The Annals of Statistics
2003, Vol. 31, No. 5, 1413–1438

FINANCIAL OPTIONS AND STATISTICAL PREDICTION INTERVALS[1]

BY PER ASLAK MYKLAND

University of Chicago

The paper shows how to convert statistical prediction sets into worst case hedging strategies for derivative securities. The prediction sets can, in particular, be ones for volatilities and correlations of the underlying securities, and for interest rates. This permits a transfer of statistical conclusions into prices for options and similar financial instruments. A prime feature of our results is that one can construct the trading strategy as if the prediction set had a 100% probability. If, in fact, the set has probability $1-\alpha$, the hedging strategy will work with at least the same probability. Different types of prediction regions are considered. The starting value A_0 for the trading strategy corresponding to the $1-\alpha$ prediction region is a form of long term value at risk. At the same time, A_0 is coherent.

1. Introduction. The usual setting of options theory concerns a derivative security whose final payoff η is a function of the values of underlying market traded securities $S_t^{(1)}, \dots, S_t^{(p)}$. Most theory for setting the prices and trading strategies associated with such a setup is based on knowing the probability distribution P of the underlying securities. See, for example, Duffie (1996) and Hull (1999) for comprehensive accounts. The main device is to create a portfolio in $S_t^{(1)}, \dots, S_t^{(p)}$, with value V_t at time t, so that at maturity T, $V_T = \eta$. For simplicity, we here take T to be nonrandom. One is allowed to change the composition of the portfolio at any time, but at times of such adjustment, the total portfolio value must remain unchanged. This is what is called a self financing portfolio or trading strategy.

The question of what happens when P is unknown, however, is not fully resolved. The existing body of work would mostly appear to fall into two categories: (i) "super-hedging" or "-replication" when P is part of a class, such as a confidence set, of probability distributions, and (ii) reduction to a single probability distribution.

The first of these approaches also involves creating a trading portfolio, but now we require $V_T \geq \eta$ a.s., for all probability distributions in the relevant class. In other words, the institution that sold the option η is required to cover its liability

Received August 1998; revised March 2002.

[1]Supported in part by NSF Grants DMS-99-71738, DMS 96-26266 and DMS 97-03918, and Army Research Office Grant DAAH04-95-1-0105.

AMS 2000 subject classifications. Primary 62F25, 62G15, 62P05; secondary 60G44, 62M99.

Key words and phrases. Analytic sets, coherent measures of risk, conservative delta hedging, incompleteness, prediction, statistical uncertainty, value at risk.

always. The portfolio must also be self financing, but now we include in this concept the possibility that funds may be removed from the portfolio over time when new information shows that the payoff can still be covered. Funds cannot, however, be added; that would negate the concept of super-replication as a way of making sure that η can be paid.

There is a fine line between what is conventionally called "super-hedging," where one tries to minimize the value of V_0, and what is usually referred to as "robustness." In the latter case, one does not try to optimize V_0, but instead one takes a reasonable strategy and sees when it will cover the final liability. Work in the former direction includes Avellaneda, Levy and Paras (1995), Lyons (1995) and Mykland (2000).

Papers focusing on the latter include Bergman, Grundy and Wiener (1996), El Karoui, Jeanblanc-Picqué and Shreve (1998) and Hobson (1998).

The main way of reducing the problem to a known P would be by considering a Bayesian posterior. To further pursue the problem in this direction, one can then go to the "known P but unknown P^*" literature, which is substantial. The symbol P^* denotes the so-called risk neutral or equivalent martingale measure, as opposed to the actual probability distribution P. The measure P^*, though not something one can fully find by statistical methods, is crucial for valuing options. The concept also comes up in this paper. For a further discussion of the distinction between unknown P and unknown P^*, see the introduction to Mykland (2000).

Studies that would appear to fall outside this categorization are Artzner, Delbaen, Eber and Heath (1999) and Cvitanić and Karatzas (1999).

From a statistical point of view, therefore, there are, at least in some cases, solutions to the question of how one would hedge with either confidence regions or Bayesian posteriors. We shall argue, however, that it is more natural to hedge based on prediction sets (either frequentist or Bayesian) rather than on the two other statistical objects mentioned. A main reason for this is that prediction sets would seem to afford greater transparency of trading, and this is discussed further in Section 4.1. Prediction sets also permit a straightforward exit strategy, as discussed in Section 6.

Another advantage of prediction sets, however, is that the main known results under (i) are, in fact, for regions that are more naturally seen as prediction sets rather than confidence sets. To illustrate this point, consider the main such set encountered in Mykland (2000). This is of the form

$$\Xi^- \le \int_0^T \sigma_t^2 \, dt \le \Xi^+, \tag{1.1}$$

where σ_t is the realized volatility of a single stock S_t. This is to say that the stock price follows the diffusion

$$dS_t = m_t S_t \, dt + \sigma_t S_t \, dB_t, \tag{1.2}$$

where m and σ are unknown random processes and B is a standard Brownian motion. It is not all that natural, however, to assume that (1.1) holds with probability 1, as would be the case if this were the definition of a confidence set. It is more reasonable to think of (1.1) as a prediction region, say C, that has a certain prediction probability $1 - \alpha$. The limits in (1.1) could then be estimated from historical data or found with the help of calibration. In our setting, therefore, the interval limits in (1.1) can be either nonrandom, or they can be estimated from information available at time zero. The special issues that pertain to the latter case are explored in the first part of Section 4.

The same considerations apply to the classes of distributions discussed in Avellaneda, Levy and Paras (1995) and Lyons (1995). The former is discussed further in Section 3.1.

The main result in this paper is that one can, in fact, use regions like (1.1) as prediction intervals. Specifically, we show in Section 2 that superhedging strategies for $1 - \alpha$ probability prediction regions can be found by first considering regions that have probability 1. Hence the problem reduces to that considered in the papers cited in Mykland (2000) in connection with (i).

Prior work directly applicable to the prediction set problem includes Föllmer (1981) and Bick and Willinger (1994). The relation to the present work is discussed in Section 3.2.

Section 3 considers two main resulting examples of prediction intervals and also gives general criteria for what types of sets can be used. A modified version of our general theorem is stated as a corollary to this discussion. Section 4.1 considers statistically based regions, and also the advantages of prediction sets from a transparency point of view.

Our results permit the incorporation of many existing econometric methods into hedging via prediction intervals. Some of the literature is cited in Section 4. Also, an example of how this can be implemented is carried out in Section 4.2.

Section 5 discusses strategies for handling interest rate uncertainty. For one of these approaches, Section 5.3 provides a moderately explicit answer for general European options in the presence of an interest rate adjusted interval similar to (1.1). Section 6 describes how the superhedging method can be used as an exit strategy even if one does not wish to use it as a primary tool for hedging. Finally, proofs are contained in Section 7.

Note that the procedure we give in this paper extends to the case where one also hedges in market traded options, as, for example, in Mykland (2003).

2. Options hedging from prediction sets.

We are concerned with continuous processes $S_t^{(1)}, \ldots, S_t^{(p)}$, which are the prices of traded securities paying no dividends. r_t is an adapted process representing the risk free interest rate, and $\beta_t = \exp\{\int_0^t r_u \, du\}$ is the value at time t of one unit of currency deposited in the money market at time 0. In the account in Section 1, we just let β_t be one of the $S_t^{(i)}$'s, but here it is useful to keep it separate. Note that the investment in β_t can be negative; in practice this is achieved by using loans where the collateral is so substantial and liquid that the loan can be considered risk free. [See "Repo rate" on page 88 of Hull (1999).]

We assume in the following that little is known about the probability distribution P governing our system. We suppose that P is an element of a large class of probability distributions $\mathcal{Q}$, to be defined below. Our main information is that we have a prediction set C which we assume will occur with high probability; C can, for example, set limits on the cumulative volatility and interest rate. An example would be (1.1); the following sections provide a more thorough treatment of how to do this.

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. The proper definitions are given in Section 4.1. 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 proposing to solve is as follows. We have to cover a liability η at a nonrandom time T. Because of the size of the set $\mathcal{Q}$, a full super-replication (that works with probability 1 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 (Bayesian or frequentist) probability $1-\alpha$.

In analogy with superhedging, we do this by setting up a self-financing portfolio in $S_t^{(1)}, \dots, S_t^{(p)}$ and in β_t. We denote by V_t the value of this portfolio. In order to achieve the desired level of solvency, we require that $V_T \geq \eta$, $\mathcal{Q}$-a.s. *so long as C occurs*. By this statement, we mean $V_T \geq \eta$, on C, Q-a.s., for all $Q \in \mathcal{Q}$. This will be called a *super-replication of η on C*. The specific technical definition is like the one for a full super-replication given on the top of page 670 in Mykland (2000), except that solvency only needs to hold on C. Also, the set $\mathcal{Q}$ replaces the set $\mathcal{P}$.

Since C has probability at least $1-\alpha$, this solves the problem of guaranteeing solvency with this same probability.

We want to go beyond this, however: we would like the cheapest such super-replication.

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

$$A_0 = \inf\{V_0 : (V_t) \text{ is a super-replication on } C \text{ of the liability } \eta\}. \tag{2.1}$$

Note that in the following, V_t denotes the portfolio value of any super-replication, while A_t is the cheapest one (across $Q \in \mathcal{Q}$), provided it exists. Both are denominated in the same currency as $S_t^{(1)}, \dots, S_t^{(p)}$ and β_t, such as dollars or yen.

So what is A_0? To proceed further, we need to set up the system, and to define $\mathcal{Q}$:

ASSUMPTIONS A (System assumptions). Our probability space is the set $\Omega = \mathbb{C}[0, T]^{p+1}$, and we let $(\beta_t, S_t^{(1)}, \dots, S_t^{(p)})$ 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 nonrandom 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 bounded (above and below) by a nonrandom constant, P^*-a.s. As usual, "[,]" is the quadratic variation of the process [see pages 51 and 52 of Jacod and Shiryaev (1987), and also the discussion after formula (3.4) at the end of (our) 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 filtration containing $(\mathcal{B}_{t+})$ and all sets in $\mathcal{N}$, given by

$$\mathcal{N} = \{F \subseteq \Omega : \forall P^* \in \mathcal{Q}^* \ \exists E \varepsilon \mathcal{B} : F \subseteq E \text{ and } P^*(E) = 0\}, \tag{2.2}$$

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 $s_0 = (s_0^{(1)}, \dots, s_0^{(p)})$, we write $\mathcal{Q}_{s_0}$.

REMARK 2.1. An important fact is that $\mathcal{F}_t$ is analytic for all t, by Theorem III.10 [page 42 in Dellacherie and Meyer (1978)]. Also, the filtration $(\mathcal{F}_t)$ is right continuous by construction. $\mathcal{F}_0$ is a noninformative (trivial) σ-field. The relationship of $\mathcal{F}_0$ to information from the past (before time zero) is established in Section 4.1.

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 (1.2) 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, e.g., the version of Girsanov's theorem given in Karatzas and Shreve (1991), Theorem 3.5.1]. Also, if these requirements fail, the $S^{(i)*}$'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 (1.2).

A typical form of the prediction set C would be (1.1) and/or $R^- \leq \int_0^T r_t \, dt \leq R^+$. 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 (1.1). See the development in Sections 3.2 and 4.1.

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

$$\mathcal{P}^* = \{P^* \in \mathcal{Q}^* : P^*(C) = 1\}. \tag{2.3}$$

The price is then, from Theorem 2.1 below,

$$A_0 = \sup\{E^*(\eta^*) : P^* \varepsilon \mathcal{P}^*\}, \tag{2.4}$$

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

$$\eta^* = \exp\left\{-\int_0^T r_u \, du\right\}\eta. \tag{2.5}$$

The quantity A_0 is therefore a form of "value at risk" [see Chapter 14 (pages 342–365) of Hull (1999)] that is based on dynamic trading. At the same time, A_0 is coherent in the sense of Artzner, Delbaen, Eber and Heath (1999). The latter is because $\mathcal{P}^*$ is convex. It should be emphasized that though (2.3) 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 $1 - \alpha$, compare statement (2.8). The point of Theorem 2.1 is to reduce the problem from $1 - \alpha$ to 1, and hence to the earlier work of Avellaneda, Levy and Paras (1995), Lyons (1995) and Mykland (2000).

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 $(\beta_\cdot, S_\cdot^{(1)}, \ldots, S_\cdot^{(p)})$.

Obviously, C is $\mathcal{Q}^*$-closed if it is closed in the supremum norm, but the opposite need not be true. See Section 3.2.

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

THEOREM 2.1 (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 nonempty. Let*

$$\eta = \theta(\beta_\cdot, S_\cdot^{(1)}, \ldots, S_\cdot^{(p)}), \tag{2.6}$$

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

$$\sup_{P^* \in \mathcal{P}^*} E^*|\eta^*| < \infty. \tag{2.7}$$

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 (2.4). *Furthermore, $A_t \geq -K\beta_t$ for all t, $\mathcal{Q}$-a.s.*

In particular,

$$Q(A_T \geq \eta) \geq Q(C) \qquad \text{for all } Q \in \mathcal{Q}, \tag{2.8}$$

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 4.1. Note that both in Theorem 2.1 and in (2.8), 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 below in Section 3 for the case of the call option. More general formulae are given in Avellaneda, Levy and Paras (1995), Lyons (1995), Mykland (2000, 2003) and in Section 5.3.

The condition that θ be bounded below can be seen as a restriction on credit. Since K is arbitrary, this is not severe. It does, however, preclude certain types of trading that earn money with probability 1, using infinite credit. An example of such trading would be a doubling strategy. See Duffie [(1996), Chaper 6.C, pages 103–105] for a discussion. Credit constraints are frequently used in the literature; see, for example, Kramkov (1996) and Karatzas and Shreve (1998). See also Section 5 of Mykland (2000) for a way of weakening the credit restriction while still avoiding arbitrage. Note that the credit limit is more naturally stated on the discounted scale: $\eta^* \geq -K$, and $A_t^* \geq -K$.

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 2.1 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. This is because of Theorem 3.1 in Mykland (2000), the conditions being satisfied in view of Proposition 3.2 in the same paper. In itself, this is a curiosity, since this expression depends on K while A_0 does not.

3. Prediction sets and options prices. To show how sets lead to starting values A_0, Section 3.1 considers payoffs of the call option type, or more generally, ones that are convex functions of the final stock price. This sets the stage for the inclusion of a data application in Section 4.2. Section 3.2 considers theoretical requirements on prediction regions, leading to modifications that are calculationally immaterial, but important as a matter of principle.

3.1. *Two types of sets.* We assume in the following that the interest rate r is constant and known in advance. The main prediction sets considered in the literature are pointwise bounds

$$\sigma_-(S_t, t) \leq \sigma_t \leq \sigma_+(S_t, t) \tag{3.1}$$

[Avellaneda, Levy and Paras (1995) and Lyons (1995)], and the integral bounds (1.1) advocated here and in Mykland (2000).

For the pointwise bounds, the simplest case is just to take the bounds to be constants σ_- and σ_+, in which case the prices of European options become the solutions of the Barenblatt equation [Barenblatt (1979)].

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_0.

For the integral bounds of the form (1.1), calculation will normally involve stopping time arguments; see Theorem 5.1.

To illustrate the implications of these prediction devices, consider first the price of a European call option with expiration T. In this case, $\eta = (S_T - K)^+$. The Black–Scholes (1973)–Merton (1973) price (where σ and r are fixed) at time zero is given by $B(S_0, rT, \sigma^2 T)$, where

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

and where

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

and $d_2 = d_1 - \sqrt{\Xi}$. For more general convex options, the form of B is given by (5.10). The starting values A_0 and the hedge ratios ("deltas") for the three approaches considered, are given in Table 1. The *delta* at time t is, by definition, the number of stocks one would hold at time t to implement the super-replication. For (3.1), it is assumed that σ_- and σ_+ are constants.

To compare these three approaches, note that the function $B(S, R, \Xi)$ is increasing in its last argument. As $\sigma^2 T \le \Xi^+ \le \sigma_+^2 T$, 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.

It is important to see Table 1 in context. 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), page 538] on volatility would appear to 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), page 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 $(S_T - K)^+$ on the prediction set C, but the actual A_T is a combination of option on the security S and swaption on the volatility, in both cases European.

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	$\Xi^- \le \int_0^T \sigma_u^2\,du \le \Xi^+$	$B(S_0, rT, \Xi^+)$	$\frac{\partial B}{\partial S}(S_t, r(T-t), \Xi^+ - \int_0^t \sigma_u^2\,du)$
Extremes based	$\sigma_- \le \sigma_t \le \sigma_+$	$B(S_0, rT, (\sigma^+)^2 T)$	$\frac{\partial B}{\partial S}(S_t, r(T-t), \sigma_+^2(T-t))$

* The function B is defined in (3.2) and (3.3) for call options, and more generally in (5.10); A_0 is the conservative price (2.1). Delta is the hedge ratio (the number of stocks held at time t to superhedge the option).

The results in the table for the extremes based procedure are from Avellaneda, Levy and Paras (1995), and the ones for the average based procedure are from Theorem 5.1, which gives the form of A_0 for a more general European payoff $\eta = g(S_T)$. The delta for the latter case follows from (5.12) at the end of Section 5.3.

The hedge ratio (delta) at time t for the average based set (1.1) 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 $0 = t_0 < \cdots < t_k \le t$. The observed quadratic variation is then

$$\hat{\Xi}_t = \sum_{i=1}^{k} \left(\log S_{t_i} - \log S_{t_{i-1}}\right)^2. \tag{3.4}$$

Note that this quantity converges in probability to $[\log S, \log S]_t$; compare Theorem I.4.47 (page 52) of Jacod and Shiryaev (1987). The natural hedge ratio at time t for the average based procedure would then be

$$\frac{\partial B}{\partial S}\left(S_t, r(T-t), \Xi^+ - \hat{\Xi}_t\right). \tag{3.5}$$

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 (2001b). 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.

3.2. *General form of the prediction set.* 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 the quadratic variations $[\log S^{(i)*}, \log S^{(j)*}]_T$. For the cumulative interest, the application is straightforward. For example, $\{R^- \le -\log \beta_T \le R^+\}$ 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 $\mathbb{C}[0,T]^q$, where $q = \frac{1}{2}p(p-1)+1$ and

$$C_Q = \{(\beta_t, [\log S^{(i)*}, \log S^{(j)*}]_t, i \le j)_{0 \le t \le T} \in F\}, \tag{3.6}$$

then, as the notation suggests, C_Q will depend on $Q \in \mathcal{Q}$. This is not allowed by Theorem 2.1. The trading strategy cannot be permitted 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 3.1 shows that C_Q has a modification that is independent of Q, and that satisfies the conditions of Theorem 2.1.

PROPOSITION 3.1. *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* (3.6). *Then there is a $\mathcal{Q}^*$-closed set C in $\mathcal{F}_T$ so that, for all $Q \in \mathcal{Q}$,*

$$Q(C \Delta C_Q) = 0, \tag{3.7}$$

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 (3.7) implies that C_{P^*} and C_Q can replace C in (2.3) and (2.8), respectively. For the two prediction sets on which our discussion is centered, (1.1) 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 (3.1) 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 2.1 and Proposition 3.1 immediately yields such a result:

THEOREM 3.1 (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* (3.6), *for every $Q \in \mathcal{Q}$. Replace C by C_{P^*} in* (2.3), *and suppose that $\mathcal{P}^*$ is nonempty. Impose the same conditions on $\theta(\cdot)$ and $\eta = \theta(\beta_\cdot, S^{(1)}_\cdot, \ldots, S^{(p)}_\cdot)$ as in Theorem* 2.1. *Then there exists a self financing portfolio (A_t), valid for all $Q \in \mathcal{Q}$, whose starting value is A_0 given by* (2.4), *and which satisfies* (2.8). *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 still there, underlying Theorem 3.2. 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 (3.1) or (1.1), one *can* take the set to be given independently of probability. In fact, this is how Proposition 3.1 is proved.

In the case of European options, this may provide a "probability free" derivation of Theorem 2.1. Under the assumption that the volatility is defined independently of probability distribution, Föllmer (1981) and Bick and Willinger (1994) provide a nonprobabilistic derivation of Itô's formula, and this can be used to show Theorem 2.1 in the European case. Note, however, that this nonprobabilistic 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 page 350 of Bick and Willinger (1994)].

4. Prediction regions from historical data. Until now, we have discused prediction sets without considering two issues. One is how to actually obtain such a prediction set. As a proof of principle we shall, in Section 4.2, discuss a fairly simple example of how to do this. First, however, is another problem. We have behaved as if the prediction sets or prediction limits were nonrandom, fixed and not based on data. This, of course, would not be the case with statistically obtained sets. Section 4.1 faces up to this issue.

4.1. *A decoupled procedure.* A main application of Theorem 2.1 is for statistical prediction sets. Consider 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 (1.1), 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 (1.1) or (3.1), is not quite well defined, but this is solved through Proposition 3.1 and Theorem 3.2. 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 that are available at time 0. 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, then $P \in \Theta$ could be probabilities on $\mathbb{C}[T_-, T]^{p+1}$. More generally, we suppose that the P's are distributions on $\mathcal{S} \times \mathbb{C}[0, T]^{p+1}$, 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 $S_0 = (S_0^{(1)}, \ldots, S_0^{(p)})$ 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, e.g., page 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 (3.6), yielding

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

for each $Q \in \mathcal{Q}_{S_0}$. Then proceed via Proposition 3.1 and Theorem 2.1, or use Theorem 3.2 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.

DEFINITION. Specifically, $\hat{F}$ is a $1 - \alpha$ *prediction set*, provided

$$P(\{(-\log \beta_t, [\log S^{(i)*}, \log S^{(j)*}]_t, i \le j)_{0 \le t \le T} \in \hat{F}\} | \mathcal{H}) \ge 1 - \alpha. \tag{4.1}$$

Here, either (i), in the frequentist setting, (4.1) 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 (2.8), $P(\hat{A}_T \ge \eta | \mathcal{H}) \ge 1 - \alpha$, again for all $P \in \Theta$. Or (ii), $P(\cdot|\mathcal{H})$ is a Bayesian posterior given the data at time 0. In this case, $\mathcal{H} = \mathcal{G}_0$, and $P(\cdot|\mathcal{H})$ is a mixture of P_ω's with respect to the posterior distribution $\hat{\pi}$ at time 0. Since $\mathcal{Q}_{S_0}$ is convex, the mixture would again be in $\mathcal{Q}_{S_0}$, subject to some regularity. Again, (2.8) would yield that $P(\hat{A}_T \ge \eta | \mathcal{H}) \ge 1 - \alpha$, a.s.

In 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. To see it in context, call this 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 2 suggests the operation of the three schemes.

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 (1.1), one can at any time t see how far the realized $\int_0^t \sigma_u^2\,du$ [or, rather, (3.4)] 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 is further discussed in Section 6.

Third, 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 4.2.

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 (1997), Lo (1987) and Zhang (2001). This is, of course, only a small sample of the literature available. The

TABLE 2
*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

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

forthcoming handbook edited by Aït-Sahalia and Hansen (2002) may provide a useful reference.

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 or 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 (2.4) may, therefore, vary according to Θ and to the data. This is because $\hat{\Theta}^*$, and $\mathcal{P}^* = \mathcal{P}^*_{S_0}$ from (2.3), 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 2.1 for subsets of $\mathcal{Q}$, so we do not have an explicit expression for A_0. However, the infimum in (2.1) 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 (2.4). As outlined above, however, this approach is not as robust as the one we have been advocating.

To round off this discussion, we return to the question of measurability. There are (at least) four functions of the data where measurability is in question: (i) the prediction set $\hat{F}$, (ii) the prediction probabilities (4.1), (iii) the starting value $\hat{A}_0$, and (iv) the process $(\hat{A}_t)_{0\le t\le T}$.

We here only consider (ii) and (iii). The first question is heavily dependent on Θ and $\mathcal{S}$. In fact, we shall take the measurability of $\hat{F}$ for granted. We omit discussing question (iv) to not be overly tedious about measurability in this paper.

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

$$\begin{aligned}\mathbf{F}\times\mathcal{S}\to\mathbb{R}:(F,\omega)&\\ &\to P_\omega(\{(-\log\beta_t,[\log S^{(i)*},\log S^{(j)*}]_t, i\le j)_{0\le t\le T}\in F\})\end{aligned}\tag{4.2}$$

and

$$\mathbf{F}\times\mathbb{R}^{p+1}\to\mathbb{R}:(F,x)\to A_0=A_0^F(x).\tag{4.3}$$

To set a σ-field on $\mathbf{F}$, 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 pointwise. On $\mathbf{F}$, this convergence is metrizable (see the proof of Proposition 4.1 for one such metric). Hence $\mathbf{F}$ has a Borel σ-field. This is our σ-field.

PROPOSITION 4.1. *Let Assumptions* A *hold. Impose the same conditions on* $\theta(\cdot)$ *and* $\eta=\theta(\beta_\cdot, S_\cdot^{(1)},\dots,S_\cdot^{(p)})$ *as in Theorem* 2.1. *Then the maps* (4.2) *and* (4.3) *are measurable.*

If we now *assume* that the map $\mathcal{S}\to\mathbf{F}$, $\omega\to\hat{F}$, is measurable, then standard considerations yield the measurability of $\mathcal{S}\to\mathbb{R}$, $\omega\to P_\omega((-\log\beta_t,[\log S^{(i)*},\log S^{(j)*}]_t, i\le j)_{0\le t\le T}\in\hat{F})$ 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 (4.2) equals the expected value of $P_\omega((-\log\beta_t,[\log S^{(i)*},\log S^{(j)*}]_t, i\le j)_{0\le t\le T}\in\hat{F})$, given $\mathcal{H}$, both in the Bayesian and frequentist cases.

4.2. *An implementation.* We here demonstrate by example that the analysis we are proposing can, in fact, be carried out. What we have chosen to use as basis for our development are the results of Jacquier, Polson and Rossi (1994), which analyzes (among other series) the S&P 500 data recorded daily. The authors consider a stochastic volatility model that is linear on the log scale,

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

a.k.a., by exact discretization,

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

where W is a standard Brownian motion and the εs are consequently i.i.d. standard normal. We shall suppose in the following that the effects of interest rate

TABLE 3
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 of posterior interval $\sqrt{\Xi}$	0.168	0.187	0.202	0.217	0.257
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 (1.1) with $\Xi^- = 0$. A 5% p.a. known interest rate is assumed; $S_0 = 100$.

uncertainty are negligible. With some assumptions, their posterior distribution, as well as our corresponding options price, are given in Table 3. Note that it is customary to state the volatility *per annum* and on a square root scale.

In the above, we are bypassing the issue of conditioning on σ_0^2. Our excuse for this is that σ_0^2 appears to be approximately observable in the presence of high frequency data. Following Foster and Nelson (1996), Zhang (2001) and Mykland and Zhang (2001a), the error in observation is of the order $O_p(\Delta t^{1/4})$, where Δt is the average distance between observations. See also Andersen, Bollerslev, Diebold and Labys (2001). 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.

5. The effect of interest rates and a general formula for European options.

5.1. *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. We suppose that intervals are set on integral form, in the style of (1.1). 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,

$$R^- \le \int_0^T r_u\,du \le R^+. \tag{5.1}$$

In combination with (1.1), this gives $A_0 = B(S_0, R^+, \Xi^+)$; compare Section 2 of Mykland (2000).

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 things are rarely that simple.

To increase the complexity, suppose that one of the securities available in the market is a zero coupon bond maturing with value \$1 at the time T of maturity of the option. If the price at time t of this bond is Λ_t, then $\Lambda_T = 1$. This security, therefore, has a riskless final payoff, but it is risky from one day to the next, and it can lose value. This is as opposed to β_t, where the immediate return is riskless, and it cannot lose value, but the return from holding it over time is nonetheless random (typically).

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 will, in practice, be substantial room for judgement in these matters.

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 3.1, different prediction regions give different values of A_0. If one combines (1.1) and (5.1), the form of A_0, given on page 668 of Mykland (2000), is somewhat unpleasant. 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. Consider the stock price discounted (or rather, blown up) by the zero coupon bond,

$$S_t^{(*)} = S_t / \Lambda_t. \tag{5.2}$$

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 (1.1), in the form

$$\Xi^{*-} \le \int_0^T \sigma_t^{*2}\, dt \le \Xi^{*+}. \tag{5.3}$$

We shall see in Section 5.3 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.

TABLE 4
Comparative prediction sets: *r nonconstant**
(*convex European options*, *including calls*)

Λ_t **available?**	A_0 **from (1.1) and (5.1)**	A_0 **from (5.3)**
No	$B(S_0, R^+, \Xi^+)$	not available
Yes	unaesthetic (see text)	$B(S_0, -\log\Lambda_0, \Xi^{*+})$

* The function B is defined in (3.2) and (3.3) for call options, and more generally in (5.10).

The situation for the call option is summarized in Table 4. The value A_0 depends on two issues: is the zero coupon bond available and which prediction region should one use?

Table 4 follows directly from the development in Section 5.3. The hedge ratio corresponding to (5.3) is given in (5.12) below.

5.2. *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 (5.3). In the following, we suppose that interest rates follow a linear model [introduced in the interest rate context by Vasicek (1977)],

$$dr_t = a_r(b_r - r_t)\,dt + c_r\,dV_t, \tag{5.4}$$

where V is a Brownian motion independent of B in (1.2).

The choice of interest rate model highlights the point made in Section 4.1: 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 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*, November 14, 1998, page 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

$$\int_0^T r_t\,dt \qquad \text{has distribution } N(\nu, \gamma^2). \tag{5.5}$$

It follows that

$$\log\Lambda_0 = -\nu + \tfrac{1}{2}\gamma^2. \tag{5.6}$$

In this case, if we suppose that the stock follows (1.2), then (5.4) and (5.5) yield

$$\int_0^T \sigma_u^{*2}\, du = \int_0^T \sigma_u^2\, du + \gamma^2. \tag{5.7}$$

Prediction intervals can now be adjusted from (1.1) to (5.3) 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.

5.3. *General European options.* We here focus on the single prediction set (5.3). The situation of constant interest rate (Section 3.1) is a special case of this, where the prediction set reduces to (1.1).

THEOREM 5.1. *Under Assumptions* A, *and with prediction set* (5.3), *if one hedges liability $\eta = g(S_T)$ in S_t and Λ_t, the quantity A_0 in* (2.1) *has the form*

$$A_0 = \sup_{\tau} \tilde{E} \Lambda_0 g\left(\frac{1}{\Lambda_0} \tilde{S}_\tau\right), \tag{5.8}$$

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

$$d\tilde{S}_t = \tilde{S}_t\, d\tilde{W}_t \qquad \textit{with } \tilde{S}_0 = s_0, \tag{5.9}$$

where s_0 is the actual observed value of S_0.

The proof is given in Section 7.

If one compares this with the results concerning nonconstant interest in Mykland (2000), the above would seem to be more elegant, and it typically yields lower values for A_0. It is also easier to implement since $\tilde{S}$ is a martingale.

To consider the case of convex or concave options, write

$$B(S, R, \Xi) = \exp(-R) E g\big(S \exp(R - \Xi/2 + \sqrt{\Xi} Z)\big), \tag{5.10}$$

where Z is standard normal. As in Section 3.1, the Black–Scholes (1973)–Merton (1973) price at time t for stock price S and nonrandom constant interest r and volatility σ^2 can be written $B(S_t, r(T-t), \sigma^2(T-t))$.

In our case, if g is convex (e.g., call and put options), then the martingale property of $\tilde{S}$ yields that the A_0 in (5.8) has the value

$$A_0 = B\big(S_0, -\log \Lambda_0, \Xi^{*+}\big). \tag{5.11}$$

In the case of concave g, one similarly gets that $A_0 = B(S_0, -\log \Lambda_0, \Xi^{*-})$.

It is shown in Section 7 that the delta hedge ratio for convex g is

$$\frac{\partial B}{\partial S}\left(S_t, -\log \Lambda_t, \Xi^{*+} - \int_0^t \sigma_u^{*2}\, du\right). \tag{5.12}$$

In practice, one has to make an adjustment similar to that at the end of Section 3.1.

6. The practical role 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; compare 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. Suppose that a market participant implements a self financing trading strategy with portfolio value V_t. The usual approach if such a strategy goes wrong in the sense of losing substantial amounts of money, is to liquidate the holding: buy back the option and sell the hedging portfolio. This can be quite expensive.

As an alternative, one can do the following. While trading with portfolio V, also monitor the evolution of a conservative value A_t, based on a $1-\alpha$ prediction interval. Also, set aside reserves of K dollars, with $K > A_0 - V_0$. The exit strategy is then to switch from portfolio V to portfolio A if V_t^* goes so low that it hits $A_t^* - K$ (the superscript "$*$" refers, as usual, to discounting). This provides an orderly and presumably less expensive exit, as it avoids liquidation.

7. Proofs.

PROOF OF THEOREM 2.1. Assume the conditions of Theorem 2.1. Let $m \geq K$, and define $\theta^{(m)}$ by

$$\begin{aligned}\theta^{(m)}&(\beta_\cdot, S_\cdot^{(1)}, \dots, S_\cdot^{(p)})\\ &= \theta(\beta_\cdot, S_\cdot^{(1)}, \dots, S_\cdot^{(p)}) I_C(\beta_\cdot, S_\cdot^{(1)}, \dots, S_\cdot^{(p)}) - m\beta_T I_{\tilde{C}}(\beta_\cdot, S_\cdot^{(1)}, \dots, S_\cdot^{(p)}),\end{aligned}$$

where $\tilde{C}$ 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

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

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 Assumptions 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 of Jacod and Shiryaev [(1987), pages 318 and 322].

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

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

Also, set for $m \geq K$,

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

The supremum is $\mathcal{F}_t$-measurable since this σ-field is analytic (see Remark 2.1), 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 of Dellacherie and Meyer [(1978), pages 42 and 43]; see also the treatment in Pollard [(1984), pages 196 and 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, $A_t^{(m)*} = \beta_t^{-1} \theta_t^{(m)}(\beta_\cdot, S_\cdot^{(1)}, \dots, S_\cdot^{(p)})$ 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 (pages 16 and 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 $\tilde{A}_t^{(m)*}$ the resulting supermartingale. Note that $\tilde{A}_t^{(m)*} \leq -K$ on the entire space, and set

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

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

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

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

In view of Proposition I.3.14 (again) in Karatzas and Shreve (1991), this does not interfere with the super-martingale 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

$$A_0^{(K)} = \sup_{P^* \in \mathcal{P}_c^*} E^*(\eta^*), \tag{7.1}$$

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

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 $m_k \ge m$, and with $\tilde{C}$ denoting the complement of C,

$$
\begin{aligned}
A_0^{(K)*} &= A_0^{(m)*} \\
&= \theta_0^{(m)}(b_0, s_0^{(1)}, \dots, s_0^{(p)}) \\
&\le \theta_0^{(m_k)}(b_0, s_0^{(1)}, \dots, s_0^{(p)}, P_{m_k}^*) + m_k^{-1} \\
&\le E_{m_k}^* \beta_T^{-1} \theta(\beta_\cdot, S_\cdot^{(1)}, \dots, S_\cdot^{(p)}) + P_{m_k}^*(\tilde{C})(K - m_k) + m_k^{-1} \\
&\le E_{m_k}^* \beta_T^{-1} \theta(\beta_\cdot, S_\cdot^{(1)}, \dots, S_\cdot^{(p)}) + P_{m_k}^*(\tilde{C})(K - m) + m_k^{-1} \\
&\le E^* \beta_T^{-1} \theta(\beta_\cdot, S_\cdot^{(1)}, \dots, S_\cdot^{(p)}) + \limsup_{k \to +\infty} P_{m_k}^*(\tilde{C})(K - m) + o(1)
\end{aligned} \tag{7.2}
$$

as $k \to \infty$. The first term on the right-hand side of (7.2) is bounded by the weak compactness of $\mathcal{Q}_c^*$. The left-hand side is a fixed, finite number. Hence, $\limsup P_{m_k}^*(\tilde{C}) = 0$. By the $\mathcal{Q}^*$-closedness of C, it follows that $P^*(C) = 1$.

Hence, (7.2) yields that the right-hand side in (7.1) is an upper bound for $A_0^{(K)*} = A_0^{(m)*}$. Since this is also trivially a lower bound, (7.1) 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 $s \le t$,

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

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

For the case of $t = 0$, (7.1) yields that

$$
A_0^{(m,\infty)} = \sup_{P^* \in \mathcal{P}^*} E^*(\eta^*), \tag{7.3}
$$

where the nonobvious inequality ($\ge$) follows from monotone convergence and assumption (2.7). Since one can choose $m = K$, Theorem 2.1 is proved. □

PROOF OF PROPOSITION 3.1. Extend the space $\mathbb{C}^{p+1}$ to $\mathbb{C}^{p+q}$. Consider the set $\tilde{\mathcal{Q}}$ of probabilities $\tilde{Q}$ on $\mathbb{C}^{p+q}$ for which the projection onto $\mathbb{C}^{p+1}$ is in $\mathcal{Q}$ and so that $([\log S^{(i)*}, \log S^{(j)*}]_t, i \le j)$ are indistinguishable from $(x_t^{(k)}$, $k = p+2, \dots, p+q)$. Now consider the set (in $\mathbb{C}^{p+q}$) $F' = \{\omega : (\beta, x^{(p+2)}, \dots, x^{(p+q)}) \in F\}$. For every $\tilde{Q} \in \tilde{\mathcal{Q}}$, let the set C_Q be given by (3.6). Then $\tilde{Q}(C_Q \Delta F') = 0$. Hence F' is in the completion of $\mathcal{F}_t \otimes \{\mathbb{C}^{q-1}, \varnothing\}$ with respect to $\tilde{\mathcal{Q}}$. It follows that there is a C in $\mathcal{F}_T$ so that $P^*(C\Delta F') = 0$ for all $P^* \in \mathcal{Q}^*$. 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 $\tilde{P}_n^*$ and $\tilde{P}^*$ in $\tilde{\mathcal{Q}}$. By Corollary VI.6.7 (page 342) in Jacod and Shiryaev (1987), $\tilde{P}_n^* \to \tilde{P}^*$ weakly. Hence, since F and hence F' is closed, if $\tilde{P}_n^*(F') \to 1$, then $\tilde{P}^*(F') = 1$. The same property must then also hold for C. □

PROOF OF PROPOSITION 4.1. Let d be the uniform metric on $\mathbb{C}^q$, that is, $d(x, y) = \sum_{i=1,\dots,q} \sup_{t\in[0,T]} |x_t^i - y_t^i|$. Let $\{z_n\}$ be a countable dense set in $\mathbb{C}^q$ with respect to this metric. It is then easy to see that

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

is a metric on $\mathbf{F}$ 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 $\mathbf{F} \times \mathbb{C}[0, T]^q \to \mathbb{R}$. From this, the indicator function $I_F(x) = \inf_{m\in\mathbb{N}} f(x)$ is upper semicontinuous, and hence measurable. The result for (4.2) then follows from Exercise 1.5.5 (page 43) in Stroock and Varadhan (1979). The development for (4.3) is similar. □

PROOF OF THEOREM 5.1. The A_t be a self financing trading strategy in S_t and Λ_t that covers payoff $g(S_T)$. In other words,

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

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

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

By using the Dambis (1965) or Dubins and Schwarz (1965) time change, the result follows. □

Derivation of the hedging strategy (5.12). As discussed in Mykland (2000), the function $B(S, R, \Xi)$ defined in (5.10) satisfies two partial differential equations, namely, $\frac{1}{2} B_{SS} S^2 = B_\Xi$ 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:

$$dB(S_t, \Xi_t, -\log \Lambda_t) = B_S\, dS_t - B_R \frac{1}{\Lambda_t}\, d\Lambda_t + B_\Xi (d\langle \log S^* \rangle_t + d\Xi_t). \tag{7.4}$$

If one looks at the right-hand side of (7.4), 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 (7.4) to be nonpositive. In the case of a convex or concave payoff, one just uses $d\Xi_t = -d\langle \log S^* \rangle_t$, with Ξ_0 as Ξ^{*+} or Ξ^{*-}, as the case may be. □

Acknowledgments. The author thanks the two referees, an Associate Editor and Jin Qian, Seongjoo Song, Jian Zhang and Lan Zhang, for comments and corrections.

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 HANSEN, L. (eds.) (2002). *Handbook of Financial Econometrics*. To appear.

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. Econom. Statist.* **18** 146–153.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. and LABYS, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.* **96** 42–55.

ARTZNER, P., 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.

AVELLANEDA, M., LEVY, A. and PARAS, A. (1995). Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Mathematical Finance* **2** 73–88.

BARENBLATT, G. I. (1979). *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-Gausian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** 167–241.

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., GRUNDY, B. D. and WIENER, Z. (1996). General properties of option prices. *J. Finance* **51** 1573–1610.

BIBBY, B. M. and SØRENSEN, M. (1995). Martingale estimation 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. *Theory of Stochastic Processes* **18** 49–56.

BIBBY, B. M. and SØRENSEN, M. (1996b). A hyperbolic diffusion model for stock prices. *Finance Stoch.* **1** 25–41.

BICK, A. and REISMAN, H. (1993). Generalized implied volatility. Preprint.

BICK, A. and WILLINGER, W. (1994). Dynamic spanning without probabilities. *Stochastic Process. Appl.* **50** 349–374.

BLACK, F. and SCHOLES, M. (1973). The pricing of options and corporate liabilities. *J. Political Economy* **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. *J. Financial Economics* **62** 131–167.

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. (1999). On dynamic measures of risk. *Finance Stoch.* **3** 451–482.

DACUNHA-CASTELLE, D. and 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 measures. *Ann. Appl. Probab.* **5** 926–945.

DELLACHERIE, C. and MEYER, P.-A. (1978). *Probabilities and Potential* **1**. North-Holland, Amsterdam.

DUBINS, L. E. and SCHWARZ, 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 Univ. Press.

EL KAROUI, N., JEANBLANC-PICQUÉ, M. and SHREVE, S. E. (1998). Robustness of the Black and Scholes formula. *Math. Finance* **8** 93–126.

FLORENS-ZMIROU, D. (1993). On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.* **30** 790–804.

FÖLLMER, H. (1981). Calcul d'Itô sans probabilités. *Seminaire de Probabilités XV*. *Lecture Notes in Math.* **850** 143–150. Springer, Berlin.

FOSTER, D. P. and NELSON, D. B. (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica* **64** 139–174.

GENON-CATALOT, V. and JACOD, J. (1994). Estimation of the diffusion coefficient for diffusion processes: Random sampling. *Scand. J. 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. (1998). Spectral methods for identifying scalar diffusions. *J. Econometrics* **86** 1–32.

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.

HULL, J. C. (1999). *Options, Futures and Other Derivatives*, 4th ed. Prentice-Hall, Upper Saddle River, NJ.

JACOD, J. (1979). *Calcul stochastique et problèmes de martingales*. *Lecture Notes in Math.* **714**. Springer, Berlin.

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

JACOD, J. and PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theory for Stochastic Processes*. Springer, Berlin.

JACQUIER, E., POLSON, N. and ROSSI, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *J. Bus. Econom. Statist.* **12** 371–417.

KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, Berlin.

KARATZAS, I. and SHREVE, S. E. (1998). *Methods of Mathematical Finance*. Springer, Berlin.

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 decomposition of supermartingales and hedging contingent claims in incomplete security markets. *Probab. Theory Related Fields* **105** 459–479.

KÜCHLER, U. and SØRENSEN, M. (1997). *Exponential Families of Stochastic Processes*. Springer, Berlin.

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. *Applied Mathematical Finance* **2** 117–133.

MERTON, R. C. (1973). Theory of rational option pricing. *Bell J. Econ.* **4** 141–183.

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

MYKLAND, P. A. (2003). The interpolation of options. *Finance Stoch.* **7** 417–432.

MYKLAND, P. A. and ZHANG, L. (2001a). Inference for volatility-type objects and implications for hedging. Technical Report 506, Dept. Statistics, Univ. Chicago.

MYKLAND, P. A. and ZHANG, L. (2001b). ANOVA for diffusions. Technical Report 507, Dept. Statistics, Univ. Chicago.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

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.) 589–612. Oxford Univ. Press.

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

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

ZHANG, L. (2001). From martingales to ANOVA: Implied and realized volatility. Ph.D. dissertation, Dept. Statistics, Univ. Chicago.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637
E-MAIL: mykland@galton.uchicago.edu