}$$

where, for $i = 1, 2$, $h_i(s) = \exp\{-rT\} f_i(\exp\{rT\}s) = (s - K_i')^+$, with $K_i' = \exp\{-rT\}K_i$.

We now give an algorithm for finding A.

For this purpose, let $B(S, \Xi, R, K)$ be as defined in (3.1) for $f(s) = (s-K)^+$ (in other words, the Black-Scholes-Merton price for a European call with strike price K). Also define, for $\Xi \leq \widetilde{\Xi}$,

$$\widetilde{B}(S, \Xi, \widetilde{\Xi}, K, \widetilde{K}) \ = \ \widetilde{E}((\widetilde{S}_\tau - \widetilde{K})^+ \mid S_0 = S), \tag{7.5}$$

where τ is the minimum of $\widetilde{\Xi}$ and the first time after Ξ that $\widetilde{S}_t$ hits K. An analytic expression for (7.5) is given as equation (7.15) in Section 7.5.

Algorithm 7.1.

(i) Find the implied volatilities Ξ_i^{impl} *of the options with strike price* K_i. *In other words,* $\widetilde{B}(S_0, \Xi_i^{\mathrm{impl}}, rT, K_i) = V_0^{(i)}$.

(ii) If $\Xi_1^{\mathrm{impl}} < \Xi_2^{\mathrm{impl}}$, *set* $\Xi_1 = \Xi_1^{\mathrm{impl}}$, *but adjust* Ξ_2 *to satisfy* $\widetilde{B}(S_0, \Xi_1^{\mathrm{impl}}, \Xi_2, K_1', K_2') = V_0^{(2)}$. *If* $\Xi_1^{\mathrm{impl}} > \Xi_2^{\mathrm{impl}}$, *do the opposite, in other words, keep* $\Xi_2 = \Xi_2^{\mathrm{impl}}$, *and adjust* Ξ_1 *to satisfy* $\widetilde{B}(S_0, \Xi_2^{\mathrm{impl}}, \Xi_1, K_2', K_1') = V_0^{(1)}$. *If* $\Xi_1^{\mathrm{impl}} = \Xi_2^{\mathrm{impl}}$, *leave them both unchanged, i.e.,* $\Xi_1 = \Xi_2 = \Xi_1^{\mathrm{impl}} = \Xi_2^{\mathrm{impl}}$.

(iii) Define a stopping time τ *as the minimum of* Ξ^+, *the first time* $\widetilde{S}_t$ *hits* K_1' *after* Ξ_1, *and the first time* $\widetilde{S}_t$ *hits* K_2' *after* Ξ_2. *Then*

$$A(f_0; f_1, f_2) \ = \ \widetilde{E} h_0(\widetilde{S}_\tau).$$

Note in particular that if f_0 is also a call option, with strike K_0, and still with the convention $K_0' = \exp\{-rT\}K_0$, one obtains

$$A \ = \ \widetilde{E}(\widetilde{S}_\tau - K_0')^+. \tag{7.6}$$

This is the sense in which one could consider the above an interpolation or even extrapolation: the strike prices K_1 and K_2 are given, and K_0 can now vary.

THEOREM 7.2. *Suppose that $\Xi^- \leq \Xi_1^{\text{impl}}, \Xi_2^{\text{impl}} \leq \Xi^+$. Then the A found in Algorithm 1 coincides with the one given by (7.4). Furthermore, for $i = 1, 2$,*

$$\Xi_i^{\text{impl}} \leq \Xi_i. \tag{7.7}$$

Note that the condition $\Xi^- \leq \Xi_1^{\text{impl}}, \Xi_2^{\text{impl}} \leq \Xi^+$ must be satisfied to avoid arbitrage, assuming one believes the bound (7.3). Also, though Theorem 7.2 remains valid, no-arbitrage considerations impose constraints on Ξ_1 and Ξ_2, as follows.

COROLLARY 7.3. *Assume $\Xi^- \leq \Xi_1^{\text{impl}}, \Xi_2^{\text{impl}} \leq \Xi^+$. Then Ξ_1 and Ξ_2 must not exceed Ξ^+. Otherwise there is arbitrage under the condition (7.3).*

We prove the algorithm and the corollary in Section 7.5. Note that $\widetilde{B}(S, \Xi, \widetilde{\Xi}, K, \widetilde{K})$ in (7.5) is a down-and-out type call for $\widetilde{K} \geq K$, and can be rewritten as an up-and-out put for $\widetilde{K} < K$, and is hence obtainable in closed form – cf. equation (7.15) in Section 7.5. A in (7.6) has a component which is on the form of a double barrier option, so the analytic expression (which can be found using the methods in Chapter 2.8 (p.94-103) in Karatzas and Shreve (1991)) will involve an infinite sum (as in *ibid*, Proposition 2.8.10 (p. 98)). See also Geman and Yor (1996) for analytic expressions. Simulations can be carried out using theory in Asmussen, Glynn and Pitman (1995), and Simonsen (1997).

The pricing formula does not explicitly involve Ξ^-. It is implicitly assumed, however, that the implied volatilities of the two market traded options exceed Ξ^-. Otherwise, there would be arbitrage opportunities. This, obviously, is also the reason why one can assume that $\Xi_i^{\text{impl}} \leq \Xi^+$ for both i.

How does this work in practice? We consider an example scenario in figures 7.1 and 7.2. We suppose that market traded calls are sparse, so that there is nothing between $K_1 = 100$ (which is at the money), and $K_2 = 160$. Figure 7.1 gives implied volatilities of A as a function of the upper limit Ξ^+. Figure 7.2 gives the implied volatilities as a function of K_0. As can be seen from the plots, the savings over using volatility Ξ^+ are substantial.

[figures 1 and 2 approximately here]

All the curves in Figure 7.1 have an asymptote corresponding to the implied volatility of the price $A_{\text{crit}} = \lambda_1^{(0)} V_0^{(1)} + (1 - \lambda_1^{(0)}) V_0^{(2)}$, where $\lambda_1^{(0)} = (K_2 - K_0)/(K_2 - K_1)$. This is known as the Merton bound, and holds since, obviously, $\lambda_1^{(0)} S_t^{(1)} + (1 - \lambda_1^{(0)}) S_t^{(2)}$ dominates the call option with strike price K_0, and is the cheapest linear combination of $S_t^{(1)}$ and $S_t^{(2)}$ with this property. In fact, if one denotes as A_{Ξ^+} the quantity from (7.6), and if the Ξ_i^{impl} are kept fixed, it is easy to see that, for (7.6),

$$\lim_{\Xi^+ \to +\infty} A_{\Xi^+} = A_{\text{crit}}. \tag{7.8}$$

Figures 7.1 and 7.2 presuppose that the implied volatility of the two market traded options are the same ($\sqrt{\Xi_1^{\text{impl}}} = \sqrt{\Xi_2^{\text{impl}}} = 0.2$). To see what happens when the out of the money option increases its implied volatility, we fix $\sqrt{\Xi_1^{\text{impl}}} = 0.2$, and we show in the following the plot of $\sqrt{\Xi_2}$ as a function of $\sqrt{\Xi_2^{\text{impl}}}$. Also, we give the implied volatilities for the interpolated option (7.6) with strike price $K_0 = 140$. We see that except for high $\sqrt{\Xi_2^{\text{impl}}}$, there is still gain by a constraint on the form (7.3).

[figures 3 and 4 approximately here]

It should be noted that there is similarly between the current paper and the work by Bergman (1995). This is particularly so in that he finds an arbitrage relationship between the value of two options (see his Section 3.2 (pp. 488-494), and in particular Proposition 4). Our development, similarly, finds an upper limit for the price of a third option given two existing ones. As seen in Corollary 7.3, it can also be applied to the relation between two options only.

The similarly, however, is mainly conceptual, as the model assumptions are substantially different. An interest rate interval (Bergman's equations (1)-(2) on p. 478) is obtained by differentiating between lending and borrowing rates (as also in Cvitanić and Karatzas (1993)), and the stock price dynamic is given by differential equations (3)-(4) on p. 479. This is in contrast to our assumptions (7.1). It is, therefore, hard to compare Bergman's and our results in other than conceptual terms.

7.4. The usefulness of interpolation.

We have shown in the above that the interpolation of options can substantially reduce the length of intervals for prices that are generated under uncertainty in the predicted volatility and interest rates. It would be natural to extend the approach to the case of several securities, and this is partially carried out in Mykland (2003c). Also one should confront the common reality that the volatility itself is quite well pinned down, whereas correlations are not. An even more interesting question is whether this kind of nonparametrics can be used in connection with the interest rate term structure, where the uncertainty about models is particularly acute.

7.5. Proofs for Section 7.

Proof of Theorem 7.1. This result follows in a similar way to the proof of Theorem 6.2, with the modification that $\mathcal{Q}^*$ is now the set of all probability distributions Q^* so that (7.1) is satisfied, so that Λ_t^* *and* the $V_t^{(i)*}$ $(i = 1, ..., p)$ are martingales, and so that $dS_t^* = \sigma_t S_t^* dW_t$, for given S_0. ∎

Before we proceed to the proof of Theorem 7.2, let us establish the following set of inequalities for $\Xi < \widetilde{\Xi}$,

$$B(S, \Xi, R, K_2) < \widetilde{B}(S, \Xi, \widetilde{\Xi}, K_1', K_2') < B(S, \widetilde{\Xi}, R, K_2). \tag{7.9}$$

The reason for this is that $\widetilde{B}(S, \Xi, \widetilde{\Xi}, K_1', K_2') = \widetilde{E}((\widetilde{S}_\tau - K_2')^+)$ is nondecreasing in both Ξ and $\widetilde{\Xi}$, since $\widetilde{S}$ is a martingale and $x \to x^+$ is convex, and also that $\widetilde{B}(S, \Xi, \Xi, K_1', K_2') = B(S, \Xi, 0, K_2') = B(S, \Xi, R, K_2)$. The inequalities are obviously strict otherwise.

Proof of Theorem 7.2 (and Algorithm 7. 1). We wish to find (7.4).First fix λ_1 and λ_2, in which case we are seeking $\sup_\tau \widetilde{E} h_{\lambda_1,\lambda_2}(\widetilde{S}_\tau)$, where $h_{\lambda_1,\lambda_2} = h_0 - \lambda_1 h_1 - \lambda_2 h_2$. This is because the $V_0^{(i)}$ are given. We recall that h_0 is (non strictly) convex since f_0 has this property, and that $h_i(s) = (s - K_i')^+$. It follows that h_{λ_1,λ_2} is convex except at points $s = K_1'$ and $= K_2'$.

Since $\widetilde{S}_t$ is a martingale, $h_{\lambda_1,\lambda_2}(\widetilde{S}_t)$ is therefore a submartingale so long as $\widetilde{S}_t$ does not cross K_1' or K_2' (see Proposition I.3.6 (p. 13) in Karatzas and Shreve (1991)). It follows that if τ_0 is a stopping time, $\Xi^- \leq \tau_0 \leq \Xi^+$, and we set

$$\tau = \inf\{\ t \geq \tau_0 : \widetilde{S}_t = K_1' \text{ or } K_2'\ \} \wedge \Xi^+,$$

then $\widetilde{E}h_{\lambda_1,\lambda_2}(\widetilde{S}_{\tau_0}) \leq \widetilde{E}h_{\lambda_1,\lambda_2}(\widetilde{S}_\tau)$. It follows that the only possible optimal stopping points would be $\tau = \Xi^+$ and τs for which $\widetilde{S}_\tau = K_i'$ for $i = 1, 2$.

Further inspection makes it clear that the rule must be on the form given in part (iii) of the algorithm, but with Ξ_1 and Ξ_2 as yet undetermined. This comes from standard arguments for American options (see Karatzas (1988), Myneni (1992), and the references therein), as follows. Define the *Snell envelope* for h_{λ_1,λ_2} by

$$\mathrm{SE}(s, \Xi) = \sup_{\Xi \leq \tau \leq \Xi^+} \widetilde{E}(h_{\lambda_1,\lambda_2}(\widetilde{S}_\tau) \mid S_\Xi = s).$$

The solution for American options is then that

$$\tau \;=\; \inf\{\ \xi \geq \Xi^- \ :\ \ \mathrm{SE}(\widetilde{S}_\xi, \xi) \;=\; h_{\lambda_1,\lambda_2}(\widetilde{S}_\xi)\ \}$$

Inspection of the preceding formula yields that $\tau = \tau_1 \wedge \tau_2$, where

$$\begin{aligned}
\tau_i &= \inf\{\ \xi \geq \Xi^- \ :\ \{\ \mathrm{SE}(\widetilde{S}_\xi, \xi) \;=\; h_{\lambda_1,\lambda_2}(\widetilde{S}_\xi)\ \} \cap \{\ \widetilde{S}_\xi = K_i'\ \}\ \} \wedge \Xi^+ \\
&= \inf\{\ \xi \geq \Xi^- \ :\ \{\ \mathrm{SE}(K_i', \xi) \;=\; h_{\lambda_1,\lambda_2}(K_l')\ \} \cap \{\ \widetilde{S}_\xi = K_i'\ \}\ \} \wedge \Xi^+ \\
&= \inf\{\ \xi \geq \Xi_i \ :\ \widetilde{S}_\xi = K_i'\ \} \wedge \Xi^+,
\end{aligned}$$

where $\Xi_i = \inf\{\ \xi \geq \Xi^- \ :\ \ \mathrm{SE}(K_i', \xi) \;=\; h_{\lambda_1,\lambda_2}(K_i')\ \} \wedge \Xi^+$.

Since the system in linear in λ_1 and λ_2, and in analogy with the discussion in Remark 7.1, it must be the case that

$$\widetilde{E}(\widetilde{S}_\tau - K_i')^+ = V_0^{(i)} \text{ for } i = 1, 2. \tag{7.10}$$

Hence the form of A given in part (iii) of the algorithm must be correct, and one can use (7.10) to find Ξ_1 and Ξ_2. Note that the left hand side of (7.10) is continuous and increasing in Ξ_1 and Ξ_2, (again since $\widetilde{S}$ is a martingale and $x \to x^+$ is convex). Combined with our assumption in Theorem 7.2 that $\Xi^- \leq \Xi_1^{\mathrm{impl}}, \Xi_2^{\mathrm{impl}} \leq \Xi^+$, we are assured that (7.10) has solutions Ξ_1 and Ξ_2 in $[\Xi^-, \Xi^+]$.

Let (Ξ_1, Ξ_2) be a solution for (7.10) (we have not yet decided what values they take, or even that they are in the interval $[\Xi^-, \Xi^+]$).

Suppose first that $\Xi_1 < \Xi_2$.

It is easy to see that

$$\widetilde{E}[(\widetilde{S}_\tau - K_1')^+ \mid \widetilde{S}_{\Xi_1}] = (\widetilde{S}_{\Xi_1} - K_1')^+. \tag{7.11}$$

This is immediate when $\widetilde{S}_{\Xi_1} \leq K_1'$; in the opposite case, note that $(\widetilde{S}_\tau - K_1')^+ = \widetilde{S}_\tau - K_1'$ when $\widetilde{S}_{\Xi_1} > K_1'$, and one can then use the martingale property of $\widetilde{S}_t$. Taking expectations in (7.11) yields from (7.10) that Ξ_1 must be the implied volatility of the call with strike price K_1.

Conditioning on $\mathcal{F}_{\Xi_2}$ is a little more complex. Suppose first that $\inf_{\Xi_1 \leq t \leq \Xi_2} \widetilde{S}_t > K_1'$. This is equivalent to $\tau > \Xi_2$, whence

$$\widetilde{E}[(\widetilde{S}_\tau - K_2')^+ \mid \mathcal{F}_{\Xi_2}] = (\widetilde{S}_{\Xi_2} - K_2')^+,$$

as in the previous argument (separate into the two cases $\widetilde{S}_{\Xi_2} \leq K_2'$ and $\widetilde{S}_{\Xi_2} > K_2'$). Hence, incorporating the case where $\tau \leq \Xi_2$, we find that

$$\widetilde{E}(\widetilde{S}_\tau - K_2')^+ = \widetilde{E}(\widetilde{S}_{\Xi_2 \wedge \tau} - K_2')^+,$$

thus showing that Ξ_2 can be obtained from $\widetilde{B}(S_0, \Xi_1^{\text{impl}}, \Xi_2, K_1', K_2') = V_0^{(2)}$. In consequence, from the left hand inequality in (7.9),

$$\begin{aligned} B(S_0, \Xi_1^{\text{impl}}, rT, K_2) &< \widetilde{B}(S_0, \Xi_1^{\text{impl}}, \Xi_2, K_1', K_2') \\ &= V_0^{(2)} = B(S_0, \Xi_2^{\text{impl}}, rT, K_2) \end{aligned}$$

Since, for call options, $B(S, \Xi, R, K_2)$ is increasing in Ξ, it follows that $\Xi_2^{\text{impl}} > \Xi_1^{\text{impl}}$.

Hence, under the assumption that $\Xi_1 < \Xi_2$, Algorithm 7.1 produces the right result.

The same arguments apply in the cases $\Xi_1 > \Xi_2$ and $\Xi_1 = \Xi_2$, in which cases, respectively, $\Xi_1^{\text{impl}} > \Xi_2^{\text{impl}}$ and $\Xi_1^{\text{impl}} = \Xi_2^{\text{impl}}$. Hence, also in these cases, Algorithm 7.1 provides the right solution.

Hence the solution to (7.10) is unique and is given by Algorithm 7.1.

The uniqueness of solution, combined with the above established fact that there are solutions in $[\Xi^-, \Xi^+]$, means that our solution must satisfy this constraint. Hence, the rightmost inequality in (7.7) must hold. The other inequality in (7.7) follows because the adjustment in (ii) increases the value of of the Ξ_i that is adjusted. This is because of the rightmost inequality in (7.9).

The result follows. ∎

An analytic expression for equation (7.5). To calculate the expression (7.5), note first that

$$\widetilde{B}(S, \Xi, \widetilde{\Xi}, K, \widetilde{K}) \;=\; \widetilde{E}[\widetilde{B}(S_{\Xi}, 0, \widetilde{\Xi} - \Xi, K, \widetilde{K}) | S_0 = S]$$

We therefore first concentrate on the expression for $\widetilde{B}(s, 0, T, K, \widetilde{K})$. For $K < \widetilde{K}$, this is the price of a down and out call, with strike $\widetilde{K}$, barrier K, and maturity T. We are still under the $\widetilde{P}$ distribution, in other words, $\sigma = 1$ and all interest rates are zero. The formula for this price is given on p. 462 in Hull (1997), and because of the unusual values of the parameters, one gets

$$\widetilde{B}(s, 0, T, K, \widetilde{K}) \;=\; \widetilde{E}((S_T - \widetilde{K})^+ | S_0 = s) \;-\; \frac{\widetilde{K}}{K}\widetilde{E}((S_T - H)^+ | S_0 = s) \;+\; \frac{\widetilde{K}}{K}(s - H)$$

for $s > K$, while the value is zero for $s \leq K$. Here, $H = K^2/\widetilde{K}$.

Now set

$$D(s, \Xi, \widetilde{\Xi}, K, X) \;=\; \widetilde{E}[(S_{\widetilde{\Xi}} - X)^+ I\{S_{\Xi} \geq K\} | S_0 = s]$$

and let BS_0 be the Black-Scholes formula for zero interest rate and unit volatility, $BS_0(s, \Xi, X) \;=\; \widetilde{E}[(S_{\Xi} - X)^+ | S_0 = s]$, in other words,

$$BS_0(s, \Xi, X) \;=\; s\Phi(d_1(s, X, \Xi)) \;-\; X\Phi(d_2(s, X, \Xi)), \tag{7.12}$$

where Φ is the cumulative standard normal distribution, and

$$d_i = d_i(s, X, \Xi) = (\log(s/X) \pm \Xi/2)/\sqrt{\Xi} \text{ where } \pm \text{ is } + \text{ for } i = 1 \text{ and } - \text{ for } i = 2. \tag{7.13}$$

Then, for $K < \widetilde{K}$,

$$\begin{aligned} \widetilde{B}(s, \Xi, \widetilde{\Xi}, K, \widetilde{K}) = D(s, \Xi, \widetilde{\Xi}, K, \widetilde{K}) \;&-\; \frac{\widetilde{K}}{K}D(s, \Xi, \widetilde{\Xi}, K, H) \;+\; \frac{\widetilde{K}}{K}BS_0(s, \Xi, K) \\ &+ (\widetilde{K} - K)\Phi(d_2(s, K, \Xi)). \end{aligned} \tag{7.14}$$

Similarly, for $K \geq \widetilde{K}$, a martingale argument and the formula on p. 463 in Hull (1997) gives that

$$\begin{aligned}
\widetilde{B}(s,0,T,K,\widetilde{K}) &= s - \widetilde{K} \ + \ \text{value of up and out put option with strike } \widetilde{K} \text{ and barrier } K \\
&= \widetilde{E}((S_T - \widetilde{K})^+ | S_0 = s) \ - \ \text{value of up and in put option with strike } \widetilde{K} \\
&\quad \text{and barrier } K \\
&= \widetilde{E}((S_T - \widetilde{K})^+ | S_0 = s) \ - \ \frac{\widetilde{K}}{K}\widetilde{E}((S_T - H)^+ | S_0 = s) \text{ for } s < K.
\end{aligned}$$

On the other hand, obviously, for $s \geq K$, $\widetilde{B}(s,0,T,K,\widetilde{K}) = (s - \widetilde{K})$ by a martingale argument.

Hence, for $K \geq \widetilde{K}$, we get

$$\begin{aligned}
\widetilde{B}(s,\Xi,\widetilde{\Xi},K,\widetilde{K}) &= BS_0(s,\widetilde{\Xi},\widetilde{K}) \ - \ \frac{\widetilde{K}}{K}BS_0(s,\widetilde{\Xi},H) - \ D(s,\Xi,\widetilde{\Xi},K,\widetilde{K}) \\
&\quad + \frac{\widetilde{K}}{K}D(s,\Xi,\widetilde{\Xi},K,H) \ + \ BS_0(s,\Xi,K) \ + \ (K - \widetilde{K})\Phi(d_2(s,K,\Xi)). \quad (7.15)
\end{aligned}$$

The formula for D is

$$D(s,\Xi,\widetilde{\Xi},K,X) = s\Phi(d_1(s,X,\widetilde{\Xi}),d_1(s,K,\Xi);A) \ - \ X\Phi(d_2(s,X,\widetilde{\Xi}),d_2(s,K,\Xi);A), \quad (7.16)$$

where

$$\Phi(x,y;A) \ = \ \text{cumulative bivariate normal c.d.f. with covariance matrix } A \quad (7.17)$$

and A is the matrix with diagonal elements 1 and off diagonal elements ρ,

$$\rho \ = \ \sqrt{\frac{\Xi}{\widetilde{\Xi}}}. \quad (7.18)$$

∎

Proof of Corollary 7.3. It is easy to see that Theorem 7.2 goes through with $K_1 = K_2$ (in the case where the implied volatilities are the same). Using formula (7.7), we get from Algorithm 7.1 that

$$A((s - K_0)^+; (s - K_1)^+) = \widetilde{C}(S_0, \Xi_1^{\mathrm{impl}}, \Xi^+, K_1', K_1'). \quad (7.19)$$

The result then follows by replacing "0" by "2" in (7.19). ∎

8. Bounds that are not based on prediction sets. It may seem odd to argue, as we have in Section 5.3, for an approach that uses different models for inference and trading, even if the first is nested in the other. To see it in context, recall that we referred to this procedure as the *decoupled prediction approach*. Now consider two alternative devices. One is a *consistent prediction approach*: use the prediction region obtained above, but also insist for purposes of trading that $P \in \Theta$. Another alternative would be to find a *confidence or credible set* $\hat{\Theta} \subseteq \Theta$, and then do a super-replication that is valid for all $P \in \hat{\Theta}$. The starting values for these schemes are considered below.

Table 4 suggests the operation of the three schemes.

TABLE 4

Three approaches for going from data to hedging strategies

approach	product of statistical analysis	hedging is valid and solvent for
confidence or credible sets	set $\hat{\Theta}$ of probabilities	probabilities in $\hat{\Theta}$
consistent prediction set method	set C of possible outcomes	probabilities in Θ outcomes in C
decoupled prediction set method	set C of possible outcomes	probabilities in $\mathcal{Q}$ outcomes in C

Θ is the parameter space used in the statistical analysis, which can be parametric or nonparametric. $\mathcal{Q}$ is the set of distributions defined in Assumption (A). C is a prediction set, and $\hat{\Theta}$ is a confidence or credible set.

The advantages of the decoupled prediction set approach are the following. First, transparency. It is easy to monitor, *en route*, how good the set is. For example, in the case of (2.3), one can at any time t see how far the realized $\int_0^t \sigma_u^2 du$ (or, rather, (3.15)) is from the prediction limits Ξ^- and Ξ^+. This makes it easy for both traders and regulators to anticipate any disasters, and, if possible, to take appropriate action (such as liquidating the book).

Second, the transparency of the procedure makes this approach ideal as an exit strategy when other schemes have gone wrong. This can be seen from the discussion in Section 2.3.

Thirdly, and perhaps most importantly, the decoupling of the inferential and trading models respects how these two activities are normally carried out. The statistician's mandate is, usually, to find a model Θ, and to estimate parameters, on the basis of whether these reasonably fit the data. This is different from finding a probability distribution that works well for trading. For example, consider modeling interest rates with an Ornstein-Uhlenbeck process. In many cases, this will give a perfectly valid fit to the data. For trading purposes, however, this model has severe drawbacks, as outlined in Section 6.2 above.

With the decoupling of the two stages, therefore, the statistical process can concentrate on good inference, without worrying about the consequences of the model on trading. For inference, one can use existing literature, on ARCH/GARCH or a variety of SDE type models. References include Aït-Sahalia (1996, 2002), Aït-Sahalia and Mykland (2003), Andersen (2000), Andersen, Bollerslev, Diebold, and Labys (2001), Barndorff-Nielsen and Shephard (2001), Bibby and Sørensen (1995, 1996a,b), Bollerslev, Chou, and Kroner (1992), Dacunha-Castelle and Florens-Zmirou (1986), Danielsson (1994), Florens-Zmirou (1993), Genon-Catalot and Jacod (1994), Genon-Catalot, Jeantheau and Laredo (1999, 2000), Hansen and Scheinkman (1995), Hansen, Scheinkman and Touzi (1998), Jacod (2000), Jacod and Protter (1998), Jacquier, Polson and Rossi (1994), Kessler and Sørensen (1999), Küchler and Sørensen (1998), Lo (1987), and Zhang (2001). This is, of course, only a small sample of the literature available.

To sum up, the decoupled prediction set approach is, in several ways, robust.

But is it efficient? The other two approaches, by using the model Θ for both stages, would seem to give rise to lower starting values A_0, just by being consistent and by using a smaller family Θ for trading. We have not investigated this question in any depth, but tentative evidence suggests that the consistent prediction approach will yield a cheaper A_0, while the confidence/credible approach is less predictable in this respect. Consider the following.

Using Kramkov (1996) and Mykland (2000), one can obtain the starting value for a true super-replication over a confidence/credible set $\hat{\Theta}$ for conditional probabilities P_ω. Assume the nesting condition. Let $\hat{\Theta}^*$ be the convex hull of distributions $Q^* \in \mathcal{Q}^*$ for which Q^* is mutually absolutely continuous with a $P_\omega \in \hat{\Theta}$. The starting value for the super-replication would then normally have the form

$$A_0 = \sup\{E^*(\eta^*) \ : \ P^* \in \hat{\Theta}^*\}.$$

Whether this A_0 is cheaper than the one from (5.3) may, therefore, vary according to Θ and to the data. This is because $\hat{\Theta}^*$, and $\mathcal{P}^* = \mathcal{P}^*_{S_0}$ from (5.2), are not nested one in the other, either way.

For the consistent prediction approach, we have not investigated how one can obtain a result like Theorem 5.1 for subsets of $\mathcal{Q}$, so we do not have an explicit expression for A_0. However, the infimum in (2.5) is with respect to a smaller class of probabilities, and hence a larger class of super-replications on C. The resulting price, therefore, can be expected to be smaller than the conservative ask price from (5.2). As outlined above, however, this approach is not as robust as the one we have been advocating.

Acknowledgements. I would like to thank Marco Avellaneda, and the editors and referees on my 2000 and 2003 papers, and everyone who has helped me understand this area.

REFERENCES

Aït-Sahalia, Y. (1996). Nonparametric Pricing of Interest Rate Derivative Securities. *Econometrica* **64** 527-560.

Aït-Sahalia, Y. (2002). Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. *Econometrica* 70, 223–262.

Aït-Sahalia, Y., and Mykland, P.A. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* **71** 483–549.

Andersen, T.G. (2000). Some reflections on analysis of high frequency data, *J. Bus. Econ. Statist.* **18** 146-153.

Andersen, T.G., Bollerslev, T., Diebold, F., and Labys, P. (2001). The distribution of realized exchange rate volatility. To appear in *J. Amer. Statist. Assoc.*.

Artzner, Ph., Delbaen, F. Eber, J.M., and Heath, D. (1999). Coherent measures of risk, *Math. Finance* **9** 203-228.

Ash, R.B. (1972). *Real Analysis and Probability.* Academic Press, New York.

Asmussen, S., Glynn, P., Pitman, J.: Discretization error in simulation of one-dimensional reflecting Brownian motion. Ann. Appl. Probab. **5** 875–896 (1995)

Avellaneda, M., Levy, A., and Paras, A. (1995). Pricing and hedging derivative securities in markets with uncertain volatilities, *Appl. Math. Finance* **2** 73-88.

Barenblatt, G.I. (1978). *Similarity, Self-similarity and Intermediate Asymptotics* (Consultants Bureau, New York).

Barndorff-Nielsen, O.E., and Cox, D.R. (1996). Prediction and asymptotics, *Bernoulli* **2** 319-340.

Barndorff-Nielsen, O.E., and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based modes and some of their uses in financial economics. *J. Roy. Stat. Soc. B* **63** no. 2.

Beckers, S. (1981). Standard deviations implied in option prices as predictors of future stock price variability, *J. Banking and Finance* **5** 363–382.

Bergman, Y.Z. (1995). Options pricing with differential interest rates. Rev. Financial Studies **8** 475–500.

Bergman, Y.Z., Grundy, B.D., and Wiener, Z. (1996). General properties of option prices, *J. Finance* **5** 1573-1610.

Bibby, B.M., and Sørensen, M. (1995). Martingale estimating functions for discretely observed diffusion processes. *Bernoulli* **1** 17-39.

Bibby, B.M., and Sørensen, M. (1996a). On estimation for discretely observed diffusions: a review. *Ther. Stochastic Proc.* **2** 49-56.

Bibby, B.M., and Sørensen, M. (1996b). A hyperbolic diffusion model for stock prices. *Finance and Stochastics* **1** 25-41.

Bick, A., and Reisman, H. (1993). Generalized implied volatility. (Preprint).

Bick, A., and Willinger, W. (1994). Dynamic spanning without probabilities, *Stoch. Aprox. Appl.* **50** 349-374.

Black, F., and Scholes, M. (1973). The pricing of options and corporate liabilities, *J. Polit. Econ.* **81** 637-654.

Bollerslev, T., Chou, R.Y., and Kroner, K.F. (1992). ARCH modeling in Finance, *J. Econometrics* **52** 5-59.

Carr, P., Geman, H., and Madan, D. (2001). Pricing and hedging in incomplete markets, forthcoming.

Constantinides, G.M., and Zariphopoulou, T. (1999). Bounds on prices of contingent claims in an intertemporal economy with proportional transaction cost and general preferences. *Finance and Stochastics* **3** 345–369.

Constantinides, G.M., and Zariphopoulou, T. (2001). Bounds on derivative prices in an intertemporal setting with proportional transaction cost and multiple securities. *Mathematical Finance* **11** 331–346.

Cox, J., Ingersoll, J. and Ross, S. (1985). A theory of the term structure of interest rates, *Econometrica* **53** 385-408.

Cvitanić, J., and Karatzas, I. (1992). Convex duality in constrained portfolio optimization, *Ann. Appl. Probab.* **2** 767–818.

Cvitanić, J., and Karatzas, I. (1993). Hedging of contingent claims with constrained portfolios, *Ann. Appl. Probab.* **3** 652-681.

Cvitanić, J., and Karatzas, I. (1999). On Dynamic Measures of Risk, *Finance and Stochastics* **3** 451–482.

Cvitanić, J., Pham, H., and Touzi, N. (1998). A closed form solution to the problem of super-replication under transaction cost, *Finance and Stochastics* **3** 35-54.

Cvitanić, J., Pham, H., and Touzi, N. (1999). Super-replication in Stochastic Volatility Models under Portfolio Constraints *Appl. Probab. J.* **36** 523–545.

Dacunha-Castelle, D. & Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations, *Stochastics* **19** 263–284.

Dambis, K. (1965). On the decomposition of continuous sub-martingales, *Theory Probab. Appl.* **10** 401–410.

Danielsson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood, *J. Econometrics* **64** 375-400.

Delbaen, F., and Schachermayer, W. (1995). The existence of absolutely continuous local martingale measure, *Ann. Appl. Probab.* **5** 926–945.

Delbaen, F., and Schachermayer, W. (1996). The variance-optimal martingale measure for continuous processes. *Bernoulli* **2** 81-105.

Delbaen, F., Monat, P., Schachermayer, W., Schweizer, M. and Stricker, C. (1997). Weighted norm inequalities and hedging in incomplete markets. *Finance and Stochastics* **1** 181-227.

Dellacherie, C., and Meyer, P.A. (1978). *Probabilities and Potential* (North Holland).

Dubins, L.E., and Schwartz, G. (1965). On continuous martingales, *Proc. Nat. Acad. Sci. U.S.A.* **53** 913–916.

Duffie, D. (1996). *Dynamic Asset Pricing Theory* (2nd ed.) (Princeton).

Eberlein, E., and Jacod, J. (1997). On the range of options prices, *Finance and Stoch.* **1** 131-140.

El Karoui, N., Jeanblanc-Picqué, M., and Shreve, S.E. (1998). Robustness of the Black and Scholes formula, *Math. Finance* **8** 93-126.

El Karoui, N. and Quenez, M.-C. (1995). Dynamic programming and pricing of contingent claims in an incomplete market, *SIAM J. Contr. Opt.* **33** 29-66.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations, *J. Appl. Prob.* **30** 790-804.

Föllmer, H. (1979). Calcul d'Itô sans probabilités, *Seminaire de Probabilités XV*, Lect. Notes in Math. (Springer-Verlag).

Föllmer, H. and Schied, A. (2002). *Stochastic Finance: An Introduction in Discrete Time* (Berlin, deGruyter Studies in Mathematics).

Föllmer, H., Leukert, P. (1999). Quantile Hedging. *Finance and Stochastics* **3** 251–273.

Föllmer, H., Leukert, P. (2000). Efficient hedging: Cost versus shortfall risk. *Finance and Stochastics* **4** 117–146.

Föllmer, H. and Schweizer, M. (1991). Hedging of Contingent Claims under Incomplete Information, in in *Applied Stochastic Analysis*, eds. M.H.A. Davis and R.J. Elliot (Gordon and Breach).

Föllmer, H. and Sondermann, D. (1986). Hedging of non-redundant contingent claims, in Hildebrand, W. and Mas-Colell, A. (eds) *Contributions to Mathematical Economics*, p. 205-223.

Foster, D.P., and Nelson, D.B. (1996). Continuous record asymptotics for rolling sample variance estimators, *Econometrica* **64** 139-74.

Frey, R.. (2000). Superreplication in stochastic volatility models and optimal stopping. *Finance and Stochastics* **4** 161–188.

Frey, R., and Sin, C. (1999). Bounds on European options prices under stochastic volatility. Math. Finance **9** 97–116.

Friedman, C. (2000). Confronting model misspecification in finance: Tractable collections of scenario probability measures for robust financial optimization problems (preprint).

Fritelli, M. (2000). Intrduction to a theory of value coherent with the no-arbitrage principle. *Finance and Stochastics* **4** 275–297.

Geman, H., Yor, M. (1996). Pricing and hedging double-barrier options: A probabilistic approach. *Math. Finance* **6** 365-378.

Genon-Catalot, V., and Jacod, J. (1994). Estimation of the diffusion-coefrficient for diffusion-processes – random sampling *Scand. J. of Statist.* **21** 193–221.

Genon-Catalot, V., Jeantheau, T., and Laredo, C. (1999). Parameter estimation for discretely observed stochastic volatility models *Bernoulli* **5** 855-872.

Genon-Catalot, V., Jeantheau, T., and Laredo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications *Bernoulli* **6** 1051-1079.

Hansen, L.P., and Scheinkman, J.A. (1995). Back to the future: Generating moment implications for continuous-time Markov processes, *Econometrica* **63** 767-804.

Hansen, L.P., Scheinkman, J.A., and Touzi, N. (1996). Identification and imbeddability of Markov diffusion models from discrete time data, preprint.

Harrison, J.M., and Kreps, D.M. (1979). Martingales and arbitrage in multiperiod securities markets, *J. Econ. Theory* **20** 381-408.

Harrison, J. M., and Pliskà, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading, *Stoc. Proc. Appl.* **11** 215- 260.

Heath, D., and Ku, H. (2001). Equilibrium in incomplete markets, preprint,

Hobson, D.G. (1998). Volatility misspecification, option pricing and superreplication via coupling. *Ann. Appl. Probab.* **8** 193–205.

Hofmann, N., Platen, E., and Schweizer, M. (1992). Option pricing under incompleteness and stochastic volatility, *Math. Finance* **2** 153-187.

Hull, J.C. (1999). *Options, Futures and Other Derivatives* (4rd Ed.). (Prentice Hall).

Jacod, J. (1979). *Calcul Stochastique et Problèmes de Martingales*, Lect. N. Math. 714 (Springer-Verlag).

Jacod, J. (2000). Non-parametric kernel estimation of the coefficient of a diffusion *Scand. J. of Statist.* **27** 83–96

Jacod, J. and Protter, P.(1998). Asymptotic error distributions for the Euler method for stochastic differential equations, *Annals of Probability* **26** 267-307.

Jacod, J., and Shiryaev, A. N. (1987). *Limit Theory for Stochastic Processes*, (Springer-Verlag).

Jacquier, E., Polson, N. and Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models *J. Bus. Econ. Statist.* **12** 371–389.

Karatzas, I. (1988). On the pricing of American options. Appl. Math. Optim. **17** 37–60 (1988)

Karatzas, I. (1996). *Lectures on the Mathematics of Finance* (CRM monograph series).

Karatzas, I., and Kou, S.G. (1996). On the pricing of contingent claims under constraints, *Ann. Appl. Probab.* **6** 321–369.

Karatzas, I., and Kou, S.G. (1998). Hedging American contingent claims with constrained portfolios, *Finance and Stochastics* **2** 215-258.

Karatzas, I., and Shreve, S.E. (1991). *Brownian Motion and Stochastic Calculus* (2nd Ed.) (Springer-Verlag).

Kessler, M., and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process, *Bernoulli* **5** 299-314.

Kramkov, D.O. (1996). Optional decompositions of supermartingales and hedging in incomplete security markets. *Probab. Theory Relat. Fields.* **105** 459-479.

Küchler, U. and Sørensen, M. (1998). *Exponential Families of Stochastic Processes.* Springer-Verlag.

Laurent, J.P., and Pham, H. (1999). Dynamic programming and mean-variance hedging, *Finance Stochast* **3** 83-110.

Lo, A.W. (1987). Semi-parametric upper bounds for option prices and expected payoffs, *J. Financial Economics* **19** 373-387.

Lyons, T.J. (1995). Uncertain volatility and the risk-free synthesis of derivatives. *Appl. Math. Finance* **2** 117–133.

Merton, R.C. (1973). Theory of rational options pricing, *Bell J. Econ. and Measurement Sci.* **4** (Spring) 141-183.

Mykland, P.A. (2000). Conservative delta hedging, *Ann. Appl. Probab.* **10** 664–683.

Mykland, P.A. (2003a). The interpolation of options *Finance and Stochastics* **7** 417–432.

Mykland, P.A. (2003b). Financial options and statistical prediction intervals (to appear in *Ann. Statist.* **31** # 5)

Mykland, P.A. (2003c). Statistics and the worst case state price distribution (in preparation).

Mykland, P.A., and Zhang, L. (2002a). Inference for volatility-type objects and implications for hedging. Technical report no. 506, Department of Statistics, University of Chicago.

Mykland, P.A., and Zhang, L. (2002b). ANOVA for diffusions, TR 507, Dept. of Statistics, The University of Chicago.

Myneni, R. (1992). The pricing of the American option, *Ann. Appl. Probab.* **2** 1-23.

Pollard, D. (1984). *Convergence of Stochastic Processes* (Springer-Verlag, New York).

Pham, H., Rheinländer, T., and Schweizer, M. (1998). Mean-variance hedging for continuous processes: New proofs and examples, *Finance and Stochastics* **2** 173-198.

Schweizer, M. (1990). Risk-minimality and orthogonality of martingales *Stochastics* **30** 123-131.

Schweizer, M. (1991). Option hedging for semimartingales *Stoch. Proc. Appl.* **37** 339- 363.

Schweizer, M. (1992). Mean-variance hedging for general claims *Ann. Appl. Prob.* **2** 171-179.

Schweizer, M. (1993). Semimartingales and hedging in incomplete markets *Th. Prob. Appl.* **37** 169- 171.

Schweizer, M. (1994). Approximating random variables by stochastic integrals, *Ann. Probab.* **22** 1536-1575.

Simonsen, K.K.: Simulating First Passage Times and the Maximum of Stochastic Differential Equations: An Error Analysis. Ph.D. dissertation, Department of Statistics, The University of Chicago, 1997

Smith, R.L.. (1999). Bayesian and frequentist approaches to parametric predictive inference (with discussion). In: *Bayesian Statistics* **6** (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds.), Oxford University Press, pp. 589-612.

Spivak, G., and Cvitanić (1998). Maximizing the probability of a perfect hedge. *Ann. Appl. Probab.* **9** 1303–1328.

Strook, D.W., and Varadhan, S.R.S. (1979). *Multidimensional Diffusion Processes.* (New York. Springer.)

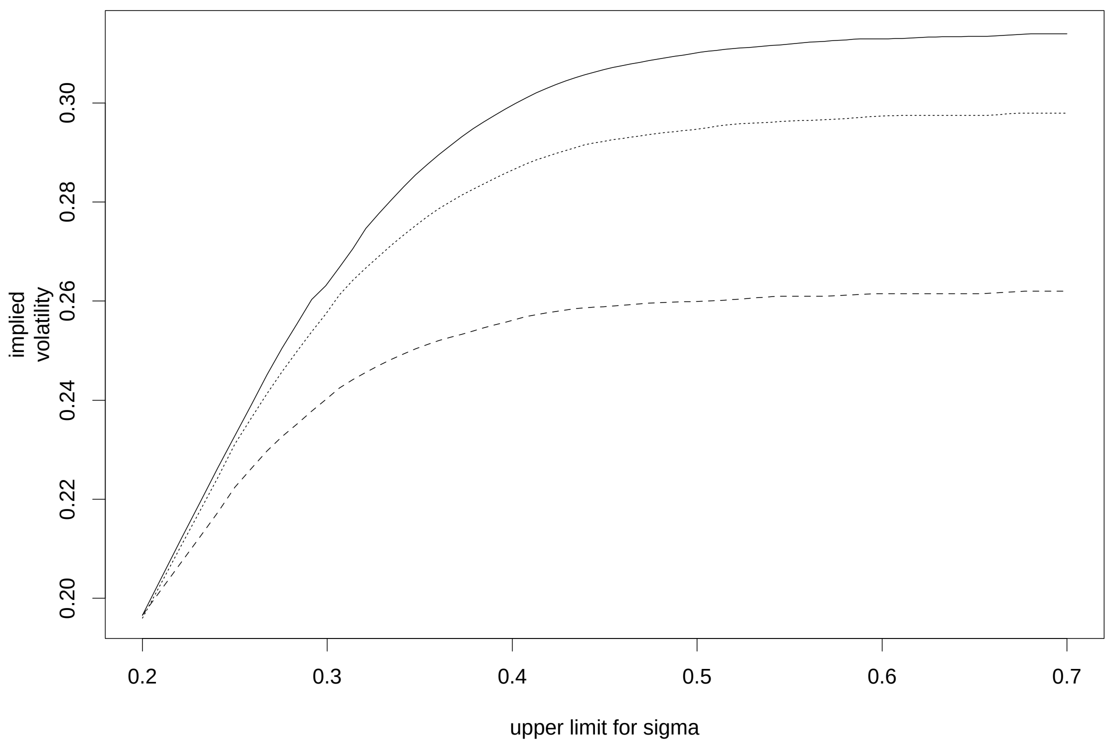
Vasicek, O.A. (1977). An equilibrium characterization of the term structure, *J. Financial Economics* **5** 177-188.

Willinger, W., and Taqqu, M.S. (1991). Toward a convergence theory for continuous stochastic securities market models, *Math. Finance* **1** no. 1, p. 55–99.

Zhang, L. (2001). From martingales to ANOVA: Implied and realized volatility. Ph.D. dissertation, Department of Statistics, University of Chicago.

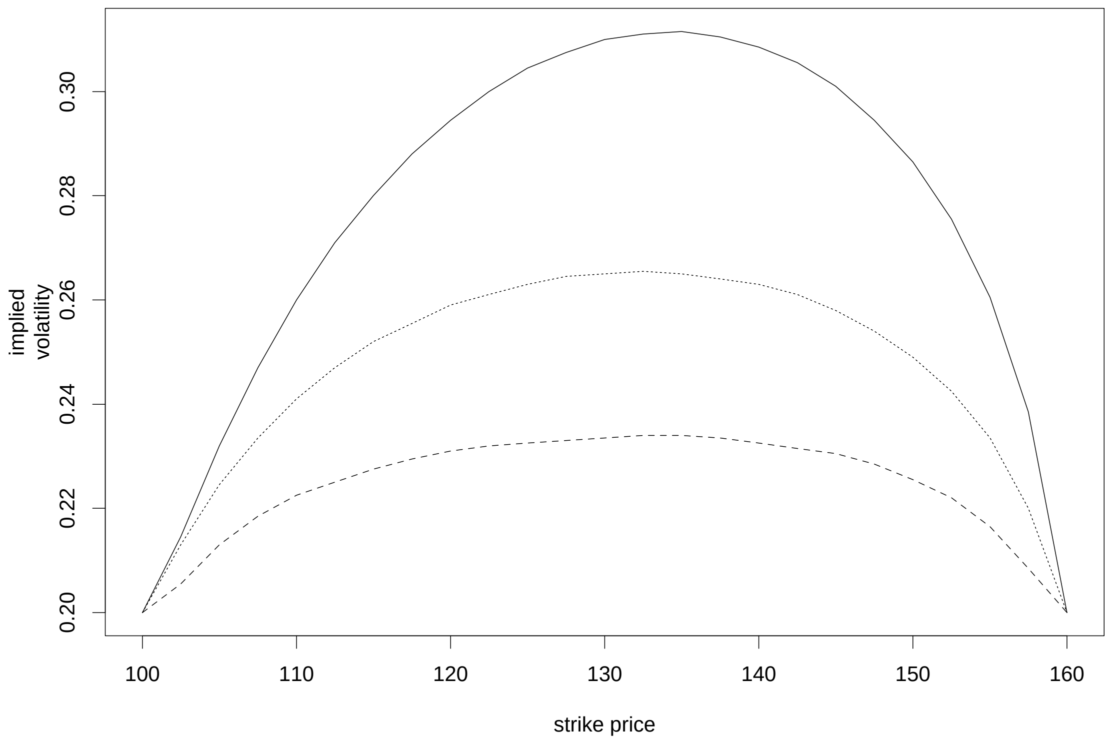
Department of Statistics
University of Chicago
5734 University Avenue
Chicago, Illinois 60637

FIGURE 7.1. Effect of interpolation: Implied volatilities for interpolated call options as a function of the upper limit of the prediction interval.



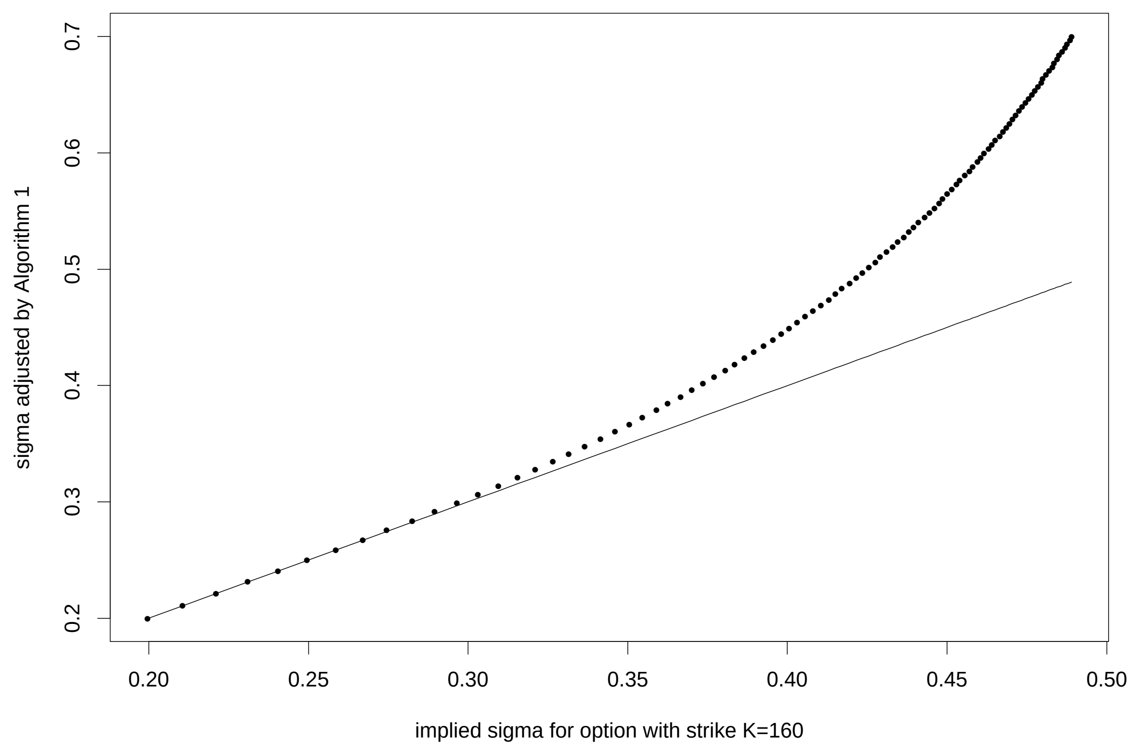
We consider various choices of strike price K_0 (from top to bottom: K_0 is 130, 120 and 110) for the option to be interpolated. The options that are market traded have strike prices $K_1 = 100$ and $K_2 = 160$. The graph shows the implied volatility of the options price A (σ_{impl} given by $B(S_0, \sigma^2_{\text{impl}}, rT, K_0) = A$ as a function of $\sqrt{\Xi^+}$. We are using square roots as this is the customary reporting form. The other values defining the graph are $S_0 = 100$, $T = 1$ and $r = 0.05$, and $\sqrt{\Xi_1^{\text{impl}}} = \sqrt{\Xi_2^{\text{impl}}} = .2$. The asymptotic value of each curve corresponds to the Merton bound for that volatility.

FIGURE 7.2. Effect of interpolation: implied volatilities for interpolated call options as a function of the strike price K_0 for the option to be interpolated.



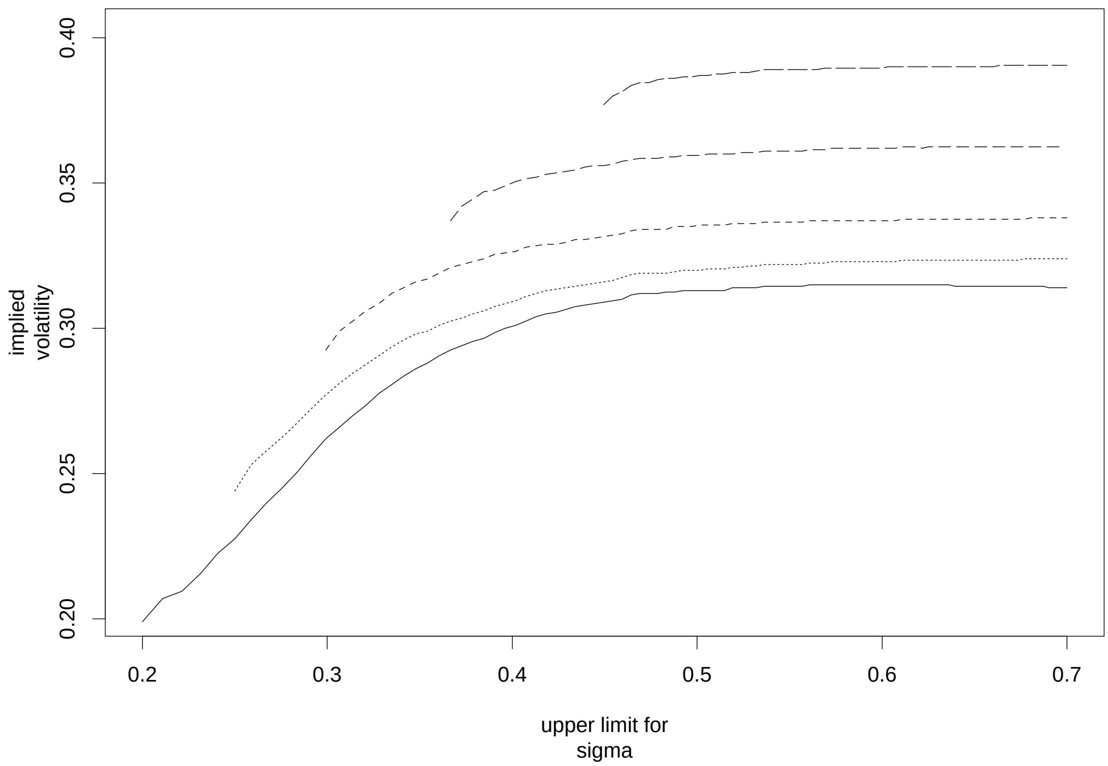
We consider various choices of maximal volatility values $\sqrt{\Xi^+}$ (from top to bottom: $\sqrt{\Xi^+}$ is .50, .40 and .25). Other quantities are as in Figure 7.1. Note that the curve for $\sqrt{\Xi^+} = .50$ is graphically indistinguishable from that of the Merton bound.

FIGURE 7.3. $\widetilde{C}$: $\sqrt{\Xi_2}$ as a function of $\sqrt{\Xi_2^{\mathrm{impl}}}$, for fixed $\sqrt{\Xi_1^{\mathrm{impl}}} = \sqrt{\Xi_2} = 0.2$.



A diagonal line is added to highlight the functional relationship.

FIGURE 7.4. Implied volatility for interpolated call option with strike price $K_0 = 140$, as the upper bound $\sqrt{\Xi^+}$ varies.



The curves assume $\sqrt{\Xi_1^{\text{impl}}} = 0.2$ and, in ascending order, correspond to $\sqrt{\Xi_2^{\text{impl}}} = 0.2,\ 0.25,$ $0.3,\ 0.35$ and 0.4. The starting point for each curve is the value $\sqrt{\Xi^+}$ (on the x axis) so that the no-arbitrage condition of Corollary 7.3 is not violated. As in Figure 7.1, the asymptotic value of each curve corresponds to the Merton bound for that volatility.