DEMS WORKING PAPER SERIES

Option pricing in an imperfect world

Gianluca Cassese

No. 277 – June 2014

OPTION PRICING IN AN IMPERFECT WORLD

GIANLUCA CASSESE

ABSTRACT. In a model with no given probability measure, we consider asset pricing in the presence of frictions and other imperfections and characterize the property of coherent pricing, a notion related to (but much weaker than) the no arbitrage property. We show that prices are coherent if and only if the set of pricing measures is non empty, i.e. if pricing by expectation is possible. We then obtain a decomposition of coherent prices highlighting the role of bubbles. Eventually we show that under very weak conditions the coherent pricing of options allows for a very clear representation from which it is possible, as in the original work of Breeden and Litzenberger, to extract the implied probability. Eventually we test this conclusion empirically via a new non parametric approach.

1. INTRODUCTION

It is unanimous the opinion that over the last decades the theory of finance has produced an impressive number of beautiful results, particularly in asset pricing. At the same time asset pricing has originated a vast empirical literature with special attention accorded to testing the classical conclusion of Breeden and Litzenberger [13] and Banz and Miller [4] – namely that option prices contain valuable information on the risk neutral density. Thus, many authors incline to believe that the beauty of modern finance, theoretical or applied, lies in the fact that the crucial property of risk neutral pricing rests ultimately on the simple and sound tenet that a market populated by rational economic agents admits no arbitrage opportunities.

Notwithstanding the general opinion, the proof of the claim that absence of arbitrage implies risk neutral pricing, a result known in the literature as the *fundamental theorem of asset pricing*, has long been a challenge for mathematical economists, from Harrison and Kreps [40] and Kreps [51] to Delbaen and Schachermayer [23]. This key result requires in fact a much more stringent condition than absence of arbitrage, requiring implicitly that agents have the ability of evaluating prospective profits from trading in a way that involves a high degree of probabilistic sophistication. Financial models are then an adequate description of markets populated by individuals whose decision process is dominated by probability assessments. A convincing amount of empirical observations seems, however, to document that the attitude of investors towards uncertainty is not easy to reconcile with

1

probabilistic rationality but rather generates puzzles which contrast with some of the principles of financial theory.

In asset pricing models probabilistic sophistication gets along with the sophistication of investors in keeping up with portfolio policies which prescribe to rebalance at a continuous pace. On the side of trading, the wedge between financial models and real markets is even more apparent as a consequence of the general neglect accorded by most models to all sorts of market imperfections, such as bid/ask spreads, transaction costs or restrictions to trade. Some easily observed facts, such as those documented by Lamont and Thaler in [52] and [53], or the high ratio of option prices violating some basic no arbitrage condition, become then difficult challenges for financial theory.

The contribution of this work to the asset pricing literature lies in the choice to describe the market mechanism under a minimal set of assumptions and yet obtain useful characterizations of financial prices and testable restrictions. A distinguishing feature of this model, first appeared in [16] (and, more recently, considered by Riedel [60]), is the absence of a given reference probability measure which is replaced here with a qualitative description of how agents rank uncertain outcomes. As in the tradition of subjective probability, we believe that probabilities should be regarded more as the outcome of choice than as an input to it.

Another distinguishing feature of the present paper is the description of the market mechanism. Not only we consider bid/ask spreads and (to a minor extent) fixed trading costs but we also allow for several additional restrictions to trade. First, agents may be prevented from forming portfolios of arbitrarily large size, so that it may not be possible to run money pumps – in case they exist. Arbitrage phenomena may thus have a minor impact on economic equilibrium and the no arbitrage principle looses, as a consequence, part of its appeal. Second, not all portfolios may be shorted, and not just as a result of trading restrictions. In the presence of credit risk, taking long or short positions should be considered as two separate investments given that the implicit level of risk depends on the reliability of the investor playing the short side. Third, we do not identify assets with their payoff so that it is possible to have two assets promising the same payoff at different prices. Fourth, the opportunity to invest in some asset may be available only if combined with other assets. Some risky positions are in fact possible only if a convenient level of collateral is put up. The impact of margins when shorting options, an empirical fact accommodated in our model, has rarely been considered in asset pricing.

Eventually, we don't assume the existence of a riskless asset but rather of a *numéraire* whose non negative payoff is used as the discount factor. In contrast with the basic principles of neoclassical economics, the choice of the *numéraire*, given the absence of a reference probability, is non neutral and an arbitrage opportunity arising with a given *numéraire* may no longer be such with a different discount factor. This point has a clear interpretation from the point of view of risk management and is also discussed by El-Karoui and Ravanelli [30].

In the framework outlined above we discuss three distinct, increasingly restrictive notions of market rationality: coherence, efficiency and absence of arbitrage opportunities. We obtain in

Theorem 3 a characterization of coherent prices in terms of a set of pricing measures. Our result is near in spirit to that obtained in the pioneering work of Jouini and Kallal [47] (in an $L^2(P)$ setting) but departs from it is several ways. First, pricing measures can only be applied to claims with limited discounted losses; second, they in our model are just finitely additive. We obtain in Theorem 4 exact necessary and sufficient conditions for countable additivity which justify the conclusion that this additional property should be regarded more as a mathematical artifact than an economic implication. In Theorem 5 we obtain a decomposition of coherent prices which highlights the role of bubbles. It is also possible to represent pricing measures with a capacity, similarly to what assumed in the work of Chateauneuf et al. [19]. The connection between with capacities is also explored in Cerreia-Vioglio et al. [18].

Of course, over the years several authors have investigated restrictions to trading similar to those considered here. The first contribution is probably due to Demsetz [25] who represents the transaction mechanism not differently from the production process and interprets its costs as the reward of a corresponding input. Leaving aside the microstructure literature, in which transaction costs are the heart of the matter (see Hasbrouck [41], Huang and Stoll [44] and Stoll [65] for a small sample), there has been a number of papers, from the pioneering work of Hahn [38] to the more recent one of Bisin [7], trying to incorporate transaction costs into general equilibrium. In these studies transaction costs act as a limit to the opportunity of exploiting potential arbitrage profits and contribute to restoring equilibrium. In modern financial theory the first papers on the matter have been those of Bensaid, Lesne, Pages and Scheinkman [5] and, most of all, of Jouini and Kallal [47]. More recent papers include Bouchard [12], Napp [58] (who first models trading restrictions via closed convex cones), Jouini and Napp [48] (who describe investments as cash flows with convex cone constraints and assume no *numéraire*), Kabanov and Stricker [49] (who consider very general forms of costs) and Schachermayer [62]. With no claim to completeness, one should also mention the work of Amihud and Mendelson [3], Dermody and Prisman [26], Prisman [59] and, most recently, Roux [61].

A limit intrinsic to this literature is that, with the noticeable exception of Luttmer [56] (who extends the classical conclusions of Hansen and Jagannathan [39]), none of the preceding papers produces testable restrictions. Aiming at applications, we focus then on the options market where trading is for the most part anonymous and our preceding results assume, as a consequence, a more tractable form. Our aim is to repropose the classical exercise of Breeden and Litzenberger [13] and of Banz and Miller [4]. A large number of papers have addressed this issue, particularly with non parametric techniques (see e.g. [1], [2], [46] and [68]). All of these papers, however, start assuming that the options market is efficient and free of arbitrage opportunities and need as a consequence a preliminary, careful filtering of the data. We highlight that, despite being generally overlooked, this first step is likely to induce a selection bias in estimates.

In our setting assuming efficiency of option prices is unnecessary. We prove in Theorem 7 that for each convex family of derivatives it is possible to extract an implicit, *countably additive* measure

associated with its superhedging price; in Lemma 6 we show how to implement this conclusion numerically. Thus, despite the weak starting assumptions and the potential role of bubbles, our result is surprisingly near to that of Breeden and Litzenberger.

We elaborate then a non parametric, two step procedure to estimate the probability implicit in option prices. In the first step, we construct a set of artificial derivatives whose payoff is (i) as smooth as required and (ii) approximates the option payoff uniformly. A control parameter, playing conceptually the same role as the bandwidth in kernel regression, calibrates smoothness *vs* goodness of fit. Second, we obtain for these derivatives the implied efficient price, which represents our non parametric estimate. The implied probability is then, as usual, simply the derivative of this newly obtained price with respect to the parameter playing the role of the strike price. Differently from other non parametric approaches – the kernel approach of Aït-Sahalia and Lo [2], the local polynomial methodology of Aït-Sahalia and Duarte [1] or the constrained least squares approximation of Yatchew and Härdle [68] – is that we do not perform smoothing of the option *prices* but rather of their *payoff* and compute the corresponding price via superhedging. We show that with an accurate choice of the smoothing parameter and even if taking into explicit account sample selection effects, our estimates have desirable convergence properties. We conclude with some empirical analyses on simulated as well as historical datasets.

The paper is structured as follows. In section 2 we describe the market and the trading mechanism. It is introduced and briefly discussed the notion of a negligible event which is crucial to our approach. In section 3 we discuss the properties of coherence, efficiency and of absence of arbitrage and, in the following section 4, we obtain an explicit characterization of coherent prices in terms of pricing measures. In section 5 we develop a decomposition of coherent asset prices from which emerges the existence of bubbles. We then move to option markets in section 6 where we prove a general representation for prices of convex derivatives, involving bubbles and an implicit pricing measure. In the following section 7 we propose our econometric strategy and prove some properties of the estimator obtained. In the concluding section 8, we illustrate some empirical investigations using Monte Carlo simulations as well as market data.

Throughout the paper we adopt the following mathematical symbols and conventions. $\mathfrak{F}(X)$ denotes the collection of real-valued functions on some space X and $\mathfrak{F}_0(X)$ designates those $f \in \mathfrak{F}(X)$ whose support $\{x \in X : f(x) \neq 0\}$ is a finite set. At times, we prefer f_x to $f(x)$. If $f \in \mathfrak{F}(X)$ and if $-X \subset X$, the symbol f^c will be used to denote the conjugate of f, that is the element of $\mathfrak{F}(X)$ defined implicitly by letting

$$(1) \qquad\qquad f^c(x) = -f(-x) \qquad x \in X$$

We set conventionally $0/0 = 0$, $\inf \varnothing = \infty$ and $\sum \varnothing = 0$. $\overline{\mathbb{R}}$ denotes the extended real numbers.

2. Markets, Prices, Investors

Assets traded on the market are identified with a *"ticker"*, $\alpha \in \mathfrak{A}$, and are associated with a corresponding payoff, $X(\alpha)$. The latter is modeled simply as a function on some given space Ω, i.e.

an element of $\mathfrak{F}(\Omega)$. No mathematical structure is imposed on the set of traded payoffs, neither topological nor measure theoretic. Although it is natural to interpret Ω as the sample space and $X(\theta)$ as a random quantity (and thus to attribute to our model an intrinsic static nature) we may as well choose $\Omega = S \times \mathbb{R}_+$, with S the sample space and $\mathbb{R}_+$ the time domain and thus give to our construction a full fledge dynamical structure.

2.1. **Trading Strategies.** Investors trade claims by taking a finite number of either long or short positions, in respect of the restrictions imposed by the market. A trading rule is then just an element of the space $\mathfrak{F}_0(\mathfrak{A})$. The trading rule which consists solely of one unit of the claim $\alpha \in \mathfrak{A}$ will be denoted by δ_α. To each trading rule θ corresponds the final gain

$$(2) \qquad\qquad X(\theta) = \sum_{\alpha \in \mathfrak{A}} \theta(\alpha) X(\alpha)$$

Of course, $X(\delta_\alpha) = X(\alpha)$.

Inspired by real markets, one may imagine several restrictions to asset trading, further to the constraints of respecting the balance of budget and of forming *finite* portfolios. These include short selling prohibitions or margin requirements and others that ultimately guarantee the enforcement of some form of bound to losses and, possibly, to prevent investors from taking positions too large. The symbol Θ, that denotes hereafter the set of all admissible trades, specifies all restrictions to asset trading. We assume the following:

Assumption 1. Θ *is a convex subset of* $\mathfrak{F}_0(\mathfrak{A})$ *containing the origin.*

Under Assumption 1 investors need not be permitted to take positions of either sign, long or short. This is consistent with the restriction to losses recalled above. Moreover, investors may encounter restrictions in the choice of the scale of the investment. On this point we depart significantly from much of the literature on asset pricing with or without transaction costs, see e.g. [47], [49] or [56]. A possible relaxation of this restriction is to allow $\lambda\theta \in \Theta$ for all $\lambda \geq 0$ whenever $\theta \in \Theta$ satisfies $X(\theta) \geq 0$. A major implication of this is that arbitrage opportunities, when available, may not have a disruptive impact on market equilibrium and the no arbitrage principle, as a consequence, may no longer be crucial. Eventually, we do not require that $\delta_\alpha \in \Theta$, i.e. that each asset may be traded individually due, e.g., to the requirement of putting up margins when taking positions on derivatives markes.

2.2. **Prices.** For each $\alpha \in \mathfrak{A}$ we denote by $q^a(\alpha)$ and $q^b(\alpha)$ its *ask* and *bid* price respectively.

Assumption 2. *The functions* $q^a, q^b \in \mathfrak{F}(\mathfrak{A})$ *are such that* $q^a(\alpha) \geq q^b(\alpha)$ *for all* $\alpha \in \mathfrak{A}$.

We highlight that in our model financial prices are defined independently of the asset payoff but depend rather on the asset name. This apparently innocuous detail makes our approach compatible with some of the pricing anomalies observed on markets by which the trading of one same asset at different market locations or simply under different names, may produce different prices (see the examples on close-end funds or of twin stocks reported in Lamont and Thaler [53]).

In order to form a given trading strategy $\theta \in \Theta$ an investor pays an ask price for each long position and earns a bid price for each short one. The corresponding cost amounts thus to

$$(3) \qquad q(\theta) = \sum_{\alpha \in \mathfrak{A}} \left[\theta(\alpha)^+ q^a(\alpha) - \theta(\alpha)^- q^b(\alpha) \right] \qquad \theta \in \Theta$$

It is clear that $q(\delta_\alpha) = q^a(\alpha)$ and, if $-\delta_\alpha \in \Theta$, that $q^c(\delta_\alpha) = -q(-\delta_\alpha) = q^b(\alpha)$.

The basic assumption behind (3) is that each position in a portfolio is priced separately. This is in accordance with trading anonymity prevailing in specialist markets but is certainly not an adequate description of OTC trading. Options markets, on which we shall focus in the last sections, are quite well represented by (3) at least for orders which fall below the size limits of the market maker. Large orders, instead, are in general processed on a separate track and the price is set *ad hoc*.

Market frictions include, further to the bid/ask spread, also some fixed costs, such as brokerage fees, which are paid by investors to have access to markets. Given that on each market investors may trade more than one asset, we may thus think of markets as a partition $\mathfrak{M}$ of subsets of $\mathfrak{A}$ and for each $M \in \mathfrak{M}$ we designate by $c(M) \geq 0$ the corresponding fixed cost. For example, all options on a given underlying are traded on the same market, independently of the strike or maturity so that any option strategy will involve the same fees. Thus the fixed cost associated with an investment strategy is

$$(4) \qquad c(\theta) = \sum_{\{M \in \mathfrak{M}: \sup_{\alpha \in M} |\theta(\alpha)| > 0\}} c(M) \qquad \theta \in \Theta$$

A realistic modeling fixed trading costs turns out to be quite difficult due to their extremely various nature[1]. (4) is just one possible model.

The total cost associated with a trading strategy θ amounts to

$$(5) \qquad t(\theta) = q(\theta) + c(\theta) \qquad \theta \in \Theta$$

It is easy to deduce from (3) and (5) some elementary properties:

Lemma 1. *The functional $q \in \mathfrak{F}(\Theta)$ defined in (3) is (i) positively homogeneous, (ii) subadditive and (iii) such that*

$$(6) \qquad q(f + g) = q(f) + q(g) \qquad for\ all \quad f, g \in \Theta \quad with \quad fg \geq 0$$

The functional $t \in \mathfrak{F}(\Theta)$ defined in (5) is subadditive and satisfies $t(0) = 0$.

In many a paper on asset pricing with frictions, starting with the seminal paper of Jouini and Kallal [47], subadditivity is the only distinguishing feature characterizing the existence of bid/ask spreads. Another paper following this choice is that of Luttmer [56]. In Chateauneuf et al. [19], the pricing functional is represented via a capacity and is then not only subadditive but even comonotonic, a property somehow akin to (6). In these papers an explicit description of the costs of trading is omitted (a remarkable exception is [49]) and in so doing, we claim, one risks to miss

---

[1]As a matter of fact transaction fees tend to be stepwise increasing with the order size rather then fixed.

important details of the price mechanism and to mix up effects that may actually originate from different sources, e.g. bounded rationality or restrictions to market participation. We will refer to property (6) as *anonymity* and, although in the following Theorems 3 through 5 this plays no role, it will be important when dealing with option prices for which, we believe, it is perfectly adequate.

2.3. **The *Numéraire* Asset.** Financial models (with the noteworthy exception of [48]) commonly assume the existence of a riskless asset, often interpreted as a bond, that may equally well serve the purpose of borrowing or lending. This assumption plays three distinct roles. First, it enables agents to move wealth back and forth in time in a safe way and thus provides a firm basis to define the present value of future wealth. Second, it allows to identify explicitly the *numéraire* of the economy, removing the arbitrariness that arises whenever several assets may play that same role. Third, if the investment in the bond is unrestricted then this asset plays a residual role in portfolio models, guaranteeing the effectiveness of portfolio constraints.

This assumption is however not only strongly counterfactual but more troublesome than it appears at first sight. First, if the bond is not fully free of risk but just evolves in a predictable way (as is often the case in continuous time models) then the role of the discounting asset is no longer neutral as its implicit risk entwines with the one originating from the underlying asset. El Karoui and Ravanelli [30] discuss this point at length and show that risk measures are affected by discounting in a significant way, when the discount factor is risky. Moreover, in equilibrium models, such as those considering the role of *noise trading* (see [24] or [63]), the riskless nature of the bond may not survive Walras law unless its elasticity of supply is infinite[2]. Eventually if investors are prone to credit risk one should consider borrowing and lending as two different financial contracts, given that the final payoff is ultimately a function of the reliability of the two intervening parts.

We summarize the preceding discussion in the following:

Assumption 3. *There exists $\alpha_0 \in \mathfrak{A}$ such that (i) $1 \geq X(\alpha_0) > 0$, (ii) if $\theta \in \Theta$ then $\theta + \lambda \delta_{\alpha_0} \in \Theta$ if and only if $\lambda \geq 0$, (iii) $c(\delta_{\alpha_0}) = 0$, (iv) $q_0 > 0$.*

We shall refer to α_0 as the *numéraire* asset and simplify δ_{α_0}, $X(\alpha_0)$ and $q(\alpha_0)$ as δ_0, X_0, q_0. Assumption 3.(i) is fairly general to allow for virtually all sorts of dynamics but it excludes the occurrence of default with no recovery value. If one visualizes the *numéraire* asset as government bonds one may perhaps consider this restriction not too far from reality given that in market economies government bonds have always been redeemed at some positive value. Alternatively, identifying the *numéraire* asset with the bank account one may argue, likewise, that the bank account is in most countries assisted by some form of deposit insurance, maybe just in the form of the role of lender of last resort played by the Central Bank. As in the real world, in our model investors are unrestricted in deciding the amount to invest in the *numéraire* asset, but they cannot take negative positions as this would more appropriately be considered a different asset, as argued

---

[2]See the criticisms to this assumption made by Lowenstein and Willard [55].

above. Property (iii) may perhaps be seen as the outcome of competition among banks, while (iv) justifies referring to α_0 as the *numéraire*.

We define normalized payoffs as

$$(7) \qquad \bar{X}(\theta) = X(\theta)/X_0$$

2.4. **Evaluating random quantities.** An agent's decision to invest in a given trading strategy $\theta \in \Theta$ is motivated, we assume, by the payoff $X(\theta)$ that it generates. However, a full description of this quantity for each possible future state $\omega \in \Omega$ is perhaps a satisfactory mathematical approach but not necessarily a correct model of how random quantities are evaluated. In other words, economic agents do not regard future outcomes as *functions* but rather as *equivalence classes*. This is clearly the case in expected utility theory and, more generally, in all models in which choice is based on a probability judgment. Equivalence, however, is not only a positive conclusion emerging from an accurate probabilistic assessment, as the classical model implicitly suggests. It often emerges from the inability of individuals to fully compare events or from their attitude to focus attention on scenarios selectively, a fact often documented in empirical decision theory and experimental psychology, and often arising in the form of market exuberance or of overly pessimistic evaluation of risky assets.

According to a well known study by Kahneman and Tversky [50, p. 282], "people are limited in their ability to comprehend and evaluate extreme probabilities" and are lead, as a consequence, to ignore highly unlikely events in order to simplify decision problems. Moreover, due to what the authors call the *isolation effect* (p. 271), in comparing alternatives people *"disregard components that the alternatives share and focus on the components that distinguish them"*. The same idea has been taken up more recently by Bordalo, Gennaioli and Shleifer [10] according to whom decision weights distort probabilities by focusing on *salient* states, i.e. those outcomes on which lotteries differ most. Here too the authors suggest that decisions are made by processing only part of the available information, ultimately, in response to the limits encountered in facing the complexity often implicit in uncertainty[3]. Gennaioli, Bordalo and Schleifer [11] explore the asset pricing implications of this approach to uncertainty, contributing to explain some known puzzles such as preference for skewness and the growth-value puzzle.

This stream of ideas suggests us to treat the equivalence among random quantities, i.e. elements of $\mathfrak{F}(\Omega)$, as a primitive element and to model it via a binary relation $\geq_*$ on $\mathfrak{F}(\Omega)$. We assume to this end[4]:

Assumption 4. *The binary relation $\geq_*$ on $\mathfrak{F}(\Omega)$ satisfies the following properties for $f, g \in \mathfrak{F}(\Omega)$: (i) $f \geq_* g$ if and only if $f - g \geq_* 0$; (ii) $f \geq 0$ implies $f \geq_* 0$; (iii) $f + 2^{-n} \geq_* 0$ for $n = 1, 2, \ldots$ implies $f \geq_* 0$; (iv) $f, g \geq_* 0$ and $a, b \geq 0$ imply $af + bg \geq_* 0$; (v) $f \geq_* 0$ and $A \subset \Omega$ imply $f\mathbf{1}_A \geq_* 0$.*

---

[3]Gennaioli and Shleifer [36] refer to this kind of decision making process as *local thinking*.

[4]We reserve the notation $f \geq g$, $f \vee g$, $f \wedge g$ or $|f|$ to pointwise ordering.

All of these properties are self explaining, save (*iii*) which states that the conclusion $f \geq_* 0$ follows whenever, for all practical purposes, f may be replaced with a quantity which satisfies this criterion.

We present some examples in which this binary relation and its properties are easily interpreted.

Example 1 (Certainty and probability). *One may let $\geq_*$ reflect certainty – by identifying $f \geq_* g$ with $f \geq g$ – or probabilistic sophistication – if, given a (countably additive) probability P, we define $f \geq_* g$ to mean $P(f \geq g) = 1$. In either case Assumption 4 is trivially satisfied.*

Example 2 (Negligible events). *In [16] an ideal $\mathcal{N}$ of subsets of Ω was introduced as a model of those sets that agents consider as negligible. Then the relationship "$f \geq g$ up to negligibility" was defined to mean that $\{f - g < -\eta\} \in \mathcal{N}$ for all $\eta > 0$. It is clear that such partial order satisfies Assumption 4. The notion of a negligible set seems to capture part of the evidence emerging from the above discussion and is best exemplified in the context of partial information. Investments involving a substantial level of credit risk are often evaluated on the basis of some external report on creditworthiness. Although the importance of credit rating is generally fully perceived, the risk of misreporting by the external analysts, originating e.g. from conflict of interests, is often overlooked.*

Example 3 (Qualitative probability). *In his pioneering work on subjective probability, de Finetti [22] introduced the idea of qualitative probability i.e. to model the judgment "event A is more likely than B" as a binary relation, $A \succeq B$ satisfying the axioms: (a) completeness, (b) transitivity, (c) $\Omega \succeq A \succeq \varnothing$ for all events A and (d) if $C \cap A = C \cap B = \varnothing$ then $A \succeq B$ if and only if $A \cup C \succeq B \cup C$. If $f \in \mathfrak{F}(\Omega)$ then one may define*

$$(8) \qquad f \geq_* g \quad \textit{if and only if} \quad A \cap \{f - g > -\eta\} \succeq A \qquad \textit{for all} \quad A \subset \Omega$$

It is easily proved that the above Assumption 4 holds.

Example 4 (Cash-subadditive risk measure). *Let $\rho : \mathfrak{F}(\Omega) \to \mathbb{R}$ be a cash-subadditive, coherent risk measure (as defined by El-Karoui and Ravanelli [30, Definition 3.1, p. 568]) so that ρ is (a) positively homogeneous, (b) subadditive, (c) inversely monotone (i.e. $f \leq g$ implies $\rho(f) \geq \rho(g)$) and such that (d) $\rho(f + \alpha) \geq \rho(f) - \alpha$ when $\alpha \in \mathbb{R}$. Then one may define*

$$(9) \qquad f \geq_* 0 \quad \textit{if and only if} \quad \rho^c(f\mathbf{1}_A) \geq 0 \qquad \textit{for all} \quad A \subset \Omega$$

Upon noting that necessarily $\rho(0) = 0$, it is easily verified that (9) satisfies Assumption 4.

We also define $f >_* g$ to mean that $f \geq_* g$ but $g \not\geq_* f$ (so that $f >_* g$ if and only if $f - g >_* 0$) and $f =_* g$ whenever $f \geq_* g$ and $g \geq_* f$. Eventually we define (with the convention $\sup \varnothing = -\infty$)

$$(10) \qquad f_* = \sup\{\alpha \in \mathbb{R} : f - \alpha \geq_* 0\} \quad \text{and} \quad f^* = -(-f)_*$$

The following facts about the partial order introduced above should be noted:

Lemma 2. *The binary relation $\geq_*$ defines a partial order on $\mathfrak{F}(\Omega)$. Moreover, (i) $f \geq_* 0$ if and only if $f_* \geq 0$; (ii) $f >_* 0$ if and only if $f_* \geq 0$ and $f^* > 0$; (iii) if $f \geq_* 0$, $b \geq 0$ and b is bounded then $bf \geq_* 0$; (iv) if $f =_* 0$ then $|f| =_* 0$; (v) $f \geq_* 0$ implies $\mathbf{1}_{\{f < -\eta\}} =_* 0$ for all $\eta > 0$; (vi) if $a \in \mathbb{R}$ then $(f + a)_* = f_* + a$ and $(f + a)^* = f^* + a$.*

Lemma 2.(iii) together with Assumption 3.(i) leads to the conclusion that $\bar{X} \geq_* 0$ implies $X \geq_* 0$ but the converse need not be true. This complication is a special feature of our model and, notwithstanding the mathematical difficulties it involves, it highlights the role of discounting in determining the overall level of risk. The statement $X \geq_* 0$, in fact, does not exclude losses but rather that losses larger than η are ignored, for any η or, yet in other words, that losses may considered as arbitrarily small. The statement $\bar{X} \geq_* 0$ means, on the other hand, that losses from X may be hedged away by investing an arbitrarily small amount in the *numéraire* asset. However, if the *numéraire* asset does not guarantee a minimum payoff the arbitrarily small losses associated with θ may require a potentially unbounded amount of the *numéraire* asset in order to be hedged. The problem arises whenever losses occur jointly with a low value for X_0, as is often the case during financial crises. We stress that the choice of the *numéraire* has risk management implications, a fact first noted by El-Karoui and Ravanelli [30] and central to their discussion of the cash additivity property of risk measures. The impact of random *numéraires* on risk measures is also discussed in Filipovic [33].

The symbol ba_* (resp. $\mathbb{P}_*$) will henceforth denote the family of finitely additive set functions m (resp. probabilities) defined on all subsets of Ω and such that $f \in L^1(m)$ and $\int |f| dm = 0$ whenever $f =_* 0$. We also define

$$(11) \qquad \mathfrak{B}_* = \{f \in \mathfrak{F}(\Omega) : \eta \geq_* |f| \text{ for some } \eta > 0\} \quad \text{and} \quad \mathcal{N}_* = \{A \subset \Omega : \mathbf{1}_A =_* 0\}$$

Remark that $f \geq_* 0$ implies $\{f < -\eta\} \in \mathcal{N}_*$ for all $\eta > 0$ but the converse is not true.

3. Coherence, Efficiency and Arbitrage

The basic tenet of financial economics is the assumption that markets do not permit arbitrage opportunities and it is the purpose of this section to discuss this fundamental principle in the current framework. In fact, if there is agreement on the general statement of such principle, its translation into a convenient mathematical notion is much less uncontroversial. Definitions vary from one another mainly for what concerns the ambient space, the choice been in general, since Harrison and Kreps [40], some topological space, often L^p, assuming the existence of an exogenously given probability measure.

Given the preceding discussion we propose the following definition:

Definition 1. *A functional $\phi \in \mathfrak{F}(\Theta)$ is said to be coherent (with the no arbitrage principle) if $\bar{X}(\theta) \geq_* 0$ implies $\phi(\theta) \geq 0$; ϕ is said to be efficient if $\theta, \theta' \in \Theta$ and $\bar{X}(\theta) \geq_* \bar{X}(\theta')$ imply $\phi(\theta) \geq \phi(\theta')$. Moreover, $\theta \in \Theta$ is an arbitrage opportunity for ϕ if*

$$(12) \qquad\qquad\qquad \bar{X}(\theta) \geq_* 0 \quad but \quad \phi(\theta) \leq 0$$

and at least one of the two inequalities is strict.

Although with no restrictions to short selling and a linear pricing functional they are equivalent properties, in the general case coherence is weaker then efficiency and it does not guarantee absence of arbitrage *per se* but it only excludes *immediate* arbitrage, i.e. the opportunity to earn money immediately with only negligible losses in the future. It is still possible, with a coherent pricing functional, that an investment which yields a strictly positive (discounted) payoff is sold for free. Coherence is thus a rather basic financial property and we shall investigate it in depth.

The inequality $\bar{X}(\theta) \geq_* 0$ may be rephrased in terms of the minimal amount $\rho(\theta)$ to be invested in the *numéraire* asset in order to hedge losses away (if possible). Formally[5],

$$(13) \qquad \rho(\theta) = \inf\{c > 0 : \bar{X}(\theta + c\delta_0) \geq_* 0\} = (-\bar{X}(\theta)_*) \vee 0$$

By Assumption 3.(ii), $\theta + \rho(\theta)\delta_0 \in \Theta$ if and only if $\rho(\theta) < \infty$ or, equivalently, if θ belongs to the set of hedgeable strategies

$$(14) \qquad \Theta_* = \{\theta : \bar{X}(\theta)_* > -\infty\}$$

In fact regulated markets do not allow investors to enter positions with unlimited potential losses so that Θ_* is often considered as the set of all *reasonable* investment strategies – see [23] where a condition akin to $\theta \in \Theta_*$ is the basis for the concept of *free lunch with vanishing risk*.

The following are some elementary properties:

Lemma 3. *The functional ρ defined in* (13) *satisfies (i) positive homogeneity, (ii) subadditivity, (iii) $\bar{X}(\theta) \geq_* \bar{X}(\theta')$ implies $\rho(\theta) \leq \rho(\theta')$, (iv) $\rho(\theta) = 0$ if and only if $\bar{X}(\theta)_* \geq 0$ and (v)*

$$(15) \qquad \rho(\theta + a\delta_0) \geq \rho(\theta) - a \qquad \text{for all } \theta \in \Theta \text{ and } a \geq 0$$

Moreover,

$$(16) \qquad \rho(\theta) = \sup\left\{\int (-\bar{X}(\theta) \vee 0)d\mu : \mu \in \mathbb{P}_*\right\}$$

If $\phi : \Theta \to \overline{\mathbb{R}}_+$ possesses properties (iv)–(v) above, then $\phi \leq \rho$.

In other words, ρ is a coherent, cash subadditive risk measure on Θ[6] and $\rho(\theta + \rho(\theta)\delta_0) = 0$.

Theorem 1. *Assume that $\theta + \lambda\delta_0 \in \Theta$ for every $\lambda > 0$ and $\theta \in \Theta_*$. The total cost functional t defined in* (5) *is coherent if and only if*

$$(17) \qquad t(\theta) + q_0\rho(\theta) \geq 0 \qquad \text{for all } \theta \in \Theta$$

It admits no arbitrage opportunity if it satisfies

$$(18) \qquad t(\theta) + q_0\rho(\theta) > 0 \qquad \text{for all } \theta \in \Theta \text{ such that } \bar{X}(\theta)^* + \rho(\theta) > 0$$

---

[5]The first equality in (13) is a definition; the second one is easily established.

[6]However the basic intuition used by El-Karoui and Ravanelli to justify cash subadditivity (namely that the discount factor is less than 1, see [30, p. 568]) does not apply as we do not impose $X_0 \geq_* 1$.

As usual, the absence of arbitrage opportunities implies no *a priori* bound on the price of claims with potentially unbounded losses as these cannot be combined in a strategy generating a $\geq_*$ non negative discounted payoff. Likewise, we have no upper bound on prices as a consequence of short selling prohibitions.

By adding convexity to fixed costs we obtain:

Theorem 2. *Suppose that the cost functional c in* (4) *is convex. Then t is coherent if and only if there exists $\mu \in \mathbb{P}_*$ such that*

$$(19) \qquad t(\theta) \geq q_0 \int \left(\bar{X}(\theta) \wedge 0 \right) d\mu \qquad \textit{for all} \quad \theta \in \Theta_*$$

The assumption that t is convex is crucial to prove Theorem 2 but is hard to justify based on the available empirical evidence which suggests, contrariwise, that fixed costs increase less than proportionally.

4. Coherent Pricing

Since the early work of Bensaid et al [5] it is known that many properties of asset prices are revealed by the super hedging functional and our model is no exception. We adapt this concept in defining the following extended real valued functional:

$$(20) \qquad \pi(f) = \inf \left\{ \lambda q(\theta) : \lambda \bar{X}(\theta) \geq_* f, \lambda \geq 0, \theta \in \Theta \right\} \qquad f \in \mathfrak{F}(\Omega)$$

Clearly, $\pi(1) \leq q_0$ and $\pi^c(1) \geq 0$; if, in addition, q is coherent, then $\pi(0) = 0$ and $\pi^c(f) \leq \pi(f)$ for all $f \in \mathfrak{F}(\Omega)$ – see Lemma 9. But even with coherence we cannot exclude the somehow abnormal situations $\pi(1) = 0$ and $\pi^c(1) = 0$ (see Example 6 below). In particular:

Lemma 4. $\pi^c(1) = 0$ *if and only if $q(\theta) \geq 0$ for every $\theta \in \Theta_*$.*

Define

$$(21) \qquad \mathscr{K} = \{f \in \mathfrak{F}(\Omega) : \pi(|f|) < \infty\} \quad \text{and} \quad \mathscr{K}_* = \{f \in \mathscr{K} : f_* > -\infty\}$$

The set $\mathscr{K}$ plays in this paper the role of the ambient space, similarly to $L^p(P)$ in the traditional setting. It should be stressed that its definition is entirely endogenous and market based and does not require the introduction of any mathematical structure.

The following is the most important result of this section.

Theorem 3. *The price functional q is coherent if and only if for each $h \in \mathfrak{B}_*$ the set*

$$(22) \qquad \mathscr{M} = \left\{ m \in ba_{*,+} : \mathscr{K} \subset L(m) \text{ and } \int f dm \leq \pi(f) \text{ for all } f \in \mathscr{K}_* \right\}$$

contains an element m_h such that $\int h dm_h = \pi(h)$.

Abusing slightly the terminology introduced in [16], we refer to $\mathscr{M}$ as the set of *pricing measures*, traditionally represented as *stochastic discount factors*. One immediately deduces from Theorem 3 that $\mathscr{M}$ is a convex set and that

$$(23) \qquad \left\{ \int h\,dm : m \in \mathscr{M} \right\} = [\pi^c(h), \pi(h)] \qquad \text{for all } h \in \mathfrak{B}_*$$

If the *numéraire* is traded with no bid/ask spread and short positions are allowed then $\{\|m\| : m \in \mathscr{M}\} = \{q_0\}$.

As in other papers in this field, Theorem 3 asserts that a coherent price system is consistent with a linear pricing rule, i.e. a pricing rule appropriate for a market free of imperfections, namely

$$(24) \qquad q^m(\theta) \equiv \int \bar{X}(\theta)\,dm \geq q^c(\theta) \qquad \text{for some } m \in \mathscr{M} \text{ and all } \theta \in \Theta_*$$

This conclusion somehow supports the view expressed in the microstructure literature that the bid and ask prices are set starting from a frictionless, linear price, the *consensus price*. It does not justify, as we shall argue later on, the common practice of overcoming market structure issues by computing mid prices.

Given the exclusive emphasis of the literature on *countably additive* pricing measures, in the next result we characterize the additional assumptions needed to this end.

Theorem 4. *Let $\pi^c(1) > 0$ and $\mathscr{A}$ be a σ algebra including $\mathcal{N}$. Write $\mathbb{P}^c_*(\mathscr{A}) = \{m \in \mathbb{P}_* : m \text{ is countably additive on } \mathscr{A}\}$. The following are equivalent:*

(i) $\mathscr{M}$ admits $\mu \neq 0$ which is countably additive in restriction to $\mathscr{A}$;

(ii) there exists $P \in \mathbb{P}^c_*(\mathscr{A})$ such that for all sequences $\langle x_n \rangle_{n \in \mathbb{N}}$ in $\mathbb{R}$ and $\langle f_n \rangle_{n \in \mathbb{N}}$ in $\mathscr{S}(\mathscr{A})$, (a) $\pi(1)x_n + \pi(f_n) \leq 0$, (b) $x_n \to x \geq 0$ and (c) $\int |f_n - f|\,dP \to 0$ imply $\int f\,dP \leq -x$;

(iii) there is $P \in \mathbb{P}^c_*(\mathscr{A})$ such that for all sequence $\langle f_n \rangle_{n \in \mathbb{N}}$ in $\mathscr{S}(\mathscr{A})_+$, $\lim_n \int f_n\,dP = 0$ implies $\lim_n \pi^c(f_n) = 0$.

Theorem 4 may be established without assuming $\pi^c(1) > 0$ upon replacing π with

$$\pi_\varepsilon(b) = \sup\left\{ \int b\,dm : m \in \mathscr{M}, \ m(\Omega) \geq \varepsilon \right\} \varepsilon^{-1} \qquad b \in \mathfrak{B}_*$$

The statement would however be less clear to interpret. Property (iii) clarifies that countable additivity requires some form of continuity of market prices with respect to the $L^1(P)$ norm, a property which has to be assumed exogenously. Condition (ii) in Theorem 4 is a specially weak version of the familiar *No-Free-Lunch* (NFL) of Kreps [51] which however, involves here only simple functions and does not require a pre assigned ambient space.

Theorem 4 provides some insight as well, suggesting cases in which $\mathscr{M}$ may not admit countably additive elements.

Example 5. *Let Ω be a separable metric space and $\mathscr{A}$ its Borel σ algebra. Assume that there exists an increasing net $\langle N_\alpha \rangle_{\alpha \in \mathfrak{A}}$ in $\mathcal{N}_*$ with N_α open and $\Omega = \bigcup_\alpha N_\alpha$. Fix $P \in \mathbb{P}^c(\mathscr{A})$. By*

[8, Proposition 7.2.2], $1 = \lim_\alpha P(N_\alpha) = \lim_k P(N_k)$ *for some suitable sequence* $\langle N_k \rangle_{k \in \mathbb{N}}$ *from* $\langle N_\alpha \rangle_{\alpha \in \mathfrak{A}}$. *Set* $f_k = \mathbf{1}_{N_k^c}$. *Then,* $\lim_k \int f_k dP = 0$ *while*

$$\pi^c(f_k) = \inf_{m \in \mathscr{M}} m(N_k^c) = \inf_{m \in \mathscr{M}} m(\Omega) = \pi^c(1)$$

so that condition (ii) of Theorem 4 fails. No pricing measure is then countably additive outside of the special case $\pi^c(1) = 0$. *Actually, decomposing each* $m \in \mathscr{M}$ *as* $m = m^c + m^\perp$, *with* m^c *countably additive and* $m^\perp$ *purely finitely additive (see* [28, III.7.8]*), and exploiting the inclusion* $\mathscr{M} \subset ba_*$, *we conclude that all pricing measures are purely finitely additive.*

The special situation illustrated in Example 5 highlights that countable additivity of the pricing measures may not only fail but actually contrast with coherence if the partial order $\geq_*$ is an *a priori* of the model.

5. Decompositions of Coherent Prices

Based on the results of the preceding section, we develop here some decompositions of coherent price functionals which highlight the role of asset bubbles.

Theorem 5. *The price* q *is coherent if and only if the set* $\mathscr{M}$ *of pricing measures is the unique non empty, convex, weak* compact subset of* ba_* *admitting the decomposition*

$$(25) \qquad \pi(f) = \beta(f) + \sup_{m \in \mathscr{M}} \int f dm \qquad \text{for all } f \in \mathscr{K}$$

where $\beta : \mathscr{K} \to \mathbb{R}$ *vanishes on* $\mathfrak{B}_*$.

For each $m \in \mathscr{M}$ the quantity $\int \bar{X}(\theta) dm$ is rightfully interpreted as the fundamental value of the portfolio θ *given* m. In order to overcome the arbitrariness implicit in having a multiplicity of possible pricing measures and obtain an unambiguous definition, it is correct to identify the fundamental value of θ with the quantity

$$(26) \qquad \sup_{m \in \mathscr{M}} \int \bar{X}(\theta) dm = \int \bar{X}(\theta) d\gamma$$

where on the right hand side we used the representation of a supremum of integrals as the Choquet integral with respect to a supermodular capacity γ having $\mathscr{M}$ as its core[7]. Differently from classical asset pricing formulas, the fundamental value is not linear here, due to transaction costs. It could be interpreted as the maximum price paid for θ in an economy identical with the one considered above but with no transaction costs. The main point is not only the multiplicity of pricing measures, which would be prevalent even in an economy with incomplete financial markets, but rather the fact that the intervening expectations do not agree on the set of traded payoffs so that the integral appearing in (26) is not invariant with respect to the choice of $m \in \mathscr{M}$.

In general, deviations of prices from fundamental values are interpreted in the literature as evidence of the existence of bubbles. See [20], [45] or [54] for examples of models dealing with bubbles

---

[7]The use of capacities in finance was introduced by Chateauneuf et al. [19, Theorem 1.1] precisely with the aim of modeling transaction costs. In their paper, however, this representation is an assumption (see also [18]).

in continuous time. In so doing, however, inefficiency phenomena and the potential contribution of asset bubbles to an efficient pricing are mixed together.

Inefficiency is measured by the quantity $q(\theta) - \pi(\bar{X}(\theta))$. The empirical literature typically reports a relatively large number of violations, e.g., of the PUT/CALL parity, by which, say, a CALL option may be replaced by a less costly synthetic constructed using the corresponding PUT, future and riskless asset. Luttmer [56], takes this mispricing as the source of subadditivity. For a coherent price system inefficiencies are a consequence of the restrictions which prevent investors to exploit them to obtain immediate profits. Empirical explanations, such as those invoked by Lamont and Thaler [52], draw attention on the fixed costs of trading which impair the arbitrage profits emerging from considering prices only. However, even fixed transaction costs would play virtually no role if investors were not somehow constrained in their ability to either take short positions or in choosing the scale for their investments arbitrarily large.

We deduce from (25) that, even in the absence of market inefficiencies and even with only two dates, prices may differ from fundamental values by a bubble component, identified with β. For a correct interpretation of the economic role of bubbles, let us recall the seminal contribution of Gilles and Leroy [37] who, in an ambient space of bounded sequences interpret deterministic bubbles as the non summable part of the asset payoff. As such, bubbles are in fact the price of the tail part of the asset discounted payoff. We can support the same interpretation here by virtue of the inequality

$$(27) \qquad \left| \beta\left(\bar{X}(\theta)\right) \right| \leq \lim_n \left\{ \pi\left(\left(\bar{X}(\theta)^+ - n\right)^+\right) + \pi\left(\left(\bar{X}(\theta)^- - n\right)^+\right) \right\}$$

By (27), $\beta\left(\bar{X}(\theta)\right)$ is rightfully viewed as the component of the price of θ which only depends on the event $\left\{ \left| \bar{X}(\theta) \right| \geq n \right\}$ for all $n \in \mathbb{N}$, i.e. on the extreme fluctuations of the portfolio discounted payoff. Observe that necessarily the *numéraire* price admits no bubbles. (27) suggests in addition that, like in other models, bubbles are related to the limit of the price of a CALL option as the strike price increases to infinity. This finding is consistent with similar conclusions linking the existence of asset bubbles to some mispricings of options (see [20] and [43]). Assuming some form of monotone continuity of the pricing functional, as in [19], excludes the existence of bubbles.

The following example illustrates the economic role of bubbles in a special case.

Example 6 (Efficient Bubbles). *Consider a market on which $\bar{X}(\theta) \geq_* 0$ for all $\theta \in \Theta$ and assume that $q(\theta) = X(\theta)^* - X(\theta)_*$. The price function is clearly subadditive and positively homogeneous. Given that $q(\theta) \geq 0$ for all $\theta \in \Theta$ it is coherent too although it will be inefficient in general. From Lemma 4 we know that $\pi^c(1) = 0$. Assume that $X_{0,*} = 0$ and $X_0^* = 1$ and that, for each $n \in \mathbb{N}$ there exists $\theta_n \in \Theta$ with $X(\theta_n) = a + X_0 2^{-n}$ with $a > 0$. Then, $\bar{X}(\theta_n) \geq a + 2^{-n}$ so that $\bar{X}(\theta_n)_* \geq a + 2^{-n}$. On the other hand $X(\theta_n)_* = a$ and $X(\theta_n)^* = a + 2^{-n}$. But then $q_0 = 1$ while*

$$(28) \qquad \pi(1) = \inf_{\bar{X}(\theta)_* > 0} q(\theta) / \bar{X}(\theta)_* \leq \inf_n (1 + a2^n)^{-1} = 0$$

Thus the numéraire is priced inefficiently and, from (25), the only possible non null efficient price is a bubble.

6. Option Pricing

In this section we apply our preceding results to option pricing, under the only assumption that option are traded anonymously (i.e. that (6) holds) and at non negative prices. $X > 0$ will be hereafter the payoff of a given underlying and $K(X)$ the set of strike prices (including $k = 0$) of all CALL options written on it. The ticker of each of these options, the corresponding strategy and price will be indicated by $\alpha_X(k) \in \mathfrak{A}$, $\theta_X(k) \in \Theta$ and $q_X(k)$ respectively. Define also

$$(29) \quad \mathfrak{A}_X = \{\alpha_X(k) : k \in K(X)\} \quad \text{and} \quad \Theta_X = \{\theta \in \Theta : \theta \geq 0 \text{ and } \theta(\alpha) = 0 \text{ whenever } \alpha \notin \mathfrak{A}_X\}$$

$$(30) \qquad \pi_X(h) = \inf\left\{\lambda q(\theta) : \lambda \frac{X(\theta)}{X \wedge 1} \geq_* h, \lambda > 0, \theta \in \Theta_X\right\} \qquad h \in \mathfrak{F}(\Omega)$$

and

$$(31) \qquad \mathscr{K}_X = \{h \in \mathfrak{F}(\Omega) : \pi_X(|h|) < \infty\}$$

Observe that the restriction of q to Θ_X is coherent, given our assumption of non negative prices. On the other hand, the change of *numéraire* implicit in (30) entails a different concept of efficiency. In particular we shall say that options are priced efficiently if

$$(32) \qquad q_X(k) = \pi_X\left(\frac{X(\theta_X(k))}{X \wedge 1}\right) \qquad k \in K(X)$$

The criterion adopted in (32) is indeed quite weak as, for example, it does not involve PUT options nor Futures or short positions. This is desirable since the larger the set of derivatives involved the more likely is it that efficiency may fail. For example, the PUT/CALL parity is well known to generate a large number of violations as well as the lower bound for CALL options[8].

Define the set[9]

$$(33) \qquad \Gamma = \left\{f \in \mathfrak{F}(\mathbb{R}_+) : f \geq 0 = f(0), \ f \text{ convex}, \ \lim_{n \to \infty} f(n)/n < \infty\right\}$$

and observe that

$$(34) \qquad g_t^0(x) = (x - t)^+ \qquad x \in \mathbb{R}_+$$

defines an element of Γ for all $t \geq 0$: let G^0 be the corresponding collection. Of course, it is possible to write other derivatives on X possessing some of the properties of options. In particular, we denote by $\mathscr{G}$ the family of all collections $G = \{g_t : t \in \mathbb{R}_+\}$ such that $g_t \in \Gamma$ and

$$(35) \qquad \alpha g_s + (1 - \alpha)g_u \geq g_t \qquad \text{whenever} \quad \alpha \in [0, 1], \ s, t, u \in \mathbb{R}_+ \quad \text{and} \quad \alpha s + (1 - \alpha)u \leq t$$

To start discussing the issue of options efficiency, denote by $J(X) \subset K(X)$ the subset of strike prices k possessing the following property

$$(36) \qquad q_X(k) \leq \inf\{a q_X(k_1) + (1 - a)q_X(k_2) : k_1, k_2 \in K(X), 0 \leq a \leq 1, \ ak_1 + (1 - a)k_2 \leq k\}$$

---

[8]Cerreia et al. [18] construct a pricing model for markets which are assumed to satisfy the PUT/CALL parity.

[9]The limit appearing in (33) exists by convexity.

i.e. which satisfy the *butterfly spread* condition. One should remark that in the present setting this is not an arbitrage restriction. In the rest of this section we shall adopt the following

Assumption 5. *Let* $\left(X\mathbf{1}_{\{X\le j\}}\right)^* = j < X^*$ *for each* $j \in J(X)$.

The first part of Assumption 5 states that the points included in $J(X)$ are not points of discontinuity for X; the second one that each option with strike in $J(X)$ is not out of the money *a priori*.

It turns out that CALL options are efficient if and only if they satisfy the butterfly restriction.

Lemma 5. Θ_X *is priced efficiently if and only if* $K(X) = J(X)$.

Carr and Madan [14], assuming the existence of an infinite collection of strike prices, obtain a result similar to the necessity claim in Lemma 5.

We turn now to the issue of derivatives which may be super hedged exactly. Of course, efficiency is important to this end as super hedging only involves options which are priced efficiently: from this perspective, inefficiency is thus comparable to market incompleteness. The next result shows that derivatives on X with payoff in Γ may hedged perfectly.

Theorem 6. *For each* $G \in \mathcal{G}$ *and* $t \in \mathbb{R}_+$ *there is* $\theta_X^G(t) \in \Theta_X$ *such that*

$$(37) \qquad \frac{X(\theta_X^G(t))}{X \wedge 1} \ge_* \frac{g_t(X)}{X \wedge 1} \qquad and \qquad q_X^G(t) \equiv q(\theta_X^G(t)) = \pi_X\left(\frac{g_t(X)}{X \wedge 1}\right)$$

In Lemma 6 we show how to compute $\theta_X^G(t)$ and $q_X^G(t)$ explicitly. When setting $G = G^0$ as in (34), the corresponding quantities will be indexed by 0 rather than G^0, for simplicity.

The following is the most important result of the paper.

Theorem 7. *Each* $G \in \mathcal{G}$ *admits* $\beta^G(X) \ge 0$ *and* $\nu_X^G \in ca(\mathscr{B}(\mathbb{R}_+))_+$ *such that*

$$(38) \qquad q_X^G(t) = \beta^G(X) + \int_t^\infty \nu_X^G(x > z)dz \qquad for\ all\ t \ge 0$$

Observe that, by standard rules,

$$(39) \qquad \int_t^\infty \nu_X^G(x > z)dz = \int (x - t)^+ d\nu_X^G(x)$$

Thus (38), represents the price of the G derivatives as the sum of a bubble part and the fundamental value. Observe that necessarily

$$(40) \qquad \beta^G(X) = \lim_{t\to\infty} q_X^G(t)$$

so that, upon choosing $G = G^0$, the term $\beta^0(X)$ represents the option price as the strike approaches infinity and contributes to explaining the overpricing of deeply out of the money CALL's often documented empirically in some form of the smile effect.

The representation obtained in (38) is akin in principle to (25) but much clearer. The difference is that the option portfolio $\theta_X^G(t)$ is priced efficiently by construction. One deduces immediately the following:

Corollary 1. *Option prices are efficient if and only if there are $\beta^0(X) \geq 0$ and $\nu_X^0 \in ca(\mathscr{B}(\mathbb{R}_+))_+$ such that*

$$(41) \qquad\qquad q_X(k) = \beta^0(X) + \int_k^\infty \nu_X^0(x > z)dz \qquad \text{for all } k \in K(X)$$

In Corollary 1 sufficiency is an easy consequence of the fact that the pricing formula (41), which holds in general only if $k \in J(X)$, is convex in the strike price and of Lemma 5.

The first remarkable implication of Theorem 7 is that it provides a fully explicit CALL function, q_X^G, and that this is entirely market based. In empirical analyses the estimation of the CALL function, dictated by optimal statistical criteria, does not guarantee a direct market interpretation, although estimates may possess several additional regularity properties that q_X^G may fail to exhibit. A second fact arising from (38) is the representation of option prices via a *countably additive* measure ν_X^G to which we refer as the probability implicit in option prices although it is clear that this will in general not be unique but may vary with the choice of the collection G. While such multiplicity does not diminishes the significance of Theorem 7 it would be desirable to investigate the role of G more in depth.

Given (38) we may compute

$$(42) \qquad\qquad \nu_X^G(x > t) = - \left.\frac{dq_X^G(k)}{dk}\right|_{k=t} \qquad \text{for all } t \geq 0$$

as originally suggested by Breeden and Litzenberger [13] and Banz and Miller [4]. Let us remark, for the sake of comparison and clarity, that in the model of Black and Scholes, equation (42) gives the familiar result

$$(43) \qquad \nu_X^{BS}(x > k) = e^{-rT}\Phi(d_2) \quad \text{with} \quad d_2 = \frac{\ln(S_0/k) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

so that $\|\nu_X^{BS}\| = \exp(-rT)$. In more general, traditional models, comparing the investment in a CALL option with strike k and in a zero coupon bond with an investment in the underlying, efficiency implies the classical bounds for the CALL price, $q_X \geq q_X(k) \geq q_X - kq_0$, from which follows the inequality $\|\nu_X\| \leq q_0 \leq 1$. However, in a model in which the *numéraire* asset is not free of risk such constraints on option prices need no longer hold. In particular, we do not have any *a priori* bound to impose on the norm of ν_X. If, however, we assume in addition $X \geq_* \eta$, we then obtain

$$\|\nu_X^0\| = \nu_X^0(x > 0) = \lim_{k \to 0} \frac{q_X^0(0) - q_X^0(k)}{k} \leq \pi_X(1) \leq \frac{q_X(0)}{\eta}$$

A final remark concerns the general treatment of price inefficiency and of the bid/ask spread in empirical analysis. In most studies a great amount of preliminary work is devoted to cleaning the dataset from price quotes which utterly violate some basic arbitrage opportunity, often amounting up to 30% of the sample. Theorem 7 illustrates an approach which allows to obtain the same result endogenously with no preliminary purging of the data. In this respect our method is similar to that of Aït-Sahalia and Duarte [1, pp. 16-18] who replace the original prices by weighted least squares approximations under shape restrictions. For what concerns the bid/ask spread, the common

approach is to restore price uniqueness by working with mid prices. We argue that this may in principle introduce distortions in the analysis. To wit, if we assume that the options market is efficient and given the inclusion $\Theta_X \subset \Theta_*$, one concludes from (24) that for each $\theta_X(k) \in \Theta_X$ the corresponding mid price may be represented as a risk neutral price $q^{m(k)}(\theta_X(k))$ for some $m(k) \in \mathscr{M}$[10]. However, microstructure effects may induce $m(k)$ to depend on k in a relevant way so that the derivative of the mid option price with respect to the strike may not distinguish between the probability of the event $\{X > k\}$ and the distortion of the intervening pricing measure as the strike changes. As a matter of fact, market makers adjust the bid/ask spread in response to market pressure. This is the so-called inventory and adverse selection effects, well documented in the microstructure literature. Huang and Stoll [44], e.g., explicitly model the impact of past trading on changes of the mid prices[11]. Another recent paper addressing the structure of the inventory holding cost is Bollen, Smith and Whaley [9], who document the impact of liquidity on such costs through the expected time to offsetting trades.

7. Estimating the CALL function

The first step in the direction of implementing Theorem 7 empirically is to choose the collection $G \in \mathscr{G}$; the second, to compute the quantity $q_X^G(t)$ explicitly. We solve the latter first.

Lemma 6. *Let Assumption 5 hold, choose $f \in \Gamma$ and write (i) $\hat{f} = f(X^*)$ and $d_i = (X^* - j_i)$ if $X^* < \infty$ or else (ii) $\hat{f} = \lim_n f(n)/n$ and $d_i = 1$. Write also $j_{I+1} = X^*$*

$$
(44) \qquad \mathbf{q} = \begin{bmatrix} q_X(j_0) \\ q_X(j_1) \\ \vdots \\ q_X(j_I) \end{bmatrix}, \qquad \mathbf{f} = \begin{bmatrix} f(j_1) \\ f(j_2) \\ \vdots \\ f(j_I) \\ \hat{f} \end{bmatrix}, \qquad \mathbf{D} = \begin{bmatrix} (j_1 - j_0) & 0 & \ldots & 0 \\ (j_2 - j_0) & (j_2 - j_1) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ (j_I - j_0) & (j_I - j_1) & \ldots & 0 \\ d_0 & d_1 & \ldots & d_I \end{bmatrix}
$$

and $\mathbf{w} = \mathbf{D}^{-1}\mathbf{f}$. The following programs are mutually equivalent

$$
(45) \qquad \min_{\{\theta \in \Theta_X, \ \lambda > 0\}} \lambda q(\theta) \ \text{subject to} \ \lambda \frac{X(\theta)}{X \wedge 1} \geq_* \frac{f(X)}{X \wedge 1}
$$

$$
(46) \qquad \min_{\mathbf{a} \in \mathbb{R}_+^{I+1}} \mathbf{q}^T \mathbf{a} \ \text{subject to} \ \mathbf{D}\mathbf{a} \geq \mathbf{f}
$$

and admit as their unique solution

$$
(47) \qquad q_X(f) = \mathbf{q}^T \mathbf{w} \quad \text{and} \quad \theta_X(f) = [\theta_X, \theta_X(j_1), \ldots, \theta_X(j_I)]\mathbf{w}
$$

Moreover, $\mathbf{w}[1] = f(j_1)/j_1$ and

$$
(48) \qquad \mathbf{w}[i] = \sup_{x \in \mathbb{R}_+, x \leq j_i} \frac{f(x) - f(j_{i-1})}{x - j_{i-1}} - \frac{f(j_{i-1}) - f(j_{i-2})}{j_{i-1} - j_{i-2}} \qquad i = 2, \ldots, I+1
$$

---

[10]Efficiency is necessary to guarantee that the mid price falls in the interval $[\pi(\bar{X}(\theta_X(k))), \pi^c(\bar{X}(\theta_X(k)))]$

[11]See e.g. equation (3), p. 1000.

For what concerns the choice of G, it is natural to start with the family G^0. By construction $q_X^0(k)$ represents the efficient price of the corresponding option and so it coincides with $q_X(k)$ if and only if $k \in J(X)$. More generally one sees that the function q_X^0 is the highest among the positive, convex and decreasing curves passing through the knots $\{(j, q_X(j) : j \in J(X)\}$. Aït-Sahalia and Duarte [1] obtain the same values in the first step of their approach by exploiting a least square projection technique due to Dykstra [29]. On the data so transformed they apply then a local polynomial regression, a methodology which generalizes suitably the kernel by replacing constant functions with polynomials of arbitrary but preassigned degree.

By Lemma 6 we obtain

$$\theta_X^0(k) = \sum_{i=0}^{I-1} b_i(j_{i+1} - k)^+ + b_I \hat{g}_k^0$$

where the vector $\mathbf{b}$ is defined as in (74) and either $\hat{g}_k^0 = (X^* - k)^+$ if $X^* < \infty$ or else $\hat{g}_k^0 = 1$. For definiteness let's assume $X^* = \infty$. Then, by (74),

$$(49) \quad \nu_X^0(x > t) = \frac{q_X(j_n) - q_X(j_{n+1})}{j_{n+1} - j_n} \quad \text{if} \quad j_{n+1} > t \geq j_n \qquad \text{or} \qquad \nu_X^0(x > t) = 1 \quad \text{if} \quad t \geq j_I$$

In other words, the implied set function ν_X^0 coincides – not surprisingly – with the right derivative of the price function over the discrete set $J(X)$ upon a change of sign.

Despite being a theoretically exact formula, the informational content of (49) is indeed quite poor empirically, as a consequence of the limited number of options which are priced efficiently by the market – the size of $J(X)$. In fact $\nu_X^0(x > t)$ remains constant within two adjacent strike prices in $J(X)$ or, equivalently, the efficient CALL function $q_X^0(k)$ is piecewise linear. Adding the fictitious efficient prices $q_X^0(k)$ to the original data set, thus, does not improve our knowledge of ν_X. A quick look at (47) reveals that this follows from options payoffs being piecewise linear.

Obtaining smooth estimates is the goal of all econometric exercises. Although the review of the different approaches followed in the literature is beyond our scopes (interested readers should refer to Garcia, Ghysels and Renault [35]), we have a special interest here for nonparametric methods, nicely reviewed by Yatchew [67].

A forerunner of this stream of studies is the paper by Jackwerth and Rubinstein [46] where the parameters of a binomial tree are set so as to minimize the distance with actual option prices given a penalty for deviations from an initial *a priori* distribution. The absence of arbitrage assumption is absolutely crucial here and estimating the option price produces simultaneously an estimate of the risk neutral measure. Aït-Sahalia and Lo [2] estimate the CALL function using the Nadaraya-Watson kernel and recover the risk-neutral density via the second derivative. This methodology has optimal asymptotic properties but relatively poor performance in small samples, a fact common to many non parametric methods and of special concern for options market. The curse of dimensionality problem, see [67, p. 675], requires to limit as much as possible the number of state variables. The risk neutral density estimated via kernel smoothing, e.g., need not be non negative in small samples [2, footnote 11, p. 508]. To circumvent this problem the

authors propose a semi nonparametric approach in which the price is computed according to the Black and Scholes formula in which the volatility function is estimated non parametrically. This choice has the clear advantage of guaranteeing the correct shape of the CALL function. Shape restrictions are easily accommodated in parametric modeling but are much more troublesome in the non parametric approach. Papers implementing the nonparametric methodology with shape restrictions are less numerous and more recent. Further to Aït-Sahalia and Duarte [1] another example of non parametric techniques incorporating shape restrictions is the paper by Yatchew and Härdle [68] who follow a least squares approach in Sobolev spaces. Among papers applying spline techniques to option pricing one should mention Fengler [32] and Yin, Wang and Qi [69].

Although the method followed hereafter is nonparametric, it is radically different from those adopted by the aforementioned papers as we do not look for smooth estimates of option prices but rather exact, super hedging prices of smooth derivatives on X. The advantage in doing so, we believe, is that our estimates maintain a stronger than usual economic interpretation[12]. Our first step is then to construct a family $G^h \in \mathscr{G}$ of derivatives written on X each having a payoff, denoted by $g_k^h(X)$, (i) conveniently close to the corresponding CALL option but (ii) twice continuously differentiable. The latter property guarantees that the corresponding price $q_X^h(k)$ will be a smooth function. The class G^h of derivatives will depend on a control parameter, $h > 0$, which acts in much the same way as the bandwidth parameter in kernel regression. We give now a more detailed description.

7.1. **The General Approach.** Let $k, h > 0$ be given. Divide the interval $[-h, h]$ into $N - 1$ intervals of equal length, with endpoints $T^h = \{-h = t_1^h < \ldots < t_N^h = h\}$, the set of knots. Let $T_k^h = \{k + t : t \in T^h\}$. Consider then the following functional (with $D^2 f$ denoting the second derivative of f):

$$(50) \qquad I_k(h; f) \equiv \sum_{i=1}^{N} \left[g_k^0(k + t_i^h) - f(k + t_i^h) \right]^2 + \lambda h^3 \int_{k-h}^{k+h} \left[D^2 f(x) \right]^2 dx$$

and the program

$$(51) \qquad I_k(h) \equiv \min_{f \in \Gamma \cap \mathscr{C}^2} I_k(h; f)$$

It is well known, see [31, Theorem 5.2], that a solution to this problem is given by a $\mathscr{C}^2$ cubic spline which is linear outside of $[k - h, k + h]$. Based on the fact that the second derivative of a cubic spline is locally linear, the infinite dimensional problem (51) conveniently reduces to a $(2N - 2)$-dimensional one. Turlach [66] developed a methodology to compute its solution under several shape restrictions such as (i) $f, Df, D^2 f \geq 0$, i.e. positivity, monotonicity and convexity. To these constraints we add the following: (ii) $f(x) = Df(x) = 0$ for $x \leq k - h$, (iii) $f(k + h) = g_k^0(k + h)$

---

[12]An exception to this conclusion is the XMM methodology proposed by Gagliardini, Gourieroux and Renault [34] where the GMM is applied with the additional constraint of reproducing a subset of given prices which are considered *a priori* to be correct

and (iv) $Df(x) = 1$ for $x \geq k + h$. Denote

(52) $\chi(k;h) = \left\{ f \in \mathscr{C}^2(R_+) : f \text{ cubic spline meeting the constraints } (i) - (iv) \right\}$

To our purposes the choice of the set of knots plays little role, so we fix $N = 5$. The existence of a constrained solution to (51) and the properties of such solution are proved in the next:

Lemma 7. *The problem* (51) *under constraints (i)–(iv) admits one and only one solution, denoted by* $g_k^h \in \chi(k;h)$. *Let* $G^h = \{g_k^h : k \in \mathbb{R}_+\}$. *Then,* (i) $g_{k+m}^h(x) = g_k^h(x - m)$ *for all* $k, m, x \geq 0$ *(so that* G^h *satisfies* (35)), (ii) $g_k^0 \leq g_k^{h'} \leq g_k^h$ *whenever* $h' \leq h$ *and* (iii) $\lim_{h \to 0} \sup_x (g_k^h - g_k^0)(x) = 0$.

Actually, in the proof of Lemma 7 we obtain in addition

(53) $g_k^{h'}(x) = g_k^h\left(k + (x - k)\frac{h}{h'}\right)\frac{h'}{h}$

so that property (iii) follows easy from

(54) $0 \leq q_X^h(a) - q_X^0(a) \leq \pi_X\left(\frac{g_a^h(X) - g_a^0(X)}{X \wedge 1}\right) \leq \frac{g_a^h(a)}{(a - h) \wedge 1}q_X(0) \leq \frac{g_a^{h_0}(a)}{(a - h_0) \wedge 1}\frac{q_X(0)}{h_0}h$

The content of Lemma 7 are clearly illustrated in Figure 1, Panels A and B.

[Figure 1 about here.]

Figure 1, Panel C illustrates the classical trade-off between smoothness and goodness of fit associated with the choice of the bandwith h. A large value of h makes the smoothing interval $[k - h, k + h]$ larger and increases, as a consequence, the distance between g_l^0 and g_k^h. At the same time the number of efficiently priced options with strike price included in such interval, i.e. the size of the set $[k - h, k + h] \cap J(X)$, grows. This makes the super hedging of the corresponding payoff more smooth, as is clearly illustrated by the piecewise linear curves dominating the corresponding payoff. In particular, in Panel C of Figure 1 we assume to have 6 equally spaced efficient strike prices ranging from 95 to 105. Thus, in the case $h = 2$, the super hedging portfolio only contains 3 options while for the case $h = 10$ it contains all 6. As a general rule the super hedging portfolio contains 1 element more than the set $[k - h, k + h] \cap J(X)$. This suggests to choose h as a function of the mesh M of the set $J(X)$. In particular we suggest $h = \delta M$ with $\delta \geq 2$ so that super hedging will involve at least 3 options. A further advantage in so doing is that this guarantees uniform convergence of g_k^h to g_k^0 as, by (53),

7.2. **Properties of the Estimator.** A different and more classical question, somewhat extraneous to our model but crucial to the empirical option pricing literature, is whether the proposed estimator converges to the *true* price function in the presence of disturbance. The classical formulation, which requires a reference probability measure P on a σ algebra $\mathscr{F}$ of subsets of Ω, is

(55) $q_X(k_i) = F(k_i) + \varepsilon_i \qquad k_i \in K(X)$

where F is the model and is often taken to be a positive, decreasing, convex, $\mathscr{C}^2$ function whereas the errors ε_i are identically and independently distributed with zero expectation.

Lemma 6 gives an explicit functional form to our estimator:

$$(56) \qquad q_X^h(a) = \sum_{j \in J(X)} w_a^h(j) q_X(j)$$

with $w_a^h(j)$ an element of the vector $\mathbf{w}$ obtained after choosing $f = g_a^h$. Thus, $w_a^h(j) \geq 0$ and, at least in the case in which $X^* = \infty$ as we assume henceforth, $\sum_{j \in J(X)} w_a^h(j) = 1$.

At first sight (56) defines a local averaging estimator for which a well established theory suggests (see e.g. [67, p. 677]) that the *MSE* converges to 0 as the sample size diverges. However the sample $J(X)$ on which it is based is obtained after removing from the original sample $K(X)$ those prices which are not efficient. This selection mechanism introduces a strong form of correlation between weights and disturbances. A high value of ε_i makes the price $q_X(k_i)$ inefficient, affecting the selection. Not only, but a high value of $\varepsilon_{i'}$ is likely to make option with strike k_i relatively more efficient and thus to increase the value of the corresponding weight. After Heckman [42] it is known that sample selection may lead to biased and inconsistent estimates unless correcting for a term which proxies the expected value of disturbances conditional on selection. The same issue has been recently addressed by Das et al. [21] in the nonparametric setting[13].

One easily deduces, from comparison with [21, equation (2.6)], that a special feature of (56) is that the selection of an option price depends not just on variables pertinent to that option but to the whole sample. On the other hand, the weights $w_a^h(i)$ are just a function of the subsample $J(X)$, (48). It is thus natural to follow a two step procedure (similarly to Heckman and to Das et al.) by first conditioning variables on the selection mechanism and then focusing on the unconditional properties. It is useful to remark to this end that for every subset z of $K(X)$ the condition $J(X) = z$ identifies a measurable set, given the selection mechanism (36). Let $\mathscr{A} \subset \mathscr{F}$ be the corresponding (finite) σ algebra and denote by $P_z(\cdot)$ the conditional expectation given the event $J(X) = z$, for $z \subset K(X)$[14]. The variable $P_z(\varepsilon)$ will be our correction term.

Theorem 8. *Let $F : \mathbb{R}_+ \to \mathbb{R}_+$ in (55) be of class $\mathscr{C}^2$ and assume that $P(\varepsilon_i) = P(\varepsilon_i \varepsilon_j) = 0$ for $i, j = 1, 2, \ldots$ with $i \neq j$ and*

$$(57) \qquad P_z(\varepsilon_i) = P_z(\varepsilon_{i'}) \qquad P\left(\sup_i P_z(\varepsilon_i^2)\right) < \infty \qquad P_z(\varepsilon_i \varepsilon_j) = P_z(\varepsilon_{i'} \varepsilon_{k'})$$

for all $z \subset K(X)$ and $i, i', k, k' \in z$. Then, for any compact interval $I \subset \mathbb{R}_+$

$$(58) \qquad \lim_{M \to 0} \sup_{a \in I} P\left((q_X^h(a) - F(a))^2\right) = 0$$

---

[13]Let us mention incidentally that, although the sample selection issue is fully apparent in our approach, it is not less important in virtually all other approaches to empirical option pricing in which, before moving to estimates, data are filtered to remove or transform observations which utterly violate some of the shape restrictions dictated by the theory. The sample selection problem is generally dismissed by presenting data cleaning as totally unrelated to the property of the model.

[14]The symbol $P(f)$ will denote hereafter the expectation of f with respect to P.

Condition (57).(i) is akin to the assumption made by Das et al. that the conditional expectation of the disturbance given selection depends only on the propensity score, see [21, Assumptions 2.1, 2.2 and 2.3]. Condition (57).(ii) is easily satisfied if, e.g., disturbances are uniformly distributed. Thus the most delicate property in (57) seems to be the last one. Remark, eventually, that an increase in the sample size, i.e. in the size of the set $K(X)$, does not guarantee in general that the size of $J(X)$ increases as well and thus that $M \to 0$.

8. Empirical Applications

In this section we are going to test our methodology in three different ways. First, we consider the classical model of Black and Scholes and verify how close we get to it for a given choice of h. The exercise is then repeated upon adding a liquidity smile. Second, introducing a simulated, noise component we construct then confidence intervals. Eventually, we analyze briefly market data.

8.1. **Black and Scholes.** We set – as in Aït-Sahalia and Duarte [1] – the current underlying price at $1,365$, maturity equal to 3 months, the interest rate at 4% and assume no dividends. We consider 25 strike prices equally spaced between $1,000$ and $1,700$, so that $M = 28$. We use the same values of that paper for reasons of comparison. We experiment three possible values of the smoothing parameter $h = \delta M$, namely $\delta = 2$, $\delta = 5$ and $\delta = 10$. Volatility is set at 40%. The CALL function obtained and the implicit probability extracted from it are plotted in Figure 2. In panel A we represent the difference $q_X^h(k) - q_X^{B\&S}(k)$ and note that, for all choice of h, the smooth option price is indeed quite near to the Black and Scholes one (which ranges from 388.54 to 42.38). In Panel B we consider the same approximation from the perspective of implied volatilities. The approximation is even more precise for what concerns the implicit probability function, Panel C, with the possible exception of the values around the extrema of the interval, where the options available for hedging is forcefully very limited. The curve for $\delta = 2$ is plotted on the right hand axis.

[Figure 2 about here.]

The situation partly changes upon introducing a smile effect into the data. In particular, we let volatility decrease linearly with k, from 40% to 20%. Of course, the Black and Scholes implicit probability, $\nu^{B\&S}$, changes accordingly to incorporate the effect of the strike on volatility. The comparison is described in Figure 3. The price associated with the smooth options is now slightly farer away from the true price, particularly for options just out of the money. This phenomenon, quite limited, soon disappears when we move from $\delta = 10$ to $\delta = 5$. Deviations from the Black and Scholes benchmark are more interesting when considering the implicit probability as the smooth functions (deviations) lie below 0 up to 1460 for all choices of h.

[Figure 3 about here.]

8.2. **Monte Carlo Simulation.** Still inspired by Aït-Sahalia and Duarte [1], we consider, on the top the Black & Scholes model with smile, a bid ask structure with noise. In particular, we fix the

basis spread equal to 10% of the option price with a floor at 50 cents and a cap at 2 dollars. The resulting price is denoted by q^a and represents the correct ask price. We proxy liquidity of options via the factor $\ell(k) = 1 + 5|\exp(r\tau)k/S_0 - 1|$ which is 1 for options exactly at the money and gets as high as 2.3 for deeply in and out of the money options. Liquidity is multiplied by a noise factor which is uniformly distributed on the interval centered at the origin and with width equal to the spread. The noise reflects the random arrival of flow orders to market makers inducing them to modify the theoretical ask and bid price to reflect market pressure. We simulate 5,000 vectors of 25 noisy prices, one for each quoted strike. For each $s = 1, \ldots, 5000$ the resulting price is thus $q_X^s(k) = q_X(k) + 0.5spread(1 + \varepsilon_s \ell(k))$

The introduction of noise produces a number of arbitrage violations which on average amount to 25% of the sample but ranges up to 48% and is above 30% in the 20% of cases. For any simulation $s = 1, \ldots, 5000$ we compute, via Lemma 6, the corresponding CALL function $q_X^{h,s}(k)$ and the associated probability $\nu_X^{h,s}$ for different values of h, although we only plot $h = M$. For each strike price k, we then compute the 2.5%, the 97.5% quantiles and the mean for the simulated samples $\{q_X^{h,s}(k)\}_{s=1}^{5000}$ and $\{\nu_X^{h,s}(x > k)\}_{s=1}^{5000}$. In Panel A of Figure 4 we plot the results so obtained for the case $\delta = 2$ as differences from the theoretical price q^a. The corresponding outcomes for the implied volatility and probability are plotted (in absolute terms) in Panel B and C. It is noteworthy that the true value always falls inside the confidence interval, suggesting that our approach produces quite reliable predictions. One may also notice that the two confidence bounds are not symmetric, especially for options deep in the money, reflecting a sort of negative skewness. The same plots relatively to the choice $\delta = 10$ appear in Figure 5.

[Figure 4 about here.]

Comparing Figures 4 and 5 illustrates how a high value of δ determines smoother estimates and a narrower confidence interval, although the distance with respect to the model increases as well.

[Figure 5 about here.]

In the literature much emphasis is on the state price density rather than on the implicit probability so that the focus is actually on the second derivative of the CALL function. We do not pursue this exercise here, partly because the presumption that a density actually exists has no financial basis. In part, however, it is our choice to work with cubic splines that limits our ability to explore densities. Second derivatives in fact exist but are piecewise linear, making the candidate density function not particularly interesting for applications. Working with splines of higher order than three is of course a possible way out but would certainly make our results much less manageable.

8.3. **Market Data.** Eventually, we consider an application to market data by selecting an arbitrary trading day – 21st October 2010 – on the S&P 500 options market[15]. We sample ask quotes at time intervals of one minute each and disregard quotes for which the reported ask size is below

---

[15]We make use of the quote prices provided by CBOE Market Data Retrieval (MDR). The dataset contains, among other things, information on bid and ask prices and volumes. Data are sampled at a frequency of approximately 1 minute

100. On the subsample so obtained we have options quotes for 180 different strike prices – ranging from 50 to 2500 – and 12 possible maturities – from 22nd October 2010 to 22nd December 2012. We focus on options expiring in November 2010 and their quotes at 12:06,12:39 and 13:03, around the market downturn of Figure 6.

[Figure 6 about here.]

At the three selected times there are 65, 53 and 67 efficiently quoted strike prices respectively out of 90, 75 and 99. We therefore have a relatively long cross section of strikes but a rather high incidence of arbitrage violations (on average 30%). The large number of available strikes is one of the advantages of working with quoted ask prices, as dictated by our model, rather than transaction prices. We select the quoted strikes so to have the same support, ranging from 670 to 1255. At 12:39 the lesser number of strikes quoted corresponds to a larger maximum interval between consecutive strikes, i.e. $M = 70$, while at the other moments strikes do not differ by more than 25 and 30, respectively. These values influence our choice of the bandwidth parameter h for which we select three different values: $h = 70, 40, 10$. As argued above, the smaller h the less smooth will be the resulting curve. In particular, smoothness breaks down if for some x the interval $[x - h, x + h]$ does not contains a quoted strike, a situation which occurs whenever $h < M/2$. This is the case for $h = 10$ and $t = 12 : 39$. In the following picture we plot for each time t the curves corresponding to the three distinct values of h.

[Figure 7 about here.]

It is clear that the loss of smoothness, as captured by the difference among the black and the green lines, is stronger at 12:29 when the distance between quoted strikes is larger.

Eventually we plot, for the value $h = 70$, the first and last curve together in Figure 8

[Figure 8 about here.]

to capture the effect of the market downturn on ν. The downward shift of the curve reflects the impact of the changes of the underlying on the implied probability. In particular, the change of $q(X)$ from 1181.59 to 1172.57 drives the probability of ending up in the money by, say, 10 dollars from 42.4% to 48.2%.

These last remarks suggest to investigate the dependence of ν_X on the current value of the underlying, although outside of our interests here. Another issue that would be important to address is the time evolution of the pricing measure.

Appendix A. Mathematical Appendix

In this appendix we present some results which we used in the proofs of the main Theorems.

Auxiliary Results

Let us start with two general results.

Theorem 9. *Let $\Gamma \subset \mathfrak{F}(\Omega)$ be a convex cone and define $\Gamma_* = \{f \in \Gamma : f_* > -\infty\}$. Then $\sup_{f \in \Gamma} f_* \leq 1$ if and only if there exists $m \in \mathbb{P}_*$ such that*

$$(59) \qquad \Gamma \subset L^1(m) \quad and \quad \sup_{f \in \Gamma_*} \int f \, dm \leq 0$$

Proof. By assumption there is no $h \in \Gamma_1 = \{g \in \mathfrak{F}(\Omega) : f \geq_* g \text{ for some } f \in \Gamma\}$ such that $h \geq 1$. Thus by [17, Proposition 1], there exists a finitely additive probability m defined on the power set of Ω and such that $\Gamma_1 \subset L^1(m)$ and $\int h \, dm \leq 0$ for all $h \in \Gamma_1$ such that $h > -b$ for some $b \in \mathbb{R}$. $N \in \mathcal{N}_*$ implies $\mathbf{1}_N \in \Gamma_1$ so that $m(N) \leq 0$ and $m \in ba_*$. On the other hand $f \in \Gamma_*$ implies $f_n = f\mathbf{1}_{\{f \geq f_* - 2^{-n}\}} \in \Gamma_1$ so that $\int |f| dm = \int |f_n| dm < \infty$, i.e. $f \in L^1(m)$; moreover, $\int f \, dm = \int f_n dm \leq 0$. For the converse, suppose that $f \in \Gamma$ and $f_* > 1$. Then $\{f \leq 1/2\} \in \mathcal{N}_*$ and $f \in \Gamma_*$ so that $0 \geq \int f \, dm = \int_{\{f > 1/2\}} f \, dm \geq 1/2$, a contradiction. $\qquad\square$

Lemma 8. *Let $\mathscr{L} \subset \mathfrak{F}(\Omega)$ be a vector lattice containing $\mathfrak{B}$. Each positive linear functional ϕ on $\mathscr{L}$ admits the decomposition*

$$(60) \qquad \phi(f) = \phi^\perp(f) + \int f \, dm_\phi \qquad f \in \mathscr{L}$$

where $m_\phi \in ba$ and $\phi^\perp$ is a positive linear functional vanishing on $\mathfrak{B}$.

Proof. See [17, Theorem 1]. $\qquad\square$

Theorem 10. *Let $\mathscr{L} \subset \mathfrak{F}(\Omega)$ be a vector lattice containing $\mathfrak{B}_*$, $C \subset \mathscr{L}$ a convex set containing the origin and $\gamma : \mathfrak{F}(\mathscr{L})$ be $\geq_*$-monotone, subadditive and positively homogeneous. Then*

$$(61) \qquad \gamma\left(\sum_{n=1}^N f_n\right) = \sum_{n=1}^N \gamma(f_n) \qquad f_1, \ldots, f_N \in C$$

if and only if there exist (i) a positive linear functional β on $\mathscr{L}$ vanishing on $\mathfrak{B}_*$ and (ii) $m \in ba_{*,+}$ such that $\mathscr{L} \subset L^1(m)$

$$(62) \qquad \gamma(h) \geq \beta(h) + \int h \, dm \quad and \quad \gamma(f) = \beta(f) + \int f \, dm \qquad for \ all \ h \in \mathscr{L}, \ f \in C$$

Proof. (61) holds on C if and only if it holds over the whole convex cone generated by C, by positive homogeneity and the inclusion $0 \in C$. Let $f_1, \ldots, f_N, g_1, \ldots, g_K \in C$ and $\lambda_1, \ldots, \lambda_N, \alpha_1, \ldots, \alpha_K \in \mathbb{R}$ be such that $\sum_{k=1}^K \alpha_k g_k =_* \sum_{n=1}^N \lambda_n f_n$. Then, $\sum_{k=1}^K \alpha_k^+ g_k + \sum_{n=1}^N \lambda_n^- f_n =_* \sum_{k=1}^K \alpha_k^- g_k + \sum_{n=1}^N \lambda_n^+ f_n$. By (61) and $\geq_*$ monotonicity

$$\sum_{k=1}^K \alpha_k^+ \gamma(g_k) + \sum_{n=1}^N \lambda_n^- \gamma(f_n) = \gamma\left(\sum_{k=1}^K \alpha_k^+ g_k + \sum_{n=1}^N \lambda_n^- f_n\right)$$

$$= \gamma\left(\sum_{k=1}^K \alpha_k^- g_k + \sum_{n=1}^N \lambda_n^+ f_n\right)$$

$$= \sum_{k=1}^K \alpha_k^- \gamma(g_k) + \sum_{n=1}^N \lambda_n^+ \mathbf{1}(f_n)$$

i.e. $\sum_{k=1}^{K} \alpha_k \gamma(g_k) = \sum_{n=1}^{N} \lambda_n \gamma(f_n)$. Thus the quantity

$$\phi_0 \left(\sum_{n=1}^{N} \lambda_n f_n \right) = \sum_{n=1}^{N} \lambda_n \gamma(f_n) \qquad f_1, \ldots, f_N \in C, \ \lambda_1, \ldots, \lambda_N \in \mathbb{R}$$

implicitly defines a linear functional on the linear span $\mathrm{Lin}(C)$ of C. It is easy to conclude from (61) and subadditivity that

$$\phi_0 \left(\sum_{n=1}^{N} \lambda_n f_n \right) = \phi_0 \left(\sum_{n=1}^{N} \lambda_n^+ f_n \right) - \phi_0 \left(\sum_{n=1}^{N} \lambda_n^- f_n \right)$$

$$= \gamma \left(\sum_{n=1}^{N} \lambda_n^+ f_n \right) - \gamma \left(\sum_{n=1}^{N} \lambda_n^- f_n \right)$$

$$\leq \gamma \left(\sum_{n=1}^{N} \lambda_n f_n \right)$$

and thus that $\phi_0 \leq \gamma$ on $\mathrm{Lin}(C)$. By Hahn Banach we may thus find an extension ϕ of ϕ_0 to the whole of $\mathscr{L}$ such that $\phi \leq \pi$. Given that γ is $\geq_*$-monotone and positive homogeneous we conclude that ϕ is positive and, by Lemma 8, that it admits the decomposition (60). Write $\beta = \phi^{\perp}$ and $m = m_\phi$. If $N \in \mathcal{N}_*$, then $\mathbf{1}_N =_* 0$ so that $0 = \phi(\mathbf{1}_N) = \beta(\mathbf{1}_N) + m(N) = m(N)$ i.e. $m \in ba_{*,+}$. Likewise, if $g \in \mathscr{L}$ then $0 \leq \beta(|g|\mathbf{1}_N) = \phi(|g|\mathbf{1}_N) \leq \gamma(|g|\mathbf{1}_N) \leq 0$ so that β vanishes on $\mathfrak{B}_*$, as claimed. The converse is obvious. $\qquad \square$

RESULTS FROM SECTION 2

Proof of Lemma 1. In any lattice X the operation $x \to x^-$ is subadditive, that is $(x + y)^- \leq x^- + y^-$. Thus, if $\mathscr{X} = \{X(\alpha) : \alpha \in \mathfrak{A}\}$ and $f, g \in \mathfrak{F}_0(\mathscr{X})$

$$q(f + g) = \sum_{X \in \mathscr{X}} \left\{ [(f + g)(X)^+]a(X) - [(f + g)(X)^-]b(X) \right\}$$

$$= \sum_{X \in \mathscr{X}} \left\{ (f + g)(X)a(X) + (f + g)(X)^-(a(X) - b(X)) \right\}$$

$$\leq \sum_{X \in \mathscr{X}} \left\{ (f + g)(X)a(X) + (f(X)^- + g(X)^-)(a(X) - b(X)) \right\}$$

$$= q(f) + q(g)$$

Positive homogeneity is clear. Suppose now that $f, g \in \mathfrak{F}_0(X)$ satisfy $fg \geq 0$ that is $f(X)$ and $g(X)$ have the same sign for all $X \in \mathscr{X}$. It is then obvious that $(f + g)(X)^- = f(X)^- + g(X)^-$ from which the claim follows. $\qquad \square$

Proof of Lemma 2. $f \geq f$ implies $f \geq_* f$. $f \geq_* g$ and $g \geq_* h$ are equivalent to $f - g \geq_* 0$ and $g - h \geq_* 0$ so that $f - h = (f - g) + (g - h) \geq_* 0$ by Assumption 4. Let $f \geq_* 0$. Then $f - \alpha \geq_* 0$ for all $\alpha \leq 0$ so that, from (10), $f_* \geq 0$. On the other hand, $f_* \geq 0$ means that $f + 2^{-n} \geq_* 0$ for $n = 1, 2, \ldots$ so that $f \geq_* 0$ by Assumption 4. By definition $f >_* 0$ if and only if $f_* \geq 0$ and $-f \not\geq_* 0$ or, equivalently, $(-f)_* \not\geq 0$ i.e. $f^* > 0$. Let f and b be as in (iii). The

claim is clearly true if b is simple, by Assumption 4.(iii) and (iv). Denote by b^n and b_n two simple functions such that $b^n \geq b \geq b_n \geq 0$ and $\sup_\omega |b - b^n|(\omega) + |b - b_n|(\omega) < 2^{-n}$. Fix $\eta > 0$ and let $b(n) = b^n \mathbf{1}_{\{f < -\eta\}} + b_n \mathbf{1}_{\{f \geq -\eta\}}$. Then,

$$fb = (f + \eta)b - b\eta \geq (f + \eta)b_n \mathbf{1}_{\{f \geq -\eta\}} + (f + \eta)b^n \mathbf{1}_{\{f < -\eta\}} - b\eta \geq fb(n) - 2^{-n}\eta$$

Thus, by Assumption 4.(ii), $fb \geq_* fb(n)$ while $fb(n) \geq_* 0$ follows from Assumption 4.(iii)–(iv) and $b(n)$ being simple. Let $f =_* 0$, i.e. $f \geq_* 0$ and $-f \geq_* 0$. Then $-f^+, -f^- \geq_* 0$, by Assumption 4.(iv) so that $-|f| \geq_* 0$; the converse is obvious. Eventually, assume that $f \geq_* 0$. Then $-\eta \mathbf{1}_{\{f < -\eta\}} \geq f \mathbf{1}_{\{f < -\eta\}} \geq_* 0$ so that $-\mathbf{1}_{\{f < -\eta\}} \geq_* 0$. The last property is obvious. $\qquad\square$

RESULTS FROM SECTION 3.

Proof of Lemma 3. The only non trivial property is subadditivity. Let $a \leq f_*$ and $b \leq g_*$. Then $f + g - (a + b - 2^{-n}) \geq_* 0$ i.e. $a + b \leq (f + g)_*$. We conclude that $(f + g)_* \geq f_* + g_*$. But then,

$$\left(\bar{X}(\theta_1)_* \wedge 0\right) + \left(\bar{X}(\theta_2)_* \wedge 0\right) \leq \left(\bar{X}(\theta_1)_* + \bar{X}(\theta_2)_*\right) \wedge 0 \leq \bar{X}(\theta_1 + \theta_2)_* \wedge 0$$

whenever $\theta_1, \theta_2, \theta_1 + \theta_2 \in \Theta$. We conclude that ρ is indeed subadditive. Let $\phi : \Theta \to \overline{\mathbb{R}}_+$ satisfy (iv) and (v). If $\rho(\theta) = \infty$ the inequality $\rho(\theta) \geq \phi(\theta)$ is obvious; if $\theta \in \Theta_*$, then $\bar{X}(\theta + \rho(\theta)\delta_0) \geq_* 0$ so that $0 = \phi(\theta + \rho(\theta)\delta_0) \geq \phi(\theta) - \rho(\theta)$. Observe that, by (10), $f - (f_* - 2^{-n}) \geq_* 0$ so that $f \geq_* f_*$ and $\{f - f_* < -\eta\} \in \mathcal{N}_*$ and thus $\mu(f - f_* < -\eta) = 0$ when $\mu \in ba_*$. Thus, $\mu \in \mathbb{P}_*$ implies $\bar{X}(\theta) \wedge 0 \in L(\mu)$ and

$$\sup_{\eta > 0}(\bar{X}(\theta)_* \wedge \eta) - \eta \leq \sup_{\eta > 0} \int_{\{\bar{X}(\theta) \geq \bar{X}(\theta)_* - \eta\}} (\bar{X}(\theta) \wedge 0)d\mu \leq \int (\bar{X}(\theta) \wedge 0)d\mu$$

and

$$\int (-\bar{X}(\theta) \wedge 0)d\bar{\mu} \leq (-\bar{X}(\theta)_* + \eta \wedge 0)$$

from which we deduce (16). $\qquad\square$

Proof of Theorem 1. If $\theta \in \Theta_*$, let $\theta' = \theta + \rho(\theta)\delta_0 \in \Theta$. By Lemma 2 and Assumption 3.(iii),

$$(63) \qquad \bar{X}(\theta') = \bar{X}(\theta) + \rho(\theta) \geq_* 0 \qquad t(\theta') = t(\theta) + q_0\rho(\theta) \qquad \bar{X}(\theta')^* = \bar{X}(\theta)^* + \rho(\theta)$$

Let t be coherent and $\theta \in \Theta$. If $\rho(\theta) = \infty$ then (17) is trivially true; otherwise, $0 \leq t(\theta') = t(\theta) + \rho(\theta)q_0$. If, conversely, $\bar{X}(\theta) \geq_* 0$, i.e. $\rho(\theta) = 0$, then (17) implies $t(\theta) \geq 0$ so that t is coherent.

We may restrict attention to $\theta \in \Theta_*$. Assume that $\bar{X}(\theta)^* + \rho(\theta) > 0$. Then, by (63), θ' is an arbitrage opportunity unless $t(\theta) + q_0\rho(\theta) > 0$. (18) is thus necessary for absence of arbitrage. Conversely, assume that t meets (18). Choose $\theta \in \Theta$ and fix $\varepsilon \geq 0$ arbitrarily but such that $\bar{X}(\theta)^* + \rho(\theta) + \varepsilon > 0$. By assumption, $\theta_\varepsilon = \theta + [\varepsilon + \rho(\theta)]\delta_0 \in \Theta$ and $\bar{X}(\theta_\varepsilon)^* + \rho(\theta_\varepsilon) > 0$. If (18) holds then

$$0 < t(\theta_\varepsilon) + q_0\rho(\theta_\varepsilon) = t(\theta) + [\varepsilon + \rho(\theta)]q_0$$

so that $t(\theta) + \rho(\theta)q_0 > -\varepsilon q_0$ for all $\varepsilon > 0$. Thus $t(\theta) + \rho(\theta)q_0 \geq 0$ for all θ (so that t is coherent) and, if $\bar{X}(\theta) >_* 0$, then $t(\theta) = t(\theta) + \rho(\theta)q_0 > 0$. We conclude that t admits no arbitrage opportunity. $\square$

Proof of Theorem 2. Under the current assumptions t is a convex functional. Consider the sets

$$\mathcal{H}_0 = \left\{ \left(\bar{X}(\theta) \wedge 0 \right) q_0 - t(\theta) : \theta \in \Theta \right\}$$

and $\mathcal{H} = \{ f \in \mathfrak{F}(\Omega) : \lambda h \geq_* f$ for some $h \in \mathcal{H}_0$, and $\lambda \geq 0 \}$. If $\lambda_1, \ldots, \lambda_N > 0$ and $\theta^1, \ldots, \theta^N \in \Theta$ then $\sum_{n=1}^N \lambda_n \left[(\bar{X}(\theta^n) \wedge 0)q_0 - t(\theta^n) \right] \leq \lambda \left[(\bar{X}(\theta) \wedge 0)q_0 - t(\theta) \right]$ with $\lambda = \sum_{n=1}^N \lambda_n$ and $\theta = \sum_{n=1}^N (\lambda_n/\lambda)\theta^n \in \Theta$. Moreover, $h_n \geq_* f_n$ for $n = 1, \ldots, N$ implies $\sum_{n=1}^N h_n \geq_* \sum_{n=1}^N f_n$. Thus $\mathcal{H}$ is a convex cone which, by (17), contains no element f with $f_* > 1$. We thus conclude from Theorem 9 that there exists a probability $\mu \in ba_*$ such that

$$\mathcal{H} \subset L^1(\mu) \quad \text{and} \quad \sup_{\{ f \in \mathcal{H} : \inf f > -\infty \}} \int f d\mu \leq 0$$

Moreover, if $\theta \in \Theta_*$ and $f = (\bar{X}(\theta) \wedge 0) - t(\theta)$ then

$$0 \geq \int_{\{ \bar{X}(\theta) > \bar{X}(\theta)_* - \eta \}} f d\mu = \int_{\{ \bar{X}(\theta) > \bar{X}(\theta)_* - \eta \}} (\bar{X}(\theta) \wedge 0) d\mu - t(\theta) = \int (\bar{X}(\theta) \wedge 0) d\mu - t(\theta)$$

which proves the direct implication. The converse follows from Theorem 1 and (16). $\square$

RESULTS FROM SECTION 4.

Lemma 9. *The functional $\pi : \mathfrak{F}(\Omega) \to \overline{\mathbb{R}}$ is $\geq_*$-monotone, positively homogeneous and satisfies*

$$(64) \qquad\qquad \pi(\bar{X}(\theta)) \leq q(\theta) \qquad \theta \in \Theta \qquad and \qquad \pi(f + g) \leq \pi(f) + \pi(g)$$

for all $f, g \in \mathfrak{F}(\Omega)$ for which the sum $\pi(f) + \pi(g)$ is defined. Moreover, the following properties are equivalent: (i) q is coherent, (ii) $\pi(0) = 0$, (iii) $\pi^c(1) \leq q_0$ and (iv)

$$(65) \quad |\pi(b)| < \infty \quad and \quad \pi(f) + \pi(b^*) \geq \pi(f + b) \geq \pi(f) + \pi^c(b_*) \qquad for\ all\ f \in \mathfrak{F}(\Omega),\ b \in \mathfrak{B}_*$$

Proof. Monotonicity, positive homogeneity and the first part of (64) are obvious properties of π. Assume that $f, g \in \mathfrak{F}(\Omega)$ are such that $\pi(f) + \pi(g)$ is a well defined element of $\overline{\mathbb{R}}$. Thus if, say, $\pi(f) = \infty$ then $\pi(f) + \pi(g) = \infty$ and the second part of (64) is obvious. If, alternatively, $\pi(f), \pi(g) < \infty$, then there exist $\lambda_f, \lambda_g \geq 0$ and $\theta_f, \theta_g \in \Theta$ such that $\lambda_f \bar{X}(\theta_f) \geq_* f$ and $\lambda_g \bar{X}(\theta_g) \geq_* g$ so that $\lambda(\bar{X}(\theta'_f) + \bar{X}(\theta'_g)) \geq_* f + g$, with $\lambda = \lambda_f + \lambda_g$ and $\theta'_f = \theta_f \lambda_f / \lambda$ and $\theta'_g = \theta_g \lambda_g / \lambda$ (with the convention $0/0 = 0$). Given that, by Assumption 1, $\theta = \theta'_f + \theta'_g \in \Theta$ we conclude that

$$\pi(f + g) \leq \lambda q(\theta) \leq \lambda(q(\theta'_f) + q(\theta'_g)) = \lambda_f q(\theta_f) + \lambda_g q(\theta_g)$$

and, the inequality being true for all λ_f, λ_f and θ_f, θ_g as above, the second half of (64) follows. (64) also implies $\pi(0) \leq 0$. It is then clear that (ii) is equivalent to (i). If $\theta \in \Theta$ and $\lambda \geq 0$ are such that $\lambda \bar{X}(\theta) \geq_* -1$ then $(1 + \lambda)\bar{X} \left(\frac{\lambda \theta + \delta_0}{1 + \lambda} \right) \geq_* 0$ so that

$$\pi(0) \leq (1 + \lambda)q \left(\frac{\lambda \theta + \delta_0}{1 + \lambda} \right) \leq \lambda q(\theta) + q_0$$

We thus conclude that $q_0 \geq \pi(0) + \pi^c(1)$ and so that $\pi(0) = 0$ implies $\pi^c(1) \leq q_0$. If $b \in \mathfrak{B}_*$, then (64) implies $|b|^* \pi(-1) \leq \pi(b) \leq |b|^* \pi(1)$ so that from (iii) we deduce $|\pi(b)| < \infty$. But then the sums $\pi(f) + \pi(b)$ and $\pi(f + b) + \pi(-b)$ are well defined for each $f \in \mathfrak{F}(\Omega)$ and the second half of (65) follows from (64). Conversely, by (65) we conclude that $\pi(0) = n\pi(0)$ for each $n \in \mathbb{N}$ and $\pi(0) \in \mathbb{R}$ so that $\pi(0) = 0$. $\qquad\qquad\square$

Proof of Lemma 4. If $\pi^c(1) = 0$ and $\bar{X}(\theta) \geq_* 0$ then $q(\theta) \geq \pi(-1)$. If $\bar{X}(\theta)_* < 0$ then $\bar{X}(\theta)/|\bar{X}(\theta)_*| \geq_* -1$ and so $q(\theta)/|\bar{X}(\theta)_*| \geq 0$. Conversely, $\lambda\bar{X}(\theta) \geq_* -1$ and $\lambda > 0$ imply $\theta \in \Theta_*$ and thus $q(\theta) \geq 0$ so that $\pi(-1) \geq 0$. $\qquad\qquad\square$

Denote by

(66) $$\Phi(\pi) = \{\phi \in \mathfrak{F}(\mathscr{K}) : \phi \text{ positive, linear and such that } \phi \leq \pi\}$$

Adopting the notation of Lemma 8 we can also write

(67) $$\mathscr{M}(\pi) = \{m_\phi : \phi \in \Phi(\pi)\} \qquad \text{and} \qquad \Phi^\perp(\pi) = \left\{\phi^\perp : \phi \in \Phi(\pi)\right\}$$

Proof of Theorem 3. By Lemma 9, if q is coherent then the space $\mathscr{K}$ is a vector sublattice of $\mathfrak{F}(\Omega)$ containing $\mathfrak{B}_*$ and π a $\geq_*$-monotone, positively homogeneous and subadditive functional on $\mathscr{K}$. Fix $h \in \mathfrak{B}_*$ and consider the set $C_h = \{\lambda h : 0 \leq \lambda \leq 1\}$. By Theorem 10 there is a positive linear functional β_h on $\mathscr{K}$ vanishing on $\mathfrak{B}_*$ and $m_h \in ba_{*,+}$ such that $\mathscr{K} \subset L^1(m_h)$ and $\pi(f) \geq \beta_h(f) + \int f dm_h$ for all $f \in \mathscr{K}$ and such that $\pi(h) = \beta_h(h) + \int h dm_h = \int h dm_h$. Suppose that $g \in \mathscr{K}_*$. Then, $g^- \in \mathfrak{B}_*$ and thus

$$\pi(g) \geq \beta_h(g) + \int g dm_h = \beta_h(g^+) + \int g dm_h \geq \int g dm_h$$

so that $m_h \in \mathscr{M}$. Conversely, if $m \in \mathscr{M}$ and $\bar{X}(\theta)_* \geq 0$ then (22) implies

$$q(\theta) \geq \pi(\bar{X}(\theta)) \geq \int \bar{X}(\theta) dm = \int \bar{X}(\theta) \mathbf{1}_{\{\bar{X}(\theta) > \bar{X}(\theta)_* - \varepsilon\}} dm \geq [\bar{X}(\theta)_* - \varepsilon] m(\Omega) \geq -\varepsilon m(\Omega)$$

for every $\varepsilon > 0$ so that q is coherent. $\qquad\qquad\square$

Proof of Theorem 4. Choose $0 \neq m_0 \in \mathscr{M}$ to be countably additive on $\mathscr{A}$ and write $P = m_0/\|m_0\|$. If the sequences $\langle f_n \rangle_{n \in \mathbb{N}}$ and $\langle x_n \rangle_{n \in \mathbb{N}}$ satisfy the conditions in (ii), then,

$$\int f dP = \lim_n \int f_n dP \leq \|m_0\|^{-1} \lim_n \pi(f_n) \leq -\pi(1)/\|m_0\| \lim_n x_n \leq -x$$

so that (ii) holds. Fix P as in (ii) and, assuming that (iii) fails, pick a sequence $\langle f_n \rangle_{n \in \mathbb{N}}$ in $\mathscr{S}(\mathscr{A})_+$ which converge to 0 in $L^1(P)$ but such that $\inf_n \pi^c(f_n) > c$ for some $c > 0$. Write $h_n = -f_n \pi(1)/\pi^c(f_n)$. Thus h_n converges to 0 in $L^1(P)$ but $x_n = -\pi(h_n)/\pi(1) = 1$ in contrast with (ii). Let now P be as in (iii) and suppose that for each $m \in \mathscr{M}$ we may construct a sequence $\langle F_n(m) \rangle_{n \in \mathbb{N}}$ in $\mathscr{A}$ such that $\lim_n P(F_n(m)) = 0 < x(m) \equiv \inf_n m(F_n(m))$. For each such m,

choose n sufficiently large and define $h_n(m) = \mathbf{1}_{F_n(m)}x(m)^{-1}$ so that $\int h_n(m)dP < 2^{-n}$ while $\int h_n(m)dm \geq 1$. The convex set $\mathscr{H}_n = \left\{ h \in \mathscr{S}(\mathscr{A})_+ : \int h dP < 2^{-n} \right\}$ is thus such that

$$\inf_{m \in \mathscr{M}} \sup_{h \in \mathscr{H}_n} \int h dm \geq 1$$

Observe that $\mathscr{M}$ is convex and weak* compact and that $\mathscr{H}_n$ is convex. By the minimax Theorem of Sion [64, Corollary 3.3], there exists then $h_n \in \mathscr{H}_n$ such that

$$\pi^c(h_n) = \inf_{m \in \mathscr{M}} \int h_n dm \geq 1/2$$

The sequence $\langle h_n \rangle_{n \in \mathbb{N}}$ so obtained contradicts (iii).   □

Lemma 10. *If q is coherent then the set $\mathscr{M}$ defined in (22) is non empty, convex and weak* compact subset of ba_+. Moreover, $\mathscr{M} = \mathscr{M}(\pi)$ (see (67)).*

Proof. If q is coherent, $\mathscr{M}$ is non empty by Theorem 3. By (22), $\mathscr{M}(\pi) \subset \mathscr{M}$. Thus, we only need to prove that $\mathscr{M}$ is closed in the weak* topology of ba and that $\mathscr{M} \subset \mathscr{M}(\pi)$. Let m_0 be an element of the closure of $\mathscr{M}$ and $f \in \mathscr{K}$. Then $m_0 \in ba_{*,+}$ and

$$\int (|f| \wedge n)dm_0 \leq \sup_{m \in \mathscr{M}} \int (|f| \wedge n)dm \leq \sup_{m \in \mathscr{M}} \int |f| dm \leq \pi(|f|)$$

so that the sequence $\langle |f| \wedge n \rangle_{n \in \mathbb{N}}$ is Cauchy in $L^1(m_0)$. Moreover, for all $c > 0$[16]

$$v(m_0)(|f| > c + |f| \wedge n) \leq v(m_0)(|f| > c + n) \leq \frac{1}{c+n} \int [|f| \wedge (c+n)]dm_0 \leq \frac{\pi(|f|)}{c+n}$$

which proves that $|f| \wedge n$ converges to $|f|$ in $L^1(m_0)$ and so that $f \in L^1(m_0)$ [28, III.3.6] . Moreover, if $f \in \mathscr{K}_*$

$$\int f dm_0 = \lim_n \int (f \wedge n)dm_0 \leq \sup_{m \in \mathscr{M}} \int f dm \leq \pi(f)$$

which proves that $m_0 \in \mathscr{M}$. Observe that by Tychonoff Theorem [28, I.8.5], the set $\Phi(\pi)$ is compact in the topology induced on it by $\mathscr{K}$. Let $\langle m_\gamma \rangle_{\gamma \in \Gamma}$ be a net in $\mathscr{M}(\pi)$ converging to $m \in \mathscr{M}$ in the weak * topology of ba. For each $\gamma \in \Gamma$ there exists $\phi_\gamma \in \Phi(\pi)$ such that $m_\gamma = m_{\phi_\gamma}$. By moving to a subnet if necessary we obtain that the net $\langle \phi_\gamma \rangle_{\gamma \in \Gamma}$ converges in the topology induced by $\mathscr{K}$ to some limit $\phi \in \Phi(\pi)$. Denote by m_ϕ the part of ϕ representable as an integral, as in (60). If $b \in \mathfrak{B}$ we have, by the inclusion $\mathfrak{B} \subset \mathscr{K}$ that follows from q being coherent,

$$\int b dm_\phi = \phi(b) = \lim_\gamma \phi_\gamma(b) = \lim_\gamma \int b dm_\gamma = \int b dm$$

so that $m = m_\phi \in \mathscr{M}(\pi)$.   □

---

[16]By $v(m)$ we total variation of m as defined in [28, III.1.9].

Results from section 5

Proof of Theorem 5. Indeed if q is coherent then $\mathscr{M}$ is a non empty, convex and weak* compact subset of ba, by Lemma 10; moreover, if $m \in \mathscr{M}$ and $f \in \mathscr{K}$ then, by (22), $\left| \int f dm \right| \leq \int |f| dm \leq \pi(|f|) < \infty$ so that (25) may be regarded as an implicit definition of β. Observe that from (22) and Lemma 10 we obtain

$$
\begin{aligned}
\inf_{\phi \in \Phi(\pi)} \phi^{\perp}(f) &\leq \pi(f) - \sup_{m \in \mathscr{M}} \int f dm \\
&= \beta(f) \\
&\leq \sup_{\phi \in \Phi(\pi)} \phi^{\perp}(f) + \sup_{\mu \in \mathscr{M}} \int f d\mu - \sup_{m \in \mathscr{M}} \int f dm \\
&= \sup_{\phi \in \Phi(\pi)} \phi^{\perp}(f)
\end{aligned}
$$

(68)

Given that $\sup_{\phi \in \Phi(\pi)} \phi^{\perp}(f) = 0$ for all $f \in \mathfrak{B}_*$, as we showed in the proof of Theorem 3, we conclude that β vanishes on $\mathfrak{B}_*$. This proves existence. To show uniqueness, suppose that $\bar{\beta}$ and $\bar{\mathscr{M}}$ is another pair with the same properties of β and $\mathscr{M}$ and for which the decomposition (25) holds. If $\mu \in \bar{\mathscr{M}} \backslash \mathscr{M}$, then there exists $f \in \mathfrak{B}$ such that $\sup_{\bar{m} \in \bar{\mathscr{M}}} \int f d\bar{m} \geq \int f d\mu > \sup_{m \in \mathscr{M}} \int f dm$ but $\bar{\beta}(f) = \beta(f) = 0$, a contradiction of (25). To show that (25) is sufficient for q to be coherent, let $f \in \mathscr{K}_*$. Then, $f^- \in \mathfrak{B}_*$ and thus $\beta(f) = \beta(f^+) \geq 0$ and thus

$$
\pi(f) \geq \sup_{m \in \mathscr{M}} \int f dm \geq f_* \sup_{m \in \mathscr{M}} \|m\| = f_* \pi(1)
$$

Therefore, if $\bar{X}(\theta) \geq_* 0$ for some $\theta \in \Theta$ then $q(\theta) \geq \pi(\bar{X}(\theta)) \geq 0$ and q is coherent. $\square$

Corollary 2. *The functional β defined in* (25) *is positive and satisfies*

(69)
$$
-\lim_{n} \{\pi(f) - \pi(f \vee -n)\} \leq \beta(f) \leq \lim_{n} \{\pi(f) - \pi(f \wedge n)\} \qquad f \in \mathscr{K}
$$

and

(70)
$$
\beta(f) = \lim_{n} \{\pi(f) - \pi(f \wedge n)\} \qquad f \in \mathscr{K}_*
$$

Proof. Positivity of β follows from (68) and the fact that $\phi^{\perp}(f) \geq 0$ for each $f \in \mathscr{K}_*$ and $\phi \in \Phi(\pi)$. Moreover,

$$
\sup_{m \in \mathscr{M}} \int f dm = \sup_{m \in \mathscr{M}} \lim_{n} \int (f \vee -n) m \leq \lim_{n} \sup_{m \in \mathscr{M}} \int (f \vee -n) m \leq \lim_{n} \pi(f \vee -n)
$$

Likewise, given that $\phi^{\perp}(f \wedge n) \leq 0$ for all $f \in \mathscr{K}$ and $\phi \in \Phi(\pi)$ by Lemma 10

$$
\sup_{m \in \mathscr{M}} \int f dm = \lim_{n} \sup_{m \in \mathscr{M}} \int (f \wedge n) dm = \lim_{n} \sup_{\phi \in \Phi} \{\phi(f \wedge n) - \phi^{\perp}(f \wedge n)\} \geq \lim_{n} \pi(f \wedge n)
$$

and (69) is proved. (70) follows from

$$
\beta(f) = \pi(f) - \sup_{m \in \mathscr{M}} \int f dm = \pi(f) - \sup_{m \in \mathscr{M}} \lim_{n} \int (f \wedge n) dm = \pi(f) - \lim_{n} \sup_{m \in \mathscr{M}} \int (f \wedge n) dm
$$

and the fact that $\sup_{m \in \mathscr{M}} \int (f \wedge n) dm = \pi(f \wedge n)$ whenever $f \in \mathscr{K}_*$. $\square$

Results from section 6

Theorem 11. *If Θ_X is priced efficiently then there are $\beta_X \in \mathbb{R}_+$ and $\nu_X \in ca\,(\mathscr{B}(\mathbb{R}_+))_+$ such that*

$$(71) \qquad\qquad q_X(k) = \beta_X + \int_k^\infty \nu_X(x > t)dt \qquad k \in K(X)$$

Proof. Assume that Θ_X is priced efficiently as of (32). In restriction to $\mathscr{K}_X$ the functional π_X is then $\geq_*$ monotonic, positively homogeneous and subadditive. Moreover, $\mathfrak{B}_* \subset \mathscr{K}_X$. By (32) and (6), π_X is additive on the convex set $C_X = \{\frac{X(\theta)}{X \wedge 1} : \theta \in \Theta_X\}$ which contains the origin. Then Theorem 10 applies. Let $\phi_X : \mathscr{K}_X \to \mathbb{R}$ be the $\geq_*$ positive linear functional such that $\phi_X \leq \pi_X$ and that $\phi_X = \pi_X$ in restriction to C_X. Observe that $(X - k)^+/X \wedge 1 \in \mathscr{K}_X$ for all $k \geq 0$ and so it is possible to define the function $F : \mathbb{R}_+ \to \mathbb{R}_+$ implicitly by letting

$$F(k) = \phi_X\left(\frac{(X - k)^+}{X \wedge 1}\right) \qquad k \geq 0$$

Of course, $F(k) = q_X(k)$ for all $k \in K(X)$; in addition, it is decreasing and convex. By a standard result on convex functions, we may write

$$(72) \qquad\qquad F(k_2) = F(k_1) + \int_{k_1}^{k_2} f(t)dt \qquad 0 < k_1 < k_2$$

where, for definiteness, we take $f(t)$ to be the right derivative of F for $t \in \mathbb{R}_+$. Suppose that $\{u \geq X > t\} \in \mathcal{N}_*$ for some $0 \leq t < u$ and fix $0 < h \leq (u - t)/2$. There is then a negligible set outside of which each of the options with strike prices $t, t + h, u - h, u$ expires in the money if and only if all the others do. In other words

$$\frac{(X - t)^+}{X \wedge 1} + \frac{(X - u)^+}{X \wedge 1} =_* \frac{(X - (t + h))^+}{X \wedge 1} + \frac{(X - (u - h))^+}{X \wedge 1}$$

from which it follows

$$\begin{aligned} F(t) + F(u) &= \phi_X\left(\frac{(X - t)^+}{X \wedge 1} + \frac{(X - u)^+}{X \wedge 1}\right) \\ &= \phi_X\left(\frac{(X - (t + h))^+}{X \wedge 1} + \frac{(X - (u - h))^+}{X \wedge 1}\right) \\ &= F(t + h) + F(u - h) \end{aligned}$$

Thus,

$$\frac{F(u) - F(u - h)}{h} = \frac{F(t + h) - F(t)}{h}$$

i.e. the left derivative of F at u and the right derivative of F at t coincide. There exists then a set $D \subset \mathbb{R}_+$ with $\mathbb{R}_+ \backslash D$ at most countable and such that $\{X > u\} \triangle N_1 = \{X > t\} \triangle N_2$ for $t, u \in D$ and $N_1, N_2 \in \mathcal{N}_*$ imply $f(t) = f(u)$. It is therefore possible to define a positive set function $\hat{\lambda}_0$ on the collection $\{\{X > t\} \triangle N : t \in D,\ N \in \mathcal{N}_*\}$ implicitly by letting

$$(73) \qquad\qquad \hat{\lambda}_0(\{X > t\} \triangle N) = -f(t) \qquad t \in D,\ N \in \mathcal{N}_*$$

Since f is negative and increasing (by convexity of F) the identities $\lambda_0(\Omega) = -f(0)$, $\lambda_0(\varnothing) = 0$ and

(74)
$$\lambda_0(\{X > t\} \triangle N) = \sup_{\{u \in D : u \geq t\}} \hat{\lambda}_0(X > u) \qquad t \in \mathbb{R}_+, \ N \in \mathcal{N}_*$$

define a positive extension λ_0 of $\hat{\lambda}_0$ to the family $\mathscr{A}_0(X)$ of subsets of Ω formed by Ω, $\varnothing$ and all sets of the form $\{X > t\} \triangle N$ with $t \in \mathbb{R}_+$ and $N \in \mathcal{N}_*$. To see that this indeed defines an extension, observe that, if there is $t_\infty \in D$ such that $\{X > t_\infty\} = \varnothing$ then $F(t_\infty + h) = F(t_\infty)$ and so $f(t_\infty) = 0$. Likewise, if $\{X > t_0\} = \Omega$ for some $t_0 \in D$, then $\{X > t_0\} = \{X > 0\}$ so that, as seen above, $f(t_0)$ coincides with $f(0)$. If either t_0 or t_∞ do not exist, then one can choose the corresponding value of λ_0 arbitrarily. Since the elements of $\mathscr{A}_0(X)$ are linearly ordered by inclusion it follows that if $A_i = \{X > t_i\} \triangle N_i$ and $N_i \in \mathcal{N}_*$ for $i = 1, 2$ with $t_1 \geq t_2$ then

$$\lambda_0(A_1) + \lambda_0(A_2) = \lambda_0(X > t_1) + \lambda_0(X > t_2)$$
$$= \lambda_0(\{X > t_1\} \cap \{X > t_2\}) + \lambda_0(\{X > t_1\} \cup \{X > t_2\})$$
$$= \lambda_0(A_1 \cap A_2) + \lambda_0(A_1 \cup A_2)$$

as $\{X > t_1\} \triangle (A_1 \cap A_2), \{X > t_2\} \triangle (A_1 \cup A_2) \subset N_1 \cup N_2 \in \mathcal{N}_*$. It follows from [6, Theorems 3.1.6 and 3.2.10], that there exists a unique extension $\lambda \in ba(\mathscr{A}(X))_+$ of λ_0 to the algebra $\mathscr{A}(X)$ generated by $\mathscr{A}_0(X)$ and thus such that $\lambda(N) = 0$ when $N \in \mathcal{N}_*$. Let

(75)
$$\beta_X = \lim_{k \to \infty} F(k)$$

Then we obtain from (72)

$$\beta_X = F(k) - \int_k^\infty \lambda(X > t)dt \qquad k \geq 0$$

To eventually get (71), write

$$\mathscr{A} = \{A \subset \mathbb{R}_+ : X^{-1}(A) \in \mathscr{A}(X)\}$$

It is clear that $\mathscr{A}$ is an algebra containing the algebra $\mathscr{A}(\mathbb{R}_+)$ generated by the left open intervals of $\mathbb{R}_+$. Define then $\lambda_X \in ba(\mathscr{A}(\mathbb{R}_+))$ by letting $\hat{\lambda}_X(A) = \lambda(X \in A)$ and observe from (72) that $\int_{\mathbb{R}_+} \lambda_X(x > t)dt = -\int_{\mathbb{R}_+} f(t)dt \leq F(0)$ so that $\lim_t \lambda_X((t, \infty)) = 0$. Exploiting standard rules of the Lebesgue integral and integration by parts we obtain

$$\int_{\mathbb{R}_+} \lambda(X > t)dt = \int_{\mathbb{R}_+} \lambda_X(x > t)dt = \int_{\mathbb{R}_+} x d\lambda_X(x) = \int x d\lambda_X(x)$$

as $\lambda_X(x < -t) = 0$ for all $t \geq 0$. It follows from [27, Lemma 2, p. 191] that, uniquely associated with λ_X is its *conventional companion* $\nu_X \in ca(\mathscr{A}(\mathbb{R}_+))_+$ with the property that

(76)
$$\int h(x)d\lambda_X = \int h(x)d\nu_X$$

for any continuous function $h : \mathbb{R} \to \mathbb{R}$ for which either integral is well defined. The extension from $\mathscr{A}(\mathbb{R}_+)$ to the generated σ algebra $\mathscr{B}(\mathbb{R}_+)$ is standard. Thus the representation (71) is implicit in Θ_X being priced efficiently. Suppose that $\hat{\beta}_X \geq 0$ and $\hat{\nu}_X \in ca(\mathscr{B}(\mathbb{R}_+))_+$ is another pair for which

the representation (71) holds. Then, $\beta_X - \hat{\beta}_X = \int_k^\infty [\hat{\nu}_X(x > t) - \nu_X(x > t)]dt$ for all $k \geq 0$ which implies $\beta_X = \hat{\beta}_X$. $\square$

Proof of Lemma 5. The direct implication is obvious. For the converse, assume that $q_X(k)$ possesses the mentioned properties and that $K(X) = J(X)$. Represent $\theta \in \Theta_X$ as $\theta = \sum_{j \in J(X)} \theta_j \theta_X(j)$. From Lemma 6 in the Appendix we know that there exists $\theta^* = \sum_{j \in J(X)} \theta_j^* \theta_X(j) \in \Theta_X$ such that

$$\frac{X(\theta^*)}{X \wedge 1} \geq_* \frac{X(\theta)}{X \wedge 1} \quad \text{and} \quad q(\theta^*) = \pi_X \left(\frac{X(\theta)}{X \wedge 1} \right)$$

Moreover, from (47) and (79) we conclude that for $n = 0, \ldots, I - 1$

$$\sum_{i=0}^n \theta_i^* = \frac{\sum_{i \leq n} \theta_i(j_{n+1} - j_i) - \sum_{i \leq n-1} \theta_i(j_{n+1} - j_i)}{j_{n+1} - j_n} = \sum_{i=0}^n \theta_i$$

and $\theta_I^* = \theta_I$ so that $\theta = \theta^*$ and thus $q(\theta) = \pi_X(X(\theta)/(X \wedge 1))$. Θ_X is thus priced efficiently. $\square$

Proof of Theorem 7. The existence of $\theta_X(g_t)$ satisfying (37) is proved in Theorem **??** in the Appendix. Consider the function $t \to q_X^G(t)$. It is clear that $q_X^G(t) \geq 0$ and that, when a, t, u, z are as in (35) then

$$q_X^G(z) = \pi_X \left(\frac{g_z(X)}{X \wedge 1} \right)$$
$$\leq \pi_X \left(\frac{a g_t(X) + (1-a)g_u(X)}{X \wedge 1} \right)$$
$$\leq a\pi_X \left(\frac{g_t(X)}{X \wedge 1} \right) + (1-a)\pi_X \left(\frac{g_u(X)}{X \wedge 1} \right)$$
$$= a q_X^G(t) + (1-a)q_X^G(u)$$

Each $q_X^G(t)$ may equivalently be viewed as the fictitious price of a CALL option written on X with strike price t. Such artificial options market would then be priced efficiently, as the corresponding CALL function clearly satisfies the conditions of Lemma 5, with $K^G(X) = \mathbb{R}_+$. But then, the representation (38) follows directly from Theorem 7. $\square$

RESULTS FROM SECTION 8

Lemma 11. *The functional π_X defined in (30) and acting on $\mathscr{K}_X$ is $\geq_*$-monotone, positively homogenous and subadditive. Moreover, $\mathfrak{B}_* \subset \mathscr{K}_X$ and $\pi_X(1) > 0$.*

Proof. To show $\mathfrak{B}_* \subset \mathscr{K}_X$ it is enough to show that $0, 1 \in \mathscr{K}_X$ which is however obvious if options have non negative prices and in view of the inequality $\pi_X(1) \leq q_X(0)$. Then $0 \leq \pi_X(|b|) \leq b^*\pi(1)$. The remaining properties are obvious. $\square$

Lemma 12. *Let Assumption 5 hold, choose $f, g \in \Gamma$ and let either (i) $\hat{f} = f(X^*)$ if $X^* < \infty$ or else (ii) $\hat{f} = \lim_n f(n)/n$ ($\hat{g}$ being defined likewise). If $f = X(\theta)$ with $\theta \in \Theta_X$, then*

(77) $$\frac{f(X)}{X \wedge 1} \geq_* \frac{g(X)}{X \wedge 1} \quad \text{if and only if} \quad f(k_n) \geq g(k_n) \quad n = 1, \ldots, N \text{ and } \hat{f} \geq \hat{g}$$

where $K(X) = \{k_n : n = 0, \ldots, N\}$ and $k_0 = 0$.

Proof. If $f(k_n) < g(k_n) - \varepsilon$ for some $\varepsilon > 0$ and $n = 1, \ldots, I$ then by continuity there exists $k_n - 1 > \eta > 0$ such that $f < g - \varepsilon$ in restriction to the set $A_n = \{k_n - \eta < X \leq k_n\}$. Likewise, if $\hat{f} < \hat{g} - \varepsilon$ then $f < g - \varepsilon$ in restriction to the set $A = \{X > X^* - \eta\}$, if $X^* < \infty$, or $A = \{X > t\}$ for some $t \in \mathbb{N}$ if $X^* = \infty$. By Assumption 5, however, neither A_n nor A belong to $\mathcal{N}_*$; moreover, $X \wedge 1 = 1$ on these sets so that $f(X)/(X \wedge 1) \geq_* g(X)/(X \wedge 1)$ is contradicted.

Conversely, if $f(k_n) \geq g(k_n)$ holds for $n = 1, \ldots, I$, then, given that $f(0) = g(0) = 0$, that g is convex and f piecewise linear, we conclude that $f(x) \geq g(x)$ for all $0 \leq x \leq X^*$ and so that $f(X) \geq g(X)$ and thus $f(X)/(X \wedge 1) \geq_* g(X)/(X \wedge 1)$. $\square$

Proof of Lemma 6. It is, first of all, clear that in solving (45) one may restrict attention to portfolios formed with options with strike prices in $J(X)$. This implies that $j_0 = 0$ and that each $\lambda X(\theta)$ with $\lambda > 0$ and $\theta \in \Theta_X$ in (45) may be taken to be of the form $\sum_{i=0}^{I} a_i (X - j_i)^+$ so that $\lambda q(\theta) = \mathbf{q}^T \mathbf{a}$ with $\mathbf{a}^T = [a_0, a_1, \ldots, a_I] \in \mathbb{R}_+^{I+1}$. Write

$$g_{\mathbf{a}}(X) = \sum_{i=0}^{I} a_i (X - j_i)^+ \quad \text{and} \quad \hat{g}_{\mathbf{a}} = \begin{cases} g_{\mathbf{a}}(X^*) & \text{if } X^* < \infty \\ \lim_n g_{\mathbf{a}}(n)/n & \text{otherwise} \end{cases}$$

and observe that $g_{\mathbf{a}}$ is a convex function, being a linear combination of convex functions with positive weights. Remark also that $[g_{\mathbf{a}}(j_1), \ldots, g_{\mathbf{a}}(j_I), \hat{g}_{\mathbf{a}}]^T = \mathbf{D}\mathbf{a}$. Fix $f \in \Gamma$. By Lemma 12 $\frac{g_{\mathbf{a}}(X)}{X \wedge 1} \geq_* \frac{f(X)}{X \wedge 1}$ is equivalent to $\mathbf{D}\mathbf{a} \geq \mathbf{f}$.

Define the vectors $\mathbf{w}, \mathbf{b} \in \mathbb{R}^{I+1}$ implicitly by letting

$$(78) \qquad b_I d_I = q_X(j_I) \quad \text{and} \quad b_I e_I + \sum_{i=n}^{I-1} b_i = \frac{q_X(j_n) - q_X(j_{n+1})}{j_{n+1} - j_n} \qquad n = 0, \ldots, I - 1$$

with $e_I = 1$ if $X^* < \infty$ or else $e_I = 0$ and

$$(79) \qquad \sum_{i=0}^{n} w_i = \frac{f(j_{n+1}) - f(j_n)}{j_{n+1} - j_n} \qquad n = 0, \ldots, I - 1 \quad \text{and} \quad \sum_{i=0}^{I} w_i d_i = \hat{f}$$

The following properties are easily established by induction: (i) $\mathbf{b} \geq 0$ (as $j_0, \ldots, j_I \in J(X)$) (ii) $\mathbf{w} \geq 0$ (as $f \in \Gamma$), (iii) $\mathbf{b}^T \mathbf{D} = \mathbf{q}^T$ and (iv) $\mathbf{w} = \mathbf{D}^{-1}\mathbf{f}$. But then,

$$\min_{\{\theta \in \Theta_X, \lambda > 0 : \lambda X(\theta) \geq_* f(X)\}} q(\theta) = \min_{\{\mathbf{a} \in \mathbb{R}_+^{I+1} : \mathbf{D}\mathbf{a} \geq \mathbf{f}\}} \mathbf{q}^T \mathbf{a} = \min_{\{\mathbf{a} \in \mathbb{R}_+^{I+1} : \mathbf{D}\mathbf{a} \geq \mathbf{f}\}} \mathbf{b}^T \mathbf{D}\mathbf{a} \geq \mathbf{b}^T \mathbf{f} = \mathbf{q}^T \mathbf{w}$$

$\square$

RESULTS FROM SECTION 8

Proof of Lemma 7. By a result of Turlach [66, p. 85] the program (51) admits as its solution a cubic $\mathscr{C}^2$ spline of the form

$$g_k^h(x) = \sum_{i=1}^{5} \mathbf{1}_{[k+t_i^h, k+t_{i+1}^h)}(x) P_{k,i}^h \left(x - (k + t_i^h) \right) \qquad x \in \mathbb{R}_+$$

where, as in the text, $t_i^h = (i-3)h/2$ for $i = 1, \ldots, 5$, $t_6^h = \infty$ and $P_{k,i}^h$ is a polynomial of degree 3. It is clear from the constraints imposed to (51) that indeed $g_k^h \in \Gamma$. Moreover, these same constraints

imply that $g_k^h(x) = g_k^0(x)$ when $x \notin [k-h, k+h]$ while $1 \geq Dg_k^h(x) \geq 0$ on $[k-h, k+h]$. Thus, $g_k^h - g_k^0$ is increasing on $(-\infty, k]$ and decreasing afterwards, so that $\sup_x(g_k^h - g_k^0)(x) \leq (g_k^h - g_k^0)(k) = g_k^h(k)$. Moreover, the (non empty) set of solutions is clearly convex and the functional $I_k(h; f)$ is strictly convex in f so that the solution is necessarily unique. Define

$$\bar{g}_k^h(x) = g_k^h(2k - x) + (x - k) \qquad x \geq 0$$

One deduces from (50) that

$$I_k(h; \bar{g}_k^h) = \sum_{i=1}^{5} [g_k^0(k + t_i^h) - \bar{g}_k^h(k + t_i^h)]^2 + \int_{k-h}^{k+h} D^2(\bar{g}_k^h(x))^2 dx$$

$$= \sum_{i=1}^{2} [\bar{g}_k^h(k + t_i^h)]^2 + \bar{g}_k^h(k)^2 + \sum_{i=4}^{5} [\bar{g}_k^h(k + t_i^h) - t_i^h]^2 + \int_{k-h}^{k+h} D^2(\bar{g}_k^h(x))^2 dx$$

$$= \sum_{i=4}^{5} [g_k^h(k + t_i^h) - t_i^h]^2 + \bar{g}_k^h(k)^2 + \sum_{i=1}^{2} [g_k^h(k + t_i^h)]^2 + \int_{k-h}^{k+h} D^2(g_k^h(x))^2 dx$$

$$= I_k(h; \bar{g}_k^h)$$

However, since the solution is unique, we have the symmetry relation

(80) $$g_k^h(x) = g_k^h(2k - x) + (x - k) \qquad x \geq 0$$

Let $f \in \chi(k; h)$ and let $h > h'$ and

$$Tf(x) = f\left(k + (x - k)\frac{h}{h'}\right)\frac{h'}{h}$$

Observe that $T : \chi(k; h) \to \chi(k; h')$ is one to one and onto and that $D^2 Tf(x) = h/h' D^2 f(k + (x - k)h/h')$. Thus,

$$I_k(h'; Tf) = \sum_{i=1}^{5} [g_k^0(k + t_i^{h'}) - Tf(k + t_i^{h'})]^2 + (h')^3 \int_{k-h'}^{k+h'} (D^2 Tf(x))^2 dx$$

$$= (h'/h)^2 \sum_{i=1}^{5} [g_k^0(k + t_i^h) - f(k + t_i^h)]^2 + (h')^2 h \int_{k-h}^{k+h} (D^2 f(z))^2 dz$$

$$= (h'/h)^2 I_k(h; f)$$

Thus f solves the program (51) relatively to h if and only if Tf solves it relatively to h'. By uniqueness we conclude that (53) holds. Observe then, if $x \leq k$

$$g_k^{h'}(x) = g_k^h(k + (x - k)h/h')h'/h \leq g_k^h(x)h'/h \leq g_k^h(x)$$

a conclusion which extends to $x > k$ by (80). This proves (ii). Given that $0 \leq Dg_k^h(x) \leq 1$ we conclude that

$$0 \leq g_k^{h'}(x) - g_k^0(x) \leq g_k^{h'}(x) - g_k^0(x) = g_k^{h'}(k) = \frac{h'}{h}g_k^h(k)$$

so that g_k^h decreases to g_k^0 uniformly in x.

If $y = k' - k \in \mathbb{R}$, then

$$g_k^h(x) = \sum_{i=1}^{5} \mathbf{1}_{[k'+t_i^h, k'+t_{i+1}^h)}(x+y) P_{k,i}^h(x+y-(k'+t_i^h)) \equiv \gamma_k^h(x+y; k')$$

It is obvious that $\gamma_k^h(x+y; k') = \gamma_k^h(x; k)$ and that $g_{k+y}^0(x+y) = g_k^0(x)$. But then, for $i = 1, \ldots, 5$,

$$g_{k'}^0(k'+t_i^h) - \gamma_k^h(k'+t_i^h; k') = g_k^0(k+t_i^h) - \gamma_k^h(k+t_i^h; k) = g_k^0(k+t_i^h) - g_k^h(k+t_i^h)$$

and that

$$\int_{k-h}^{k+h} D^2 \gamma_k^h(x; k)^2 dx = \int_{k'-h}^{k'+h} D^2 \gamma_k^h(x+y; k')^2 dx$$

Using the notation of (50), we conclude that

$$I_k(h) = I_k(h; \gamma_k^h(\cdot; k)) = I_{k+y}(h; \gamma_k^h(\cdot; k')) \geq I_{k'}(h) \qquad k > 0, y \in \mathbb{R}$$

The same inequality holds after exchanging k for k' so that $\gamma_k^h(\cdot; k')$ and $g_{k'}^h$ are both $\mathscr{C}^2$ splines of degree 3 solving (51) and thus coincide, by uniqueness. We conclude that $g_{k+y}^h(x) = g_k^h(x-y)$ and thus, if $0 \leq a \leq 1$ and $k_3 \leq ak_1 + (1-a)k_2$,

$$ag_{k_1}^h(x) + (1-a)g_{k_2}^h(x) = ag_{k_3}^h(x+(k_1-k_3)) + (1-a)g_{k_3}^h(x+k_2-k_3)$$
$$\geq g_{k_3}^h(x+ak_1+(1-a)k_2-k_3))$$
$$\geq g_{k_3}^h(x)$$

proving (i). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proof of Theorem 8. Exploiting the inclusion $F \in \mathscr{C}^2$ we can rewrite (56) as

$$(81) \quad q_X^h(a) = F(a) + \sum_{k_i \in J(X)} w_a^h(i)[DF(a)(k_i-a) + \frac{1}{2}D^2F(a)(k_i-a)^2 + o((k_i-a)^2)] + \sum_{k_i \in J(X)} w_a^h(i)\varepsilon_i$$

From (48) we know that w_a^h is a constant conditional on each selection z; moreover, given that $g_a^h(x) = g_a^0(x)$ whenever $x \notin B_r(a)$, the open ball of radius r around x, we conclude that $w_a^h(i) = 0$ unless $k_{i-1} < a + h$ and $k_{i+1} > a - h$ i.e. unless $k_i \in B_{(\delta+1)M}(a)$. The first sum on the right hand side of (81) is then majorized by

$$|DF(a)|(\delta+1)M + \frac{1}{2}D^2F(a)(\delta+1)^2M^2 + o((\delta+1)^2M^2)$$

On the other hand by $(57).(i)$

$$P\left(\sum_{k_i \in J(X)} w_a^h(i)\varepsilon_i\right) = P\left(\sum_{k_i \in J(X)} w_a^h(i)P_z(\varepsilon_i)\right) = P\left(P_z(\varepsilon_1) \sum_{k_i \in J(X)} w_a^h(i)\right) = P(\varepsilon_1) = 0$$

Likewise, from (48) and the fact that g_a^h is symmetric around a we conclude that

$$(82) \qquad w_a^h(1) = o(M) \qquad \text{and} \qquad w_a^h(i) \leq \frac{1}{2}D^2g_a^h(a)M + o(M) \qquad \text{for} \quad i > 1$$

so that, by (57).(ii)

$$P\left(\sum_{k_i \in J(X)} w_a^h(i)^2 \varepsilon_i^2\right) \leq P\left(\sup_i P_z(\varepsilon_i^2) \sum_{k_i \in J(X)} w_a^h(i)^2\right)$$

$$\leq P\left(\sup_i P_z(\varepsilon_i^2) \sum_{k_i \in J(X)} w_a^h(i)\right) o(M)$$

$$\leq o(M) P\left(\sup_i P_z(\varepsilon_i^2)\right)$$

Eventually, if $i_0, j_0 \in z$,

$$P\left(\sum_{\{i \neq j : k_i, k_j \in J(X)\}} w_a^h(i) w_a^h(j) \varepsilon_i \varepsilon_j\right) = P\left(P_z(\varepsilon_{i_0} \varepsilon_{j_0}) \sum_{\{i \neq j : k_i, k_j \in J(X)\}} w_a^h(i) w_a^h(j)\right)$$

$$= P(\varepsilon_{i_0} \varepsilon_{j_0}) - P\left(P_z(\varepsilon_{i_0} \varepsilon_{j_0}) \sum_{k_i \in J(X)} w_a^h(i)^2\right)$$

$$= -P\left(P_z(\varepsilon_{i_0} \varepsilon_{j_0}) \sum_{k_i \in J(X)} w_a^h(i)^2\right)$$

so that, by (82), the last term is clearly less than

$$P\left(|P_z(\varepsilon_{i_0} \varepsilon_0 j)|\right) O(M)$$

This together with continuity of $D^2 g_a^h$ proves (58).   □

<div align="center">REFERENCES</div>

[1] Y. Aït-Sahalia, J. Duarte (2003), *Nonparametric Option Pricing Under Shape Restrictions*, J. Econometrics **116**, 9-47.

[2] Y. Aït-Sahalia, A. Lo (1998), *Nonparametric Estimation of State-Price-Densities Implicit in Financial Asset Prices*, J. Finance **53**, 499-547.

[3] Y. Amihud, H. Mendelson (1986), *Asset Pricing and the bid/ask Spread*, J. Finance **17**, 223-249.

[4] R. W. Banz, M. H. Miller (1978), *Prices for State-Contingent Claims: Some Estimates and Applications*, J. Business **51**, 653-672.

[5] B. Bensaid, J.-P. Lesne, H. Pagés, J. Scheinkman (1992), *Derivative Asset Pricing with Transaction Costs*, Math. Finance **2**, 63-86.

[6] K. P. S. Bhaskara Rao, M. Bhaskara Rao (1983), *Theory of Charges*, Academic Press, London.

[7] A. Bisin (1998), *General Equilibrium with Endogenously Incomplete Financial Markets*, J. Econ. Theory **82**, 19-45.

[8] V. I. Bogachev (2007), *Measure Theory*, Springer-Verlag, Berlin-Heidelberg.

[9] N. P. B. Bollen, T. Smith, R. E. Whaley (2004), *Modeling the bid/ask Spread: Measuring the Inventory-Holding Premium*, J. Financ. Econ. **72**, 97-141.

[10] P. Bordalo, N. Gennaioli, A. Shleifer (2012), *Salience Theory of Choice under Risk*, Quarterly J. Econ. **127**, 1243-1285.

[11] P. Bordalo, N. Gennaioli, A. Shleifer (2013), *Salience and Asset Prices*, Amer. Econ. Review Papers and Proceedings **103**, 623-628.

[12] B. Bouchard (2006), *No-Arbitrage in Discrete-time Markets with Proportional Transaction Costs and General Information Structures*, Finance Stoch. **10**, 276-297.

[13] D. Breeden, R. Litzenberger (1978), *Prices of State-Contingent Claims Implicit in Option Prices*, J. Business **51**, 621-651.

[14] P. Carr, D. B. Madan (2005), *A Note on Sufficient Conditions for No Arbitrage*, Fin. Research Lett. **2**, 125-130.

[15] P. Carr, H. Géman, D. B. Madan, M. Yor (2003), *Stochastic Volatility for Lévy Processes*, Math. Finance **13**, 345-382.

[16] G. Cassese (2008), *Asset Pricing with no Exogenous Probability Measure*, Math. Finance **18**, 23-54.

[17] G. Cassese (2009), *Sure wins, Separating Probabilities and the Representation of Linear Functionals*, J. Math. Anal. Appl. **354**, 558-563.

[18] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci (2012), *Put-Call Parity and Market Frictions*, IGIER W. P. No. 447.

[19] A. Chateauneuf, R. Kast, A. Lapied (1996), *Choquet Pricing for Financial Markets with Frictions*, Math. Finance **6**, 323-330.

[20] A. M. G. Cox, D. Hobson (2005), *Local Martingales, Bubbles and Option Prices*, Fin. Stoch. **9**, 477-492.

[21] M. Das, W. K. Newey, F. Vella (2003), *Estimation of Sample Selection Models*, Rev. Econ. Stud. **70**, 33-58.

[22] B. de Finetti (1937), *La Prévision: Ses Lois Logiques, ces Sources Subjectives*, Ann. I.H.P. **7**, 1-68.

[23] F. Delbaen, W. Schachermayer (1994), *A General Version of the Fundamental Theorem of Asset Pricing*, Math. Ann. **300**, 463-520.

[24] J. B. De Long, A. Shleifer, L. H. Summers, R. J. Waldmann (1990), *Noise Trader Risk in Financial Markets*, J. Political Econ. **98**, 703-738.

[25] H. Demsetz (1968), *The Cost of Transacting*, Quarterly J. Econ. **82**, 35-53.

[26] J. C. Dermody, E. Z. Prisman (1993), *No Arbitrage and Valuation in Markets with Realistic Transaction Costs*, J. Fin. Quant. Anal. **28**, 65-80.

[27] L. E. Dubins, L. J. Savage (1965), *How to Gamble if You Must*, McGraw-Hill, New York.

[28] N. Dunford, J. T. Schwarz (1988), *Linear Operators*, Wiley, New York.

[29] R. L. Dykstra (1983), *An algorithm for restricted least squares regression*, J. Amer. Stat. Ass. **78**, 837-842.

[30] N. El-Karoui, C. Ravanelli (2009), *Cash Subadditive Risk Measures and Interest Rate Ambiguity*, Math. Finance **19**, 561-590.

[31] R. L. Eubank (1999), *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York - Basel.

[32] M. R. Fengler (2009), *Arbitrage-Free Smoothing of the Implied Volatility Surface*, Quant. Finance **9**, 417-428.

[33] D. Filipovic (2008), *Optimal Numeraires for Risk Measures*, Math. Finance **18**, 333-336.

[34] P. Gagliardini, C. Gourieroux, E. Renault (2011) *Efficient Derivative Pricing by the Extended Method of Moments*, Econometrica **79**, 1181-1232.

[35] R. Garcia, E. Ghysels, E. Renault (2010), *The Econometrics of Options Pricing*, in Y. Aït-Sahalia, L. P. Hansen (Eds.) *Handbook of Fianncial Econometrics*, vol. 1, 479-552, Amsterdam North-Holland.

[36] N. Gennaioli, A. Shleifer (2010), *What Comes to Mind*, Quarterly J. Econ. **75**, 1399-1433.

[37] C. Gilles, S. F. LeRoy (1992), *Bubbles and Charges*, Int. Econ. Rev. **33**, 323-339.

[38] F. Hahn (1971), *Equilibrium with Transaction Costs*, Econometrica **39**, 417-439.

[39] L. P. Hansen, R. Jagannathan (1991), *Implications of Security Market Data for Models of Dyinamic Economies*, J. Pol. Econ. **99**, 225-262.

[40] M. J. Harrison, D. M. Kreps (1979) *Martingales and Arbitrage in Multiperiod Securities Markets*, J. Econ. Theory **20**, 381-408.

[41] J. Hasbrouck (2002), *Stalking the "Efficient Price" in Market Microstructure Specifications: an Overview*, J. Financial Markets **5**, 329-339.

[42] J. J. Heckman (1979), *Sample Selection Bias as a Specification Error*, Econometrica **47**, 153-161.

[43] S. L. Heston, M. Lowenstein, G. A. Willard (2007), *Options and Bubbles*, Rev. Financial Studies **20**, 359-390.

[44] R. D. Huang, H. R. Stoll (1997), *The Components of the bid/ask Spread: a General Approach*, Rev. Financial Studies **10**, 995-1034.

[45] Hugonnier J. (2012), *Rational Asset Pricing Bubbles and Portfolio Constraints*, J. Econ. Theory **146**, 2260-2302.

[46] J. C. Jackwerth, M. E. Rubinstein (1996), *Recovering Probability Distributions from Option Prices*, J. Finance **51**, 1611-1631.

[47] E. Jouini, H. Kallal (1995), *Martingales and Arbitrage in Securities Markets with Transaction Costs*, J. Econ. Theory **66**, 178-197.

[48] E. Jouini, C. Napp (2001), *Arbitrage and Investment Opportunities*, Finance Stochast. **5**, 305-325.

[49] Yu. Kabanov, C. Stricker (2001), *The Harrison-Pliska Pricing Theorem under Transaction Costs*, J. Math. Econ. **35**, 185-196.

[50] D. Kahneman, A. Tversky (1979), *Prospect Theory: An Analysis of Decision Under Risk*, Econometrica **47**, 263-292.

[51] D. M. Kreps (1981), *Arbitrage and Equilibrium in Economies with Infinitely Many Commodities*, J. Math. Economics **8**, 15-35.

[52] O. Lamont. R. Thaler (2003), *Can Markets Add and Subtract? Mispricing in Tech Stock Carve-outs*, J. Political Econ. **111**, 227-268.

[53] O. Lamont. R. Thaler (2003), *The Law of One Price in Financial Markets*, J. Econ. Perspectives **17**, 191-202.

[54] M. Lowenstein, G. A. Willard (2000), *Rational Equilibrium Asset Pricing Bubbles in Continuous Trading Models*, J. Econ. Theory **91**, 17-58.

[55] M. Lowenstein, G. A. Willard (2006), *The Limits of Investor Behaviour*, J. Finance **61**, 231-258.

[56] E. G. J. Luttmer (1996), *Asset Pricing in Economies with Frictions*, Econometrica **64**, 1439-1467.

[57] E. Mammen, C. Thomas-Agnan (1999), *Smoothing Splines and Shape Restrictions*, Scand. J. Statist. **26**, 239-252.

[58] C. Napp (2003), *The Dalang-Morton-Willinger Theorem under Cone Constraints*, J. Math. Econ. **39**, 111-126.

[59] E. Z. Prisman (1986), *Valuation of Risky Assets in Arbitrage Free Economies with Frictions*, J. Finance **41**, 545-557.

[60] F. Riedel (2011), *Finance without Probabilistic Prior Assumptions*, U. Bielefeld, Inst. Math. Econ. W.P. 450.

[61] A. Roux (2011), *The Fundamental Theorem of Asset Pricing in the Presence of bid/ask and Interest Rate Spreads*, J. Math. Econ. **47**, 159-163.

[62] W. Schachermayer (2004), *The Fundamental Theorem of Asset Pricing under Proportional Transaction Costs in Finite Discrete Time*, Math. Finance **14**, 19-48.

[63] A. Shleifer, R. W. Vishny (1997), *The Limits of Arbitrage*, J. Finance **52**, 35-55.

[64] M. Sion (1958), *On General Minimax Theorems*, Pacific J. Math. **8**, 171-176.

[65] H. R. Stoll (2003), *Market Microstructure*, in G. M. Constantinides, M. Harris, R. M. Stulz (eds.), *Handbook of the Economics of Finance*, Vol. 1a, 553-604, North-Holland, Amsterdam.

[66] B. A. Turlach (2005), *Shape Constrained Smoothing Using Smoothing Splines*, Comp. Stat. **20**, 81-103.

[67] A. Yatchew (1998), *Nonparametric Regression Techniques in Economics*, J. Econ. Lit. **36**, 669-721.

[68] A. Yatchew, W. Härdle (2006), *Nonparametric State Price Density Estimator Using Constrained Least Squares and Bootstrap*, J. Econometrics **133**, 579-599.

[69] H. Yin, Y. Wang, L. Qi (2009), *Shape-Preserving Interpolation and Smoothing for Options Market Implied Volatility*, J. Optim. Theory Appl. **142**, 243-266.

Università Milano Bicocca and University of Lugano

E-mail address: gianluca.cassese@unimib.it

Current address: Department of Economics, Statistics and Management, Building U7, Room 2097, via Bicocca degli Arcimboldi 8, 20126 Milano - Italy
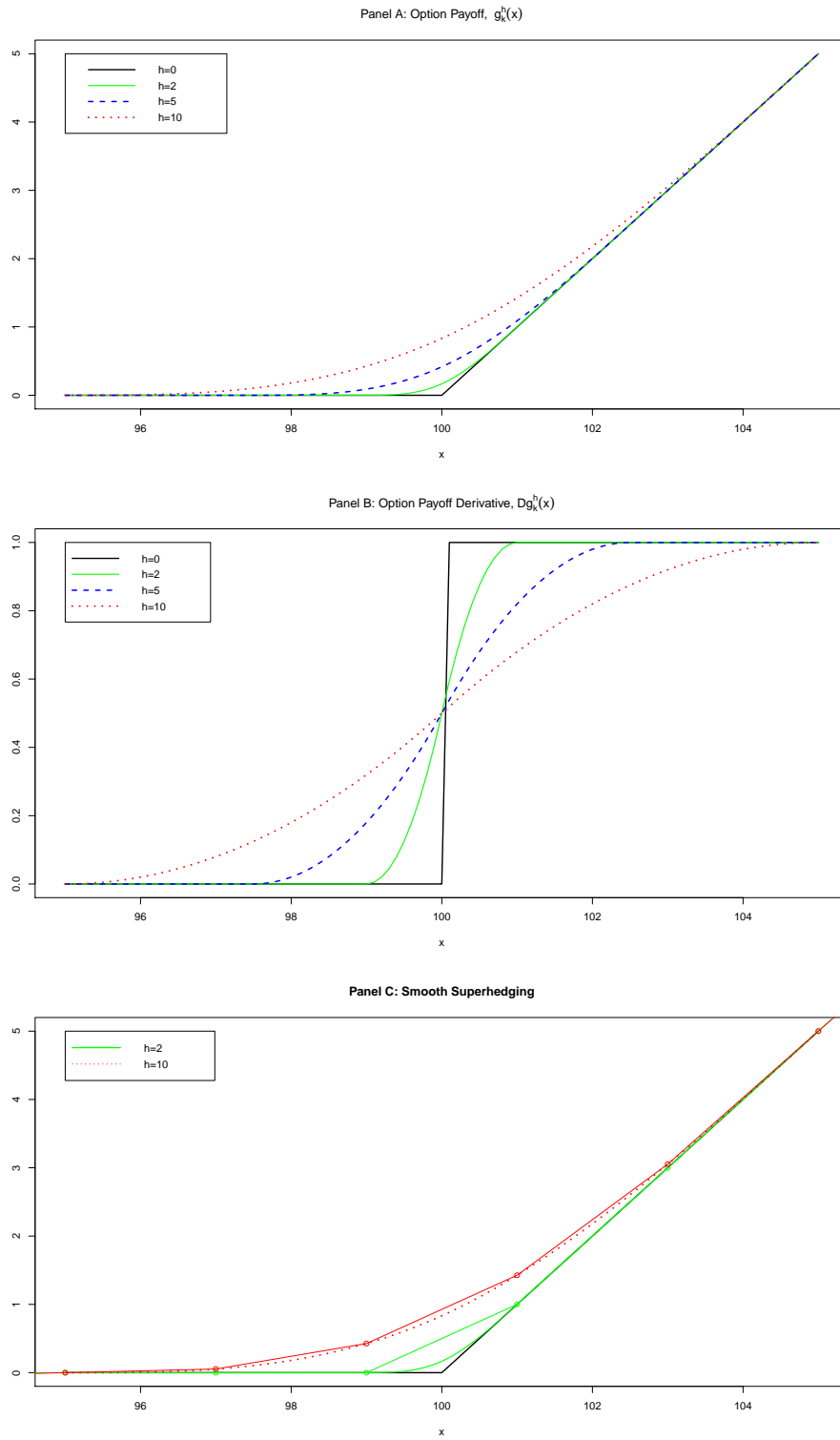
Panel A: Option Payoff, $g_k^h(x)$

Panel B: Option Payoff Derivative, $Dg_k^h(x)$

Panel C: Smooth Superhedging

FIGURE 1.   Smoothing Option Payoff by h.

In Panel A and B we plot the payoff $g_k^h(x)$ and its derivative $Dg_k^h(x)$ for $k = 100$ and different values of h. We have set $\lambda = 0.25$. In Panel C we have drawn for the cases $h = 2, 10$ the smooth payoff $g_k^h(x)$ together with the payoff of the portfolio superhedging it, assuming the existence of options with strikes 97, 99, 101, 103 and 105.

Panel A: Option Price, $q_X^{\delta M}(k)$

Panel B: Implied volatility

Panel C: Implicit Probability $\nu^{\delta M}(X > k)$
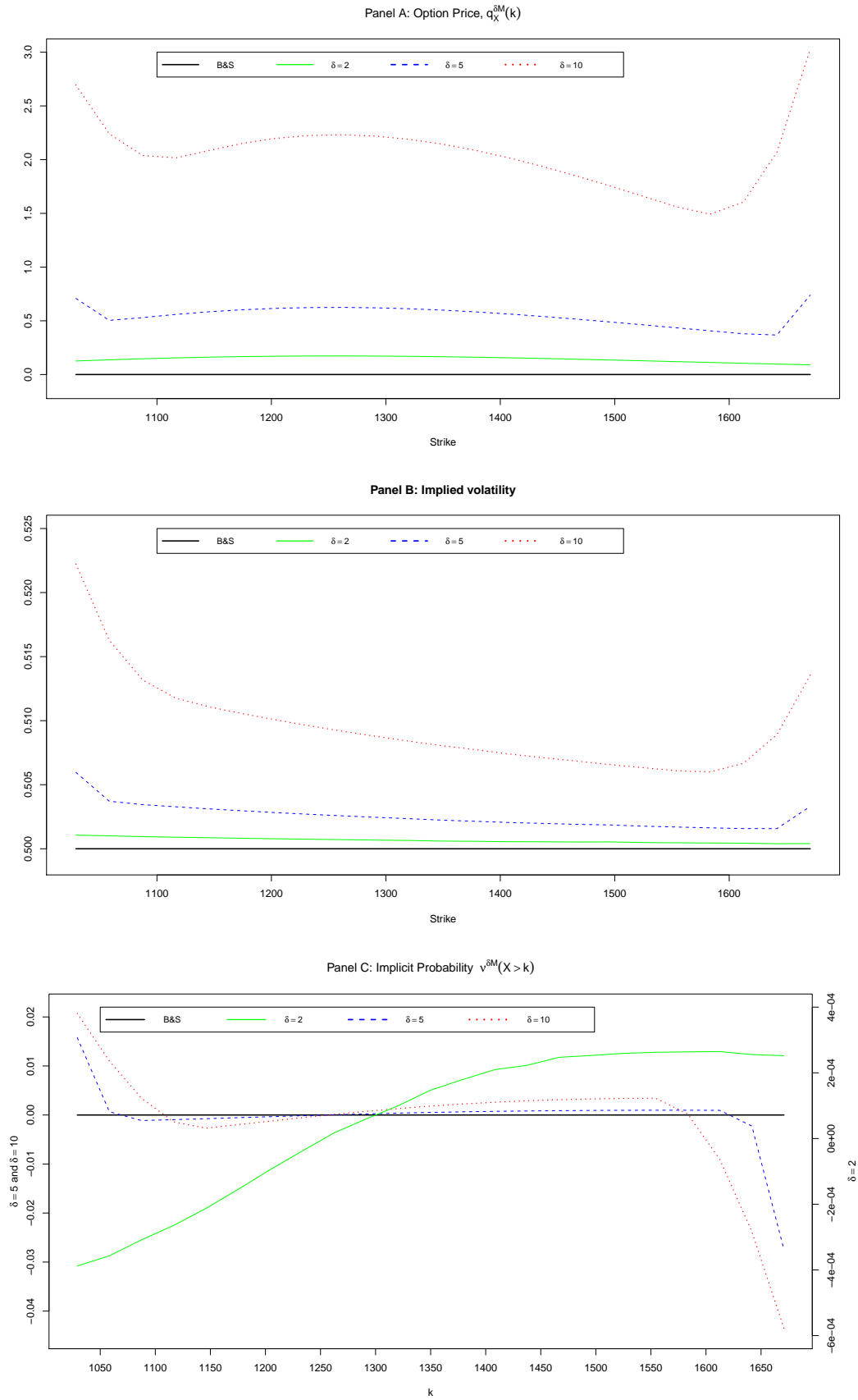
FIGURE 2. Deviations from Black and Scholes.
In Panel C the curve relative to $\delta = 2$ is represented on the right hand side scale.

Panel A: Option Price, $q_X^{\delta M}(k)$

Panel B: Implied volatility

Panel C: Implicit Probability $\nu^{\delta M}(X > k)$
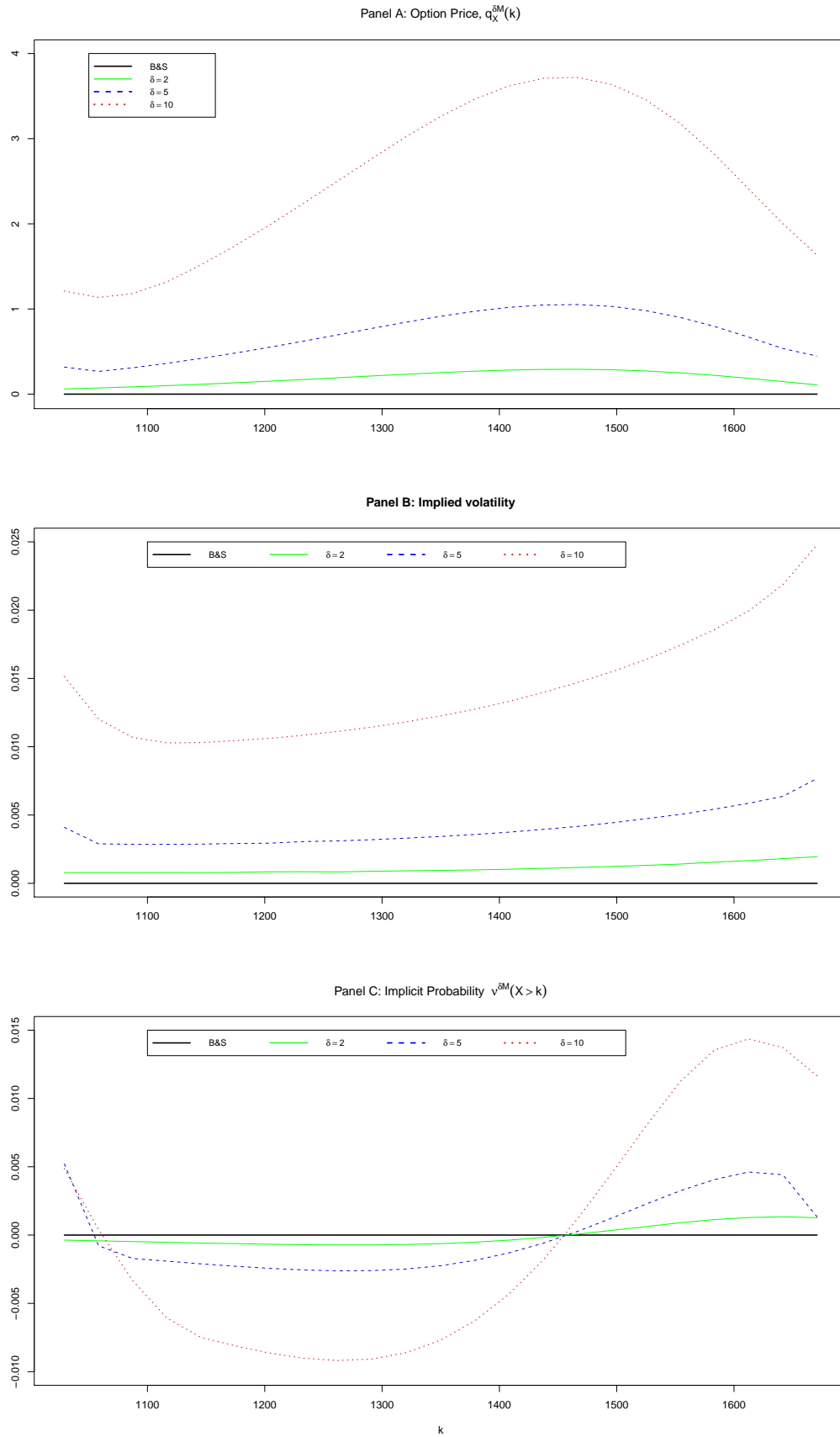
FIGURE 3.   Deviations from Black and Scholes in the Presence of a Smile Effect.
$\sigma = 0.4 - 0.2(k - 1000)/700$ and $\nu^{B\&S} = -\partial q^{B\&S}/\partial K - \partial q^{B\&S}/\partial\sigma \; d\sigma/dk$.

Panel A: Price

Panel B: Volatility

Panel C: Implicit Probability
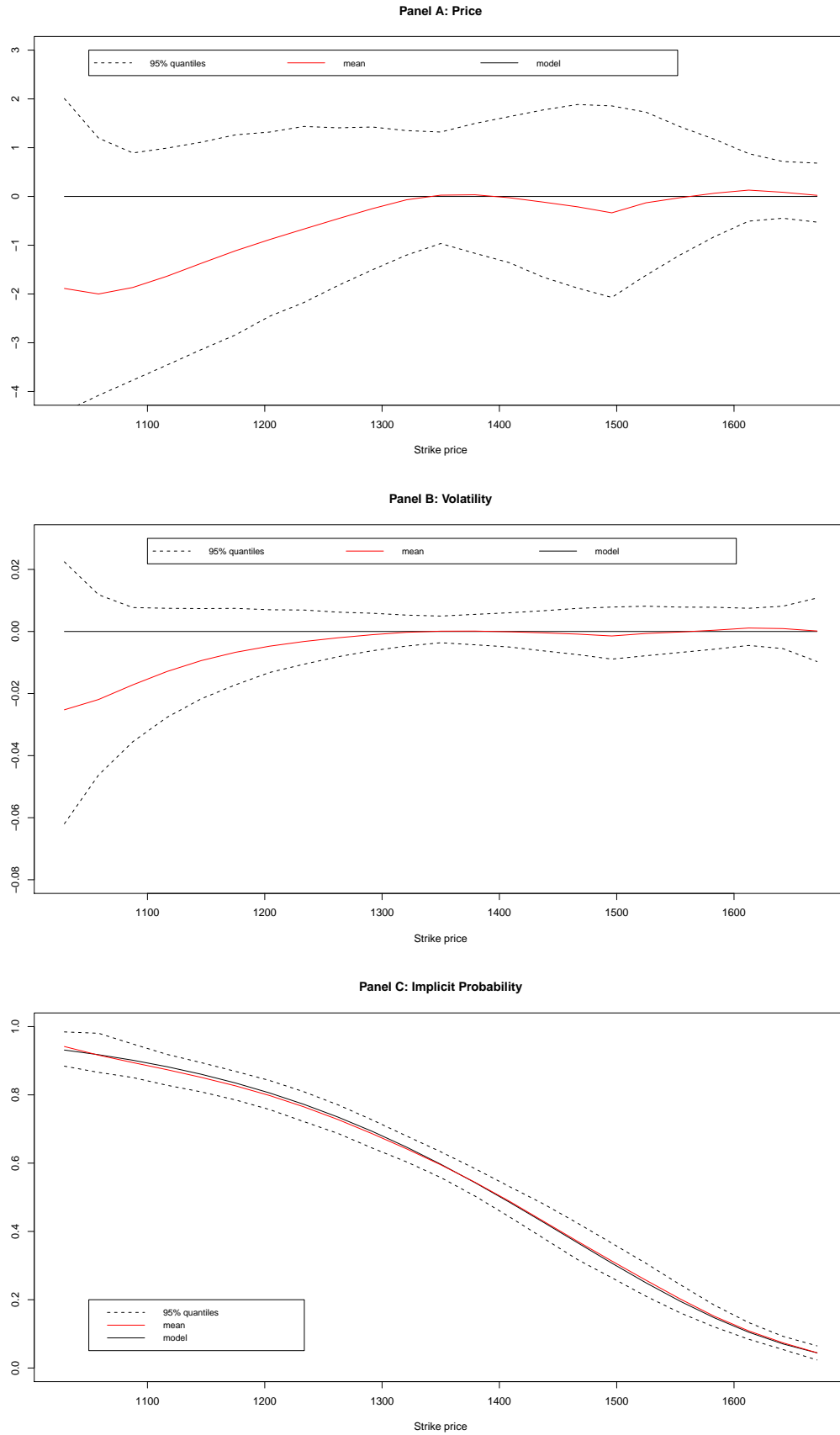
FIGURE 4. Simulated 95% Confidence Intervals for the case $\delta = 2$.
The confidence bands were obtained after 5.000 Monte Carlo simulations.

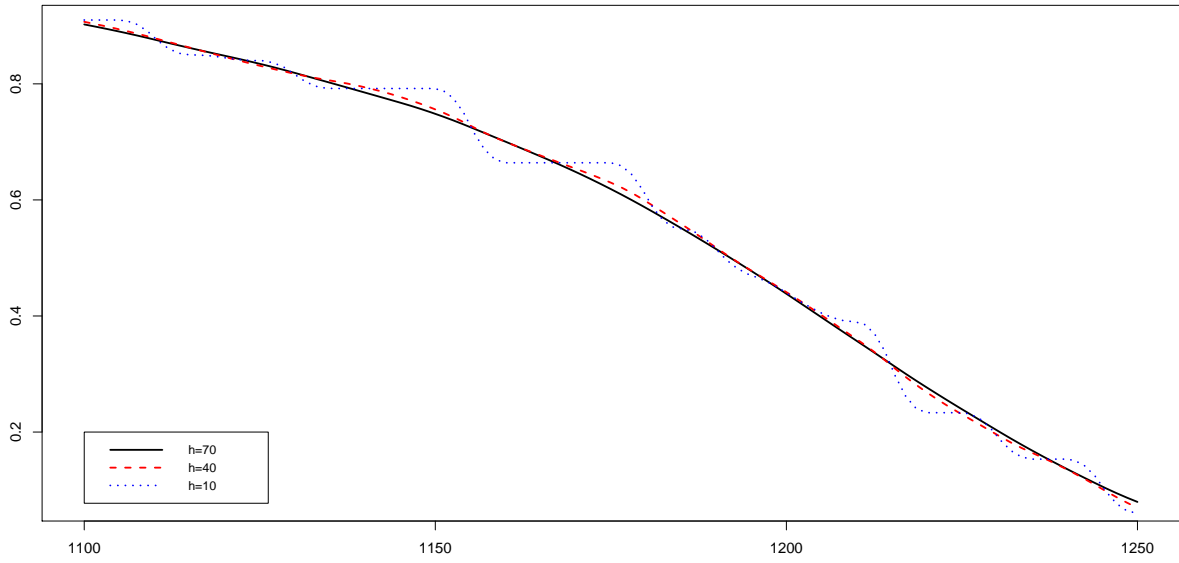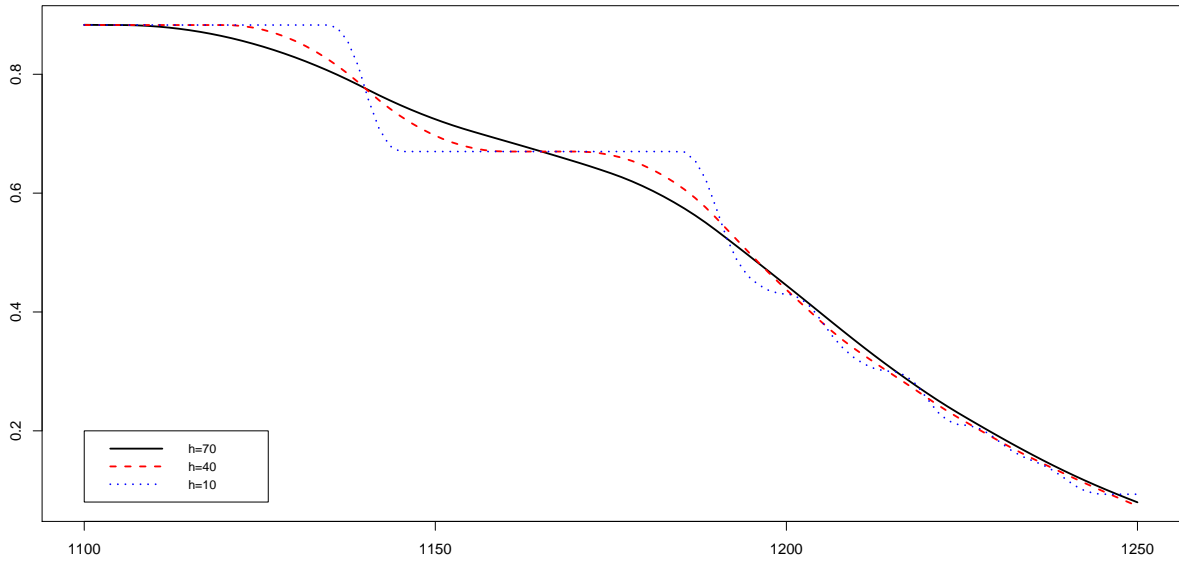**Panel A: Price**

Panel B: Volatility

Panel C: Implicit Probability

FIGURE 5.   Simulated 95% Confidence Intervals for the case $\delta = 10$.
The confidence bands were obtained after 5.000 Monte Carlo simulations.
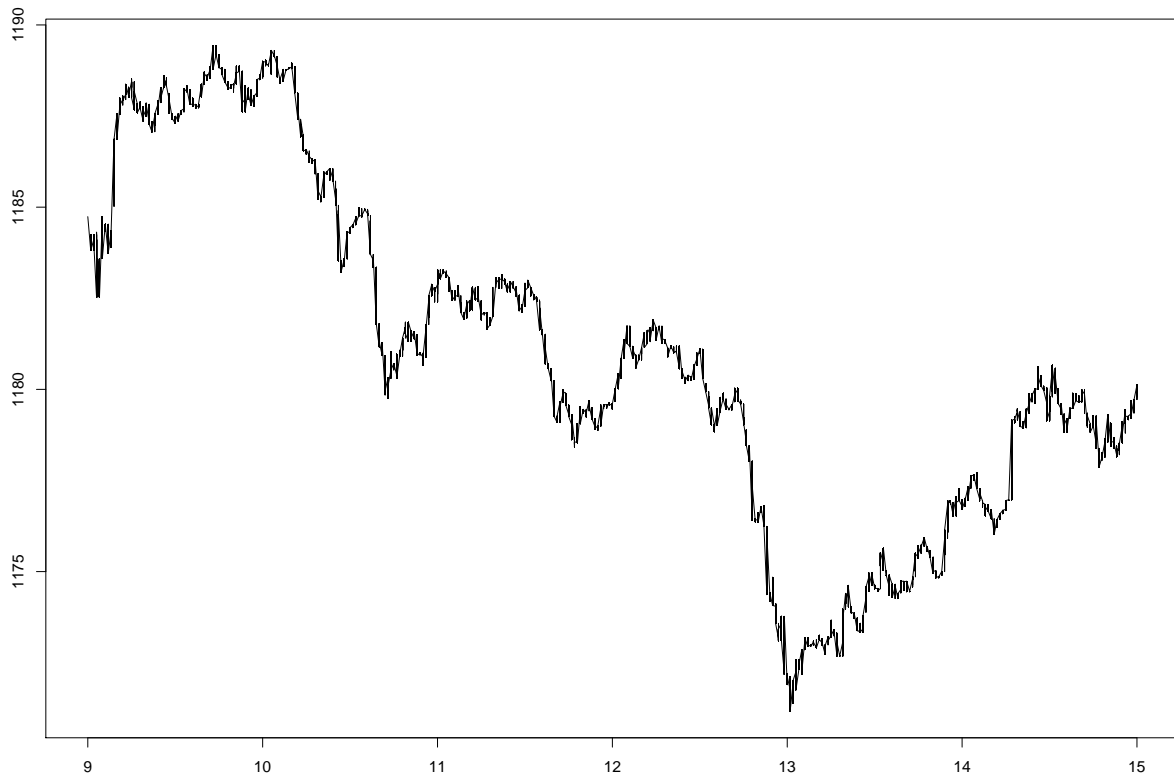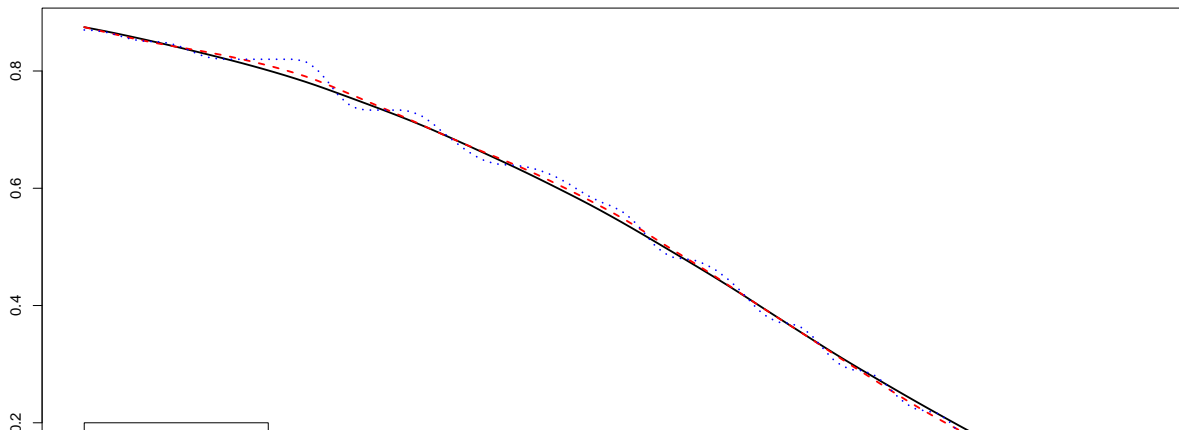
FIGURE 6. The S&P 500 Index on 21st October 2010.

Implicit probability at 12:06

Implicit Probability at 12:39

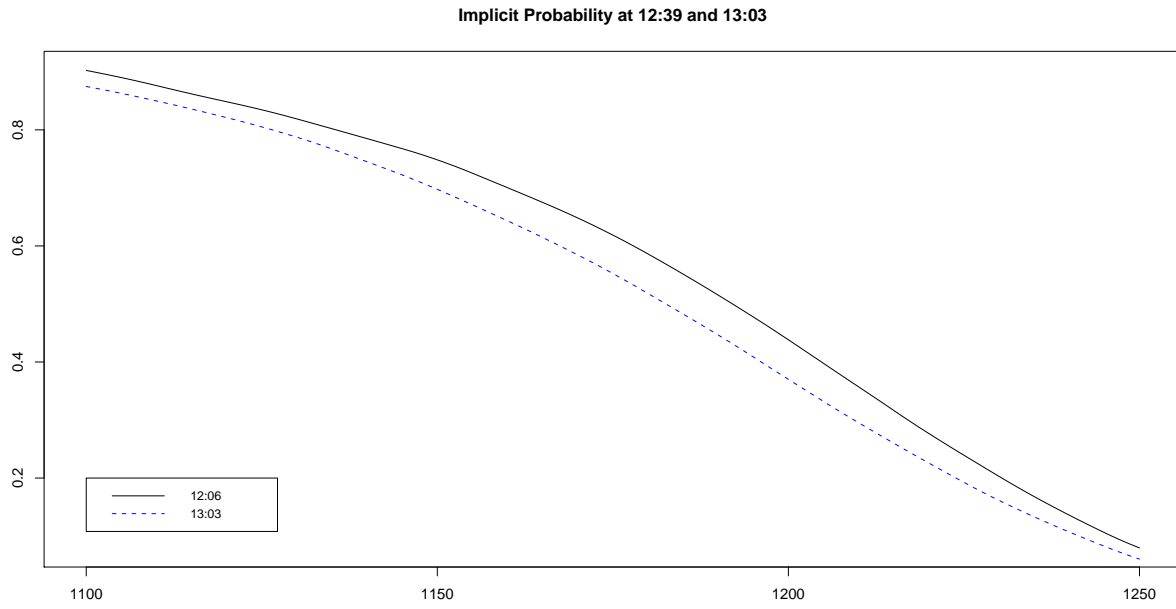**Implicit Probability at 13:03**

Implicit Probability at 12:39 and 13:03

FIGURE 8. $\nu^h(X > x)$ at 12:06 and 13:03 for the case $h = 70$.