J.R. Birge and V. Linetsky (Eds.), *Handbooks in OR & MS, Vol. 15*

DOI: 10.1016/S0927-0507(07)15012-X

Chapter 12

Incomplete Markets

Jeremy Staum

Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208-3119, USA
E-mail: j-staum@northwestern.edu
url: http://users.iems.northwestern.edu/~staum

Abstract

In reality, markets are *incomplete*, meaning that some payoffs cannot be replicated by trading in marketed securities. The classic no-arbitrage theory of valuation in a complete market, based on the unique price of a self-financing replicating portfolio, is not adequate for nonreplicable payoffs in incomplete markets. We focus on pricing over-the-counter derivative securities, surveying many proposed methodologies, drawing relationships between them, and evaluating their promise.

1 Introduction

Incomplete markets are those in which perfect risk transfer is not possible. Despite the ever-increasing sophistication of financial and insurance markets, markets remain significantly incomplete, with important consequences for their participants: workers and homeowners remain exposed to risks involving labor income, property value, and taxes, investors and portfolio managers have limited choices, and traders of derivative securities must bear residual risks. From a theoretical perspective, incomplete markets complicate the study of financial market equilibrium, portfolio optimization, and derivative securities.

Although the theory of derivative securities in complete markets is understood very well, and is the subject of numerous textbook accounts, there is as yet no fully developed, sound theoretical framework for pricing derivative securities in incomplete markets. This has profound consequences for the practice of trading, speculating, and hedging with derivative securities. This chapter surveys the topic of incomplete markets, with an emphasis on pricing and hedging derivative securities.

Other surveys have treated different aspects of incomplete markets. For portfolio optimization in incomplete markets, see Skiadas (2006). The finance literature emphasizes the existence and characteristics of equilibria, including market efficiency. Magill and Quinzii (1996) offer a book-length exposition, and Hens (1998) provides an overview with a low level of technicalities. Appendix B presents perspectives from the finance literature, not usually addressed in financial engineering, on the degree to which markets are actually incomplete, and the implications for welfare.

Surveys of derivative security pricing in incomplete markets include Jouini (2001), who covers no-arbitrage bounds, utility maximization, and equilibrium valuation, as an introduction to a special journal issue on these topics. Cont and Tankov (2004, Chapter 10) cover these approaches and others, including quadratic and entropy criteria, as well as calibration. Another survey is by Davis (2004b), whose "intention is not to aim at a maximum level of generality but, on the contrary, to concentrate on specific cases and solved problems which give insight into the nature of optimal strategies for hedging and investment." In contrast, we will cover all major approaches to pricing derivative securities in incomplete markets, as well as providing enough background to evaluate them and understand them in relation to one another.

Thus, due to limitation of space, we will not be able to concentrate on specific derivative securities or models of markets, although we will give simple examples that illustrate major ideas. Likewise, we can neither recount the development of each method nor provide an exhaustive list of references, so many significant papers will not be mentioned. Instead, we will merely provide references to the literature as a substitute for an exposition of the technical details of all the methods we survey, for which there is also not space. However, we address the technicalities of defining incompleteness in Appendix A.

We begin with background for the problem of incomplete markets. In Section 2, there is a description of the over-the-counter market for derivative securities and the financial engineering problems we will address. The causes of incomplete markets are addressed in Section 3. Next, we turn to general theoretical considerations about pricing in incomplete markets. The connections between pricing and optimization occupy Section 4, which covers no-arbitrage bounds, indifference prices, good deal bounds, and minimum-distance pricing measures. In Section 5, simple examples based on expected utility illustrate issues in pricing and optimization. Subsequent sections are devoted to various particular methods. The quadratic approach to hedging occupies Section 6. Exponential utility, with its connection to relative entropy, is the topic of Section 7. Several methods based on considering only losses, not gains, appear in Section 8: these include partial replication schemes such as quantile hedging. Restrictions on pricing kernels, including methods based on low-distance pricing kernels, are covered in Section 9. Ambiguity and robustness to model risk is the topic of Section 10. The standard practice of calibrating a model to market prices occupies Section 11. In Section 12, we offer some conclusions, evaluation, and directions for future research.

2 The over-the-counter market

Let us imagine ourselves in the position of a market-maker in an over-the-counter (OTC) derivatives market. Throughout this survey, we will consider incomplete markets from the market-maker's perspective, focusing on the financial engineering of solving the problems of pricing and risk management. The same considerations apply to customers in the OTC market.

2.1 The workings of OTC markets

Although some derivative securities, including some stock options and commodity or currency futures, are listed on exchanges and traded in the same manner as the underlying securities, many are not. A hedger or speculator who wishes to trade them must participate in the OTC market by calling OTC market-makers, usually at investment banks, and requesting a quote for bid and ask prices at which the market-makers are willing to buy or sell, respectively. Duffie et al. (2006) address the relationship of frictions and liquidity in OTC markets to valuation. We will focus on the process by which market-makers prepare these prices. Through an analogous process, the potential customer must then decide whether to sell at the highest of the quoted bid prices, buy at the lowest of the quoted ask prices, or do nothing.

If the customer indeed transacts a deal with a market-maker, the market-maker must bear risk associated with this trade, because markets are incomplete. In order to measure the risk of his portfolio and manage it through hedging, he needs to model the future value of the OTC derivatives he has traded. As time passes, he must track the profit or loss generated by his hedged portfolio, based on values of OTC derivatives updated in light of current market prices, a process known as *marking to market*. It is a matter of debate among practitioners whether and when it is appropriate to mark to market using a bid price, a "mid-market price" between the bid and ask prices, or an "unwind price" at which the derivative might be sold. Marking to market is not studied enough relative to pricing, but the risk-adjusted value processes of Artzner et al. (2007) may be useful in this regard.

Establishing bid and ask prices for an OTC derivative security is not the same as determining the equilibrium price for a new security if it were to be listed on an exchange, which is another goal often considered in the literature on incomplete markets. Determining the equilibrium price is more difficult than it might seem, because introducing a new security could alter the existing security prices (Boyle and Wang, 2001). In financial engineering, although equilibrium concepts may be useful in pricing, it is too ambitious to attempt to construct an entire equilibrium, based on the preferences and endowments of all participants. This is more appropriate in finance, where one may use a simplified model to formulate a hypothesis or explain some phenomenon.

2.2 Standard practice

Whether using a model in which markets are complete or incomplete, derivatives traders know that markets are actually incomplete, and that after trading, they will not be able to hedge away all the risk, to which they are averse. Nonetheless, their standard practice is to assign prices to OTC derivative securities primarily on the basis of consistency with the market prices of underlying and other derivative securities. We will further discuss and evaluate this standard practice in Section 11.

According to the classic theory of financial engineering, in a complete market, the unique no-arbitrage price of a derivative security whose payoff is X is the expected discounted payoff $\mathrm{E}_{\mathbf{Q}}[DX]$ under the risk-neutral probability measure $\mathbf{Q}$, under which the marketed securities' expected returns equal the risk-free rate of interest. Traders calibrate the parameters of $\mathbf{Q}$ to prices of marketed securities so as to minimize the discrepancy between these market prices and the prices given by the model, i.e. the expected discounted payoffs. To recoup their business expenses and to earn compensation for bearing the risks that they will not be able to hedge, traders establish a bid–ask interval around the expected discounted payoff. The exact level of the bid and ask depend on informal consideration of several factors, such as how the trade will affect the portfolio's Greeks, the trader's outlook on likely market events, what the competition is charging, and the relationship with the customer. One of the major challenges facing financial engineering in the area of derivative securities is to establish a sound basis for this pricing decision, based on quantitative risk assessment using models of incomplete markets.

If the market is incomplete, then pricing by calibration of a complete-market model does not systematically account for the costs of hedging or the risks that remain after hedging. This approach wrongly prices the unhedgeable part of the risk as though it too could be hedged away; it assigns to a derivative security the same price as a fictitious replicating portfolio strategy, when this strategy will not actually succeed in replicating the target payoff. As Foldes (2000) says,

> Enthusiasm for methods of hedging and valuation of derivatives in complete markets, and for associated methods of computation, seems often to obscure the fact that these techniques do not provide a general theory of valuation and that they are liable to give at best only imprecise results when applied beyond their proper domain.

The need to quantify and value residual risks motivates the search for a practical method of pricing with incomplete-markets models.

2.3 The apparent and real problems

The apparent problem of pricing in incomplete markets is mathematical: given the statistical probability measure $\mathbf{P}$, there is a set $\mathcal{Q}$ of *equivalent martingale measures* (EMMs) such that the expected discounted payoff $\mathrm{E}_{\mathbf{Q}}[DX]$ is an arbitrage-free price for X.[1] There is an interval

$$\Big(\inf_{\mathbf{Q}\in\mathcal{Q}} \mathrm{E}_{\mathbf{Q}}[DX],\ \sup_{\mathbf{Q}\in\mathcal{Q}} \mathrm{E}_{\mathbf{Q}}[DX]\Big) \tag{1}$$

of arbitrage-free prices for X, and it is usually too wide for these *no-arbitrage bounds* (Section 4.2.1) to serve as useful bid and ask prices (see e.g. Eberlein and Jacod, 1997). The problem may appear to be that we want a way of choosing one of the pricing measures $\mathbf{Q} \in \mathcal{Q}$, so that we may then assign the unique price $\mathrm{E}_{\mathbf{Q}}[DX]$ to each payoff X.

Another way to view the situation is that the no-arbitrage criterion allows a multiplicity of possible *pricing kernels* Π. A pricing kernel $\Pi = D\,\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P}$ where $\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P}$ is the likelihood ratio, i.e. Radon–Nikodym derivative, between some $\mathbf{Q} \in \mathcal{Q}$ and $\mathbf{P}$. The value $\Pi(\omega)$ of the pricing kernel in state ω can be interpreted as the price now for \$1 to be paid if state ω occurs. With no restriction on the pricing kernel, the price can be anywhere within the no-arbitrage price bounds. However, some of these pricing kernels may seem implausible from an economic perspective. See Section 9 for methodologies that work by eliminating implausible pricing kernels.

The real problem of pricing in incomplete markets depends on the objective. For example, a goal in setting bid and ask prices is to ensure that any trade undertaken at these prices is advantageous to the firm. A grounding of the pricing scheme in financial economics would be desirable. It is not clear how selecting a single pricing measure $\mathbf{Q} \in \mathcal{Q}$ will accomplish this goal; indeed, given a unique price $\mathrm{E}_{\mathbf{Q}}[DX]$, further considerations would be required to generate distinct bid and ask prices. Another objective is marking to market, in which the goal is to assign to the firm's portfolio of derivative securities a value, not a price, that is accurate from an accounting or actuarial perspective. Again, for risk management, there may be different goals that involve assessing the future value of derivative securities. However, in all cases, we want a methodology that respects the no-arbitrage bounds, is computationally efficient, and is robust to those errors that are likely in specifying its inputs, e.g. to stale prices of marketed securities, or to estimation error of statistical probabilities.

In constructing bid and ask prices, the difficulty posed by incomplete markets is more significant than it might at first seem, because of adverse selection. If the ask price is too high, few potential customers will be willing to pay so much, and the result is forgone profits. If the ask price is too low, the resulting trade is bad for the firm and good for the customer, which entices many customers to make trades that entail likely loss for the firm. For example, Dunbar (2005) describes an incident in which it was thought that a large portion of a \$200 million loss by JP Morgan could be attributed to this adverse selection:

> ... by selling a swaption straddle that expired the day before a non-farm payroll announcement and buying one that expired immediately after, a hedge fund

[1] For precise details, see Appendix A. We assume an arbitrage-free market.

> could profit from [the] potential volatility. However, a dealer on the other side of this one-day calendar spread trade might find it difficult to hedge its position over such a short interval of time, and ought to price this risk into the trade, or not undertake the trade at all. But JP Morgan seemed to lack such caution, say market sources, and in effect offered 'lottery tickets' to the market.

As we see from this example, a calibrated model can assign a wrong price. The price for the calendar spread was consistent with market prices, but did not account for an unusual feature of the statistical probability measure **P**: interest rate volatility is concentrated at the date of an important news announcement. See Section 11 for further discussion.

3 Causes of incompleteness

Several phenomena cause incompleteness. One is an insufficiency of marketed assets relative to the class of risks that one wishes to hedge, which may involve jumps or volatility of asset prices, or variables that are not derived from market prices. Market frictions, such as transaction costs and constraints on portfolios, may also cause incompleteness. A source of effective incompleteness is ambiguity, i.e. ignorance of the true stochastic model for market prices: it is effectively the same if it is impossible to transfer risk perfectly or if one merely does not know how to do so.

3.1 Insufficient span of marketed assets

Markets are incomplete with respect to payoffs that are not entirely determined by market prices: examples include weather derivatives, catastrophe bonds, and derivatives written on economic variables such as gross domestic product. Corporate investment projects provide another example; real options analysis applies to a valuation problem in an incomplete market.

Features such as jumps and stochastic volatility of marketed asset prices may also cause incompleteness, depending on the available trading opportunities. For example, in the Heston (1993) model of a stock with stochastic volatility and a bond with constant interest rate, the market is incomplete because it is not possible to hedge the risk factor associated with stochastic volatility. However, if an option on the stock were also to be marketed, both risk factors could be hedged by trading in stock and option, and the market would be complete. Jumps tend to cause incompleteness except in very simple or unusual models (see e.g. Dritschel and Protter, 1999). Whereas in the Black–Scholes model, delta is the hedge ratio that matches the locally linear dependence of an option's value on infinitesimal changes in the stock price, it is not so easy to hedge against potential jumps of various sizes, because value is not linear. To complete a market in which jumps of all sizes are possible might require many more marketed securities, for example, vanilla European options of all strikes and maturities.

Jumps and stochastic volatility are important as ways to model volatility smiles. A primary alternative is a local volatility model, in which the market is complete. However, the local volatility model is criticized (e.g. by Davis, 2004a, §2a) for the crucial, counterfactual assumption that an asset's volatility is a function of its price: precisely the absence of a second risk factor, which makes the model complete, prevents it from saying anything about volatility risk and vega hedging.

For evidence that it may be necessary to model jumps, or jumps and stochastic volatility, in describing equity or equity index returns adequately, see Andersen et al. (2002), Carr et al. (2002). The most realistic models imply incomplete markets.

3.2 Market frictions

Constraints produce incompleteness by forbidding portfolio strategies that replicate some payoffs. For example, an executive who is granted stock options is not supposed to hedge them by selling stock in the company. Different interest rates for borrowing and lending may be modeled by constraints: where $r_b > r_\ell$ are the rates for borrowing and lending respectively, only positive shares of a money market account paying rate r_ℓ and negative shares of one paying rate r_b are allowed.

Transaction costs produce incompleteness less straightforwardly. Continuous-time portfolio strategies accrue transaction costs at every instant the portfolio is rebalanced. These strategies are effectively forbidden if their costs are infinite, which can happen, for instance, in the Black–Scholes model because of the infinite first variation of geometric Brownian motion. Fixed and proportional transaction costs are the most frequently studied; the latter are equivalent to bid–ask spreads for marketed securities. There is a substantial literature on the topic, looking back to Hodges and Neuberger (1989). More recent work on the topic includes Clewlow and Hodges (1997).

Rather than model transaction costs explicitly, one might use a model in which trading is allowed only at a fixed, discrete set of times. This also eliminates continuous-time strategies that would incur infinite costs, and it can be more tractable; however, rebalancing the portfolio at fixed times is typically not as good as rebalancing at a finite number of random times.

3.3 Ambiguity

Suppose a stock index follows a geometric Brownian motion whose volatility is known to be 20%. How many years' data are required to construct a 95% two-sided confidence interval of width 1% for the drift? The answer is 6,147: this yields a width of approximately $2 \times 1.96 \times 20\%/\sqrt{6147} = 1\%$. On the other hand, according to this Black–Scholes model, knowledge of the drift is unnecessary for option pricing, and the volatility can be estimated perfectly by observing any time interval, no matter how short. This has to do with the nonequivalence of Black–Scholes models with different volatilities, but it is merely an artifact of the continuous-time model. In reality, estimating volatility from high-frequency data is quite difficult (Zhang et al., 2005). Moreover, a cursory examination of financial time series shows that, for instance, daily historical volatility has varied dramatically from year to year. Ambiguity about volatility is so important that, according to Carr (2002), a frequently asked question in option pricing is whether one should hedge at historical or implied volatility. Carr (2002, §IX) provides a formula for the error resulting from hedging a derivative security at the wrong volatility, given a diffusion model. The hedging error can be quite substantial.

4 Pricing and optimization

Pricing can be grounded in portfolio optimization (Sections 4.1–4.2) or in an optimization over pricing measures (Section 4.4).

4.1 Portfolio optimization

Conditions for the existence of optimal portfolio strategies and related probability measures have attracted much attention. There may be no optimal strategy or measure if there is a sequence of them converging to a limit point that is excluded from the feasible set, or if the optimization problem is unbounded. If the limiting strategy is infeasible, one may be satisfied to choose a nearly optimal strategy. When the problem is unbounded, usually something is wrong with the way it has been posed. For example, if there is no bound on the expected utility one can attain by investing, it may be that the set of allowed strategies is unrealistically large, the utility function is unsuitable, or the probability measure is erroneous.

In the interests of simplicity, we will not treat the question of the existence of an optimal solution: the interested reader can find precise results in the literature cited in the sections on specific methodologies. We will also speak primarily of optimizing random wealth at a fixed future date, and the connected problem of pricing payoffs at that date, although the same ideas apply to continuous consumption streams, American options, etc. We ignore the structure of the portfolio strategies, which could be a single vector of weights determining a static portfolio in a one-period problem, or a continuous-time vector stochastic process, or something in between, focusing instead on the payoffs they provide. Expository treatments of portfolio optimization include Karatzas and Shreve (1998), Schachermayer (2002), Skiadas (2006).

However, we will now consider briefly two issues in formulating an optimization underlying a pricing scheme: whether the optimization takes into account only the market risk of the OTC trade itself or also accounts for the opportunities for future trades, and whether the portfolio strategy is instantaneously, myopically optimal or optimal over an entire time interval.

4.1.1 Opportunity

Accounting for changing investment opportunities leads to better portfolio strategies. The optimal portfolio can be decomposed into a term that would be optimal if asset returns were independent, plus a term that corrects for the dependence of current asset returns and the conditional distribution of all future asset returns. For example, suppose that there is a riskless asset and one risky stock whose log price follows a diffusion with stochastic drift and volatility. A state with a higher ratio of drift to volatility constitutes a more favorable investment opportunity (cf. the Sharpe ratio) and thus a greater certainty equivalent for wealth. Suppose further that the change in this drift-volatility or *mean–variance ratio* is negatively correlated with the asset return. The optimal allocation to the stock is greater than it would be if the mean–variance ratio were deterministic: a loss from investing in the stock is cushioned by an increased certainty equivalent for each dollar of wealth. This increased demand for stock in the optimal portfolio is *hedging demand*. For a very lucid theoretical account of this phenomenon in the context of quadratic hedging (Section 6), see Schweizer (1995, especially p. 16). Extensive numerical results for hedging options occupy Brandt (2003); Example 4.1 is related.

Analogously, for an OTC market-maker, there is a stochastic process of OTC trade opportunities, i.e. requests for a quote of bid and ask prices by a potential customer, where each customer has reservation prices below or above which he is willing to buy or sell. Routledge and Zin (2004) take a step in this direction, which merits greater attention. The methods of pricing covered in this survey all focus on whether an individual OTC trade is attractive to the market-maker without considering its effect on future trades. However, a trade done now affects the portfolio the trader will have in the future, and in light of which he will evaluate future trades. For example, if there is a risk constraint, doing an OTC trade now might prevent the trader from doing a more attractive trade in the future. Therefore, each opportunity should be evaluated in light of the stochastic process of future opportunities: the compensation for doing a trade should reflect the direct cost of possible losses and also the indirect cost of lost opportunity for profit on future trades that may be passed up due to the risk associated with this trade.

4.1.2 Local vs. global

In pricing an OTC security, a *global* optimization optimizes over portfolio strategies that cover an entire time interval. This may be difficult to solve, whether numerically or analytically, or even to set up. A simpler alternative is a *local* optimization, in which the objective and constraints contain only static criteria, changes over a single time step, or instantaneous rates of change. A local optimization optimizes over the current portfolio weights only: whether one intends to hedge dynamically or not, a local optimization is a static problem, in a sense.

Global criteria include terminal wealth, total utility from consumption over an entire time interval, value at risk, and squared hedging error. The global criteria can be constraints as well as objectives, for example, the constraint that the wealth process never be negative. Local criteria include Greeks and often form pairs with global criteria. For example, in quadratic hedging there is a locally risk-minimizing strategy and a global variant, the variance-optimal hedge: see Example 6.1. Analogous to the usual expected utility maximization (Section 4.2.3) is the local utility maximization Kallsen (2002a), discussed in Section 9.2 similar to the following example based on Schweizer (1995, §5), but in continuous time.

Example 4.1. There is a riskless bank account whose value is always \$1, and a risky asset whose prices are given by Table 1. An investor has \$100 of initial wealth and utility function $u(W) = -(W/100)^{-4}$. The investor maximizes the expected utility of wealth at time 2 over self-financing strategies: the decision variables are ξ_1, the number of shares of the risky asset to hold over the first step, and $\xi_2^{(+)}$, $\xi_2^{(0)}$, and $\xi_2^{(-)}$, the number of shares to hold over the second step, respectively if the risky asset price at time 1 is 1, 0, or −1.

Local optimization of one-step expected utility in each of the four scenarios yields $\xi_1 = \xi_2^{(0)} = \xi_2^{(-)} = 0$ and $\xi_2^{(+)} = 16.13$: only when the risky asset's price is \$1 at time 1 is its one-step expected return positive, so that it is worth investing in it, from a local perspective. A global optimization of two-step expected utility yields $\xi_1 = -3.27$, $\xi_2^{(0)} = \xi_2^{(-)} = 0$, and $\xi_2^{(+)} = 15.60$: the negative position in the risky asset over the first step hedges the increase in the derived utility of wealth at time 1 if the asset's price should rise. See Example 6.1 for a continuation.

That is, local optimization ignores hedging demand, while global optimization captures it. Intermediate wealth is worth more in states with better investment opportunities, and the global optimization yields greater expected utility from terminal wealth by producing more wealth in the intermediate states with poorer investment opportunities.

Table 1.
Risky Asset Prices.

State	Probability	Time 0	Time 1	Time 2
1	1/9	\$0	\$1	\$3
2	1/6	\$0	\$1	\$2
3	1/18	\$0	\$1	\$0
4	1/6	\$0	\$0	\$1
5	1/6	\$0	\$0	−\$1
6	1/6	\$0	−\$1	\$0
7	1/6	\$0	−\$1	−\$2

4.2 Pricing via portfolio optimization

No-arbitrage bounds (Section 4.2.1) and indifference prices (Section 4.2.2) are special cases of the mathematical structure of *good deal bounds* (Section 4.2.4). Let R be the set of replicable payoffs, $\pi(Y)$ be the market price to replicate a payoff $Y \in R$, and A be an *acceptance set* of payoffs that are acceptable compared to the status quo. The lower good deal bound for a payoff X, which might be interpreted as a bid price, is

$$b(X) = \sup_{Y \in R}\{-\pi(Y) \mid Y + X \in A\}. \tag{2}$$

If we can buy X over the counter for less than $b(X)$ then there is a Y that we can buy in the market for $\pi(Y)$, such that in total we get $X + Y$, which is acceptable, for a cost $b(X) + \pi(Y) < 0$. The upper good deal bound or ask price for X is

$$a(X) = -b(-X) = \inf_{Y \in R}\{\pi(Y) \mid Y - X \in A\}. \tag{3}$$

To sell X or to buy $-X$ has the same effect. The other minus sign in $a(X) = -b(-X)$ reflects the convention that the buyer pays the price to the seller. Because of the relationship $a(X) = -b(-X)$, one may specify only b (or a), getting distinct price bounds unless b is antisymmetric.

The interpretation of $-b(X)$ is the cost of rendering X acceptable, and this can be thought of as a risk measure. As Jaschke and Küchler (2001, n. 6) say, "any valuation principle that yields price bounds also induces a risk measure and vice versa." Indeed, under some conditions, $-b$ is a coherent or convex risk measure (Artzner et al., 1999; Föllmer and Schied, 2002). The no-arbitrage bounds provide an example. For generalities, see Jaschke and Küchler (2001, Prop. 7) and Staum (2004, Prop. 4.2).

The acceptance set A must include $\{Z \mid Z \geqslant 0\}$, the set of riskless payoffs, which is the acceptance set that generates no-arbitrage bounds. It must not intersect the set $\{Z \mid Z < 0\}$ of pure losses with no chance of gain. Finally, $Z \in A$ and $Z' \geqslant Z$ must imply $Z' \in A$. These three properties correspond to a subset of the axioms defining coherent risk measures (Artzner et al., 1999).

The acceptance set A must also be consistent with market prices π, or arbitrage will result. For example, if there is an acceptable payoff $Y \in A$ with negative cost $\pi(Y) < 0$, then $b(0) > 0$, and the trader is thus expressing willingness to give money away in exchange for nothing. For a concrete example using expected utility indifference pricing, see Section 5.2.1. For general remarks, related to duality, see Section 4.2.5.

4.2.1 No-arbitrage pricing

The *no-arbitrage price bounds* are given by Eqs. (2) and (3) with the acceptance set $A = \{Z \mid Z \geqslant 0\} = \{Z \mid \operatorname{ess\,inf} Z \geqslant 0\}$:

$$b_{NA}(X) := \sup_{Y \in R}\{-\pi(Y) \mid Y + X \leqslant 0\} = -\inf_{Y \in R}\{\pi(Y) \mid Y \geqslant -X\} \tag{4}$$

and

$$a_{NA}(X) := \inf_{Y\in R}\{\pi(Y) \mid Y - X \leqslant 0\} = \inf_{Y\in R}\{\pi(Y) \mid Y \geqslant X\}. \tag{5}$$

That is, a payoff is acceptable if and only if it has no risk of loss under the statistical probability measure **P**. Buying X for less than $b_{NA}(X)$ or selling it for more than $a_{NA}(X)$ admits arbitrage. For instance, for any $\epsilon > 0$, there is a $Y_\epsilon \in R$ such that $\pi(Y) < \epsilon - b_{NA}(X)$ and $Y_\epsilon \geqslant -X$. Thus, if we buy X for $b_{NA}(X) - \epsilon$ and also buy Y_ϵ, we acquire $Y_\epsilon + X \geqslant 0$ (this is *super-replication* of $-X$) for a negative total cost: we get paid now and assume no risk of loss. El Karoui and Quenez (1995) give a dynamic programming algorithm for computing the no-arbitrage bounds.

While $-\operatorname{ess\,inf} X$ measures the worst possible loss X can yield, $-b_{NA}$ is also a risk measure, measuring the cost of hedging to prevent the worst possible loss. The solution Y^* to Problem (4) is an optimal hedge for X: it is the cheapest payoff that combines with X to produce a portfolio with zero probability of loss. The typical result for a complete market is that $Y^* = -X$, X solves Problem (5), and $b_{NA}(X) = a_{NA}(X) = \pi(X)$, the cost of replicating X. In an incomplete market, $b_{NA}(X)$ and $a_{NA}(X)$ are usually too low and too high, respectively, to be of use to an OTC market-maker: few customers would be willing to trade at such prices (see e.g. Eberlein and Jacod, 1997).

4.2.2 *Indifference pricing*

Indifference prices are good deal bounds with acceptance set $A = \{Z \mid P(Z) \geqslant P(0)\}$ in Eqs. (2) and (3), where P is a *preference function* specifying *complete preferences*. Completeness of preferences is different from completeness of markets: it means that for any pair of payoffs X and Y, either one prefers X to Y, is indifferent between X and Y, or prefers Y to X. With a preference function, these three cases correspond to $P(X) > P(Y)$, $P(X) = P(Y)$, and $P(X) < P(Y)$ respectively. Buying X for less than $b(X)$ results in a nonnegative cost to acquire a total payoff $X + Y$ that is at least as good as the status quo, i.e. $P(X + Y) \geqslant 0$.

The main point of indifference pricing is not the mathematics of a preference function versus an acceptance set; it is possible to convert between them as for risk measures and acceptance sets (Jaschke and Küchler, 2001). The point is the interpretation of the set A as the set of *all* payoffs that are at least as good as the status quo. The no-arbitrage bounds are not to be interpreted as indifference prices. They have the form of indifference prices with P equal to the essential infimum, which is far too conservative: it says that zero is preferable to any payoff with a positive probability of loss.

Indifference pricing takes place against the background of the portfolio optimization problem

$$\sup_{Y\in R}\{\tilde{P}(W + Y) \mid \pi(Y) \leqslant c\} \tag{6}$$

where the initial endowment consists of c dollars and the random wealth W, and $\tilde{P}$ is a preference function over the total random wealth. Then $V^* = W + Y^*$ is the total random wealth produced by the trader's optimal portfolio strategy. A trader who has the opportunity to purchase X over the counter formulates the problem

$$b(X) = \sup_{Y \in R} \{ -\pi(Y) \mid \tilde{P}(V^* + X + Y) \geqslant \tilde{P}(V) \} \tag{7}$$

to find the indifference bid price. If Y^* solves Problem (7) with the constraint tight, then the trader is indeed indifferent between $V^* + X + Y^*$ and V^*, i.e. between doing and not doing the trade at $b(X)$.

Problem (7) coincides with Problem (2) when $P(Z) = \tilde{P}(V^* + Z) - \tilde{P}(Z)$. That is, preferences over payoffs, which are changes in wealth, used in constructing indifference prices, are derived from more fundamental preferences over total wealth. Therefore, preferences over payoffs depend on the optimal total random wealth V^* in Problem (6). For various reasons, e.g. that the procedure takes too long or that one does not trust its results, one may wish to avoid solving Problem (6) first, instead simply solving Problem (7) with V, determined by the status quo portfolio strategy, replacing the optimal V^*. However, Problem (7) can be quite sensitive to V, and if $V \neq V^*$, the indifference price can violate the no-arbitrage principle. There is an example and further discussion in Section 5.2.1. Another way of dealing with this situation is to formulate indifference prices by incorporating the portfolio optimization problem (6):

$$b(X) = \sup_{Y \in R} \Big\{ c - \pi(Y) \mid \tilde{P}(X + Y) \geqslant \sup_{V \in R} \{ \tilde{P}(V) \mid \pi(V) \leqslant c \} \Big\} \tag{8}$$

based on an initial budget of c.

4.2.3 Expected utility

Expected utility theory specifies the preference function as $\tilde{P}(W) = \mathrm{E}[u(W)]$, where the *utility function* u is increasing because more money is better and concave because of risk aversion. It is characteristic of expected utility indifference pricing that $a(X) \neq b(X)$ for a typical nonreplicable payoff X: it leads to price bounds, not a unique price, because of aversion to risk that cannot be hedged. As Musiela and Zariphopoulou (2004b) emphasize, "no linear pricing mechanism can be compatible with the concept of utility based valuation," so we should not expect to have the ask price $a(X) = -b(-X)$ equal to the bid price $b(X)$. Marginal indifference pricing (Section 4.3) delivers a unique price based on expected utility.

Expected utility indifference pricing is difficult to implement in the context of derivative security pricing. The key inputs to expected utility maximization are the endowment V, the statistical probability measure $\mathbf{P}$, and the utility function u. As Carr et al. (2001, §1) observe,

> Unfortunately, the maximization is notoriously sensitive to these inputs, whose formulation is suspect at the outset. This shortcoming renders the methodology potentially useless ...

OTC traders prefer calibration (Section 11), which does not require them to specify the endowment, the utility function, or the parameters of **P**, but only the form of the pricing measure **Q**. In particular, it is not required to estimate the expected return of marketed assets under **P**, which is difficult (Section 3.3), but of the utmost importance for expected utility maximization. It is likewise difficult to determine an appropriate utility function in the corporate setting of making a market in OTC derivatives. What is the basis for corporate risk aversion? The view of the equity of a firm with debt as a call option on the firm's value suggests that shareholders should be risk-seeking, so as to maximize the value of this call option. Does corporate risk aversion come from regulatory capital requirements, or from financial distress costs (for which see Jarrow and Purnanandam, 2004, and references therein), and if so, how is it to be quantified? To model the firm's endowment, one ought to include not only all securities, loans, and liabilities currently on the books, but also future business earnings as a going concern: for instance, one would want to know the dependence between portfolio returns and earnings from doing advisory work on mergers and acquisitions.

Aside from these perplexities in modeling, continuous-time expected utility maximization also involves difficult technicalities. For example, it is not easy to pick a suitable set of portfolio strategies over which to optimize (Delbaen et al., 2002; Kabanov and Stricker, 2002; Schachermayer, 2003). Work has continued in this area, to clarify the conditions that are necessary for existence of optimal portfolios and unique prices (Hugonnier and Kramkov, 2004; Hugonnier et al., 2005; Karatzas and Žitković, 2003). Schachermayer (2002) and Skiadas (2006) give expository treatments of the problem of expected utility maximization in a continuous-time incomplete market, providing a basis for indifference pricing.

4.2.4 Good deal bounds

The acceptance set A for use in the good deal bounds (2) and (3) includes only payoffs that are preferable to the status quo, but possibly not all of them. At prices below $b(X)$, it is preferable to buy; at prices above $a(X)$, it is preferable to sell. In indifference pricing, A contains all payoffs preferable to the status quo, so at prices between $b(X)$ and $a(X)$ it is preferable to do nothing. Otherwise, $b(X)$ is a lower bound on the indifference bid price and $a(X)$ is an upper bound on the indifference ask price, and the best policy at prices between $b(X)$ and $a(X)$ is unknown. This is the difference of interpretation between good deal bounds and indifference prices, which are a mathematical special case of the former.

There are two alternative interpretations of good deal bounds. One treats good deal bounds as possible bid and ask prices for a market-maker, much like indifference prices: see e.g. Cochrane and Saá-Requejo (2000, p. 86), Carr et al. (2001, §7), Staum (2004), and Larsen (2005, §5). This is financial engineering, with the goal of making only subjectively good deals, by trading outside the *subjective* price bounds, buying below $b(X)$ and selling above $a(X)$. The other interpretation is that A is a subset of the payoffs that many traders prefer to the status quo, as in Section 9. It treats good deal bounds like no-arbitrage bounds, asserting that good deals should not be available, because almost everyone would be willing to take them: see e.g. Cochrane and Saá-Requejo (2000, p. 82), Carr et al. (2001, §1), and Černý and Hodges (2002). This may be mathematical finance, with the goal of making a more precise statement about observed prices in incomplete markets than does the no-arbitrage principle. However, if these *objective* good deal bounds are narrow enough, they indeed offer market-makers useful guidance about prices: they should buy below $a(X)$ and sell above $b(X)$. In such a trade, the counterparty sells below $a(X)$ and buys above $b(X)$, not receiving a good deal from the market-maker. If the counterparty also insists on buying below $a(X)$ and selling above $b(X)$, trades take place inside the price bounds, so that neither party gets a good deal.

So far we have been discussing an abstract framework. How can it be given economic content by specifying the acceptance set A? The primary approaches include restrictions on the pricing kernel (Section 9) and robustness (Section 10). A simple version involves a convex risk measure formed by a finite number of valuation measures and stress measures with floors (Carr et al., 2001; Larsen et al., 2005), which might be specified by looking at the marginal utility and the risk management constraints of several market participants (Carr et al., 2001, §2). In Section 8, we consider methods that yield price bounds and have the same mathematical form as good deal bounds, except that they use acceptance sets A that violate the axioms in Section 4.2. This causes them to be unsuitable for OTC pricing, although they have other uses.

4.2.5 Duality

Duality provides Formula (1) for no-arbitrage bounds and related expressions for a good deal bound or indifference price as in Eq. (2): see Jaschke and Küchler (2001, §4) and Staum (2004, Thm. 4.1). It yields both computational advantage and insight. For example, in pricing a path-independent European option given continuous trading, the dual optimization is taken over a set of probability measures on terminal payoffs, which is more tractable than the set of continuous-time portfolio strategies appearing in the primal problem. The two major ways of grounding pricing in optimization involve the two sides of this duality: portfolio optimization is optimization over portfolios or the payoffs they provide, while the methods of selecting minimum-distance measures or subsets of the set $\mathcal{Q}$ of EMMs involve optimization over probability measures. For more on duality in indifference pricing, see Frittelli (2000a, §3).

For an exposition of portfolio optimization in incomplete markets in terms of convex duality, including equivalent martingale measures and marginal indifference pricing, see Schachermayer (2002). Convex duality also appears in representation and optimization of risk measures (Ruszczyński and Shapiro, 2004). The conditions for the price bounds (2) and (3) to avoid arbitrage are best understood in terms of duality: for a version of the first fundamental theorem of asset pricing, see Staum (2004).

Under some conditions, including that the acceptance set A be related to a coherent risk measure, the price bounds (2), (3) have the dual representation

$$\left(\inf_{\mathbf{Q}\in\mathcal{D}} \mathrm{E}_{\mathbf{Q}}[DX],\ \sup_{\mathbf{Q}\in\mathcal{D}} \mathrm{E}_{\mathbf{Q}}[DX]\right), \tag{9}$$

where $\mathcal{D}$ is a subset of the set $\mathcal{Q}$ of EMMs (Jaschke and Küchler, 2001). For the no-arbitrage bounds, $\mathcal{D} = \mathcal{Q}$. When the price bounds coincide and are linear, $\mathcal{D}$ is a singleton, i.e. the method selects a single EMM (see Section 2.3 for a discussion). Marginal indifference pricing and minimum-distance measures are the principal methods of selecting a single EMM.

4.3 Marginal pricing

For any price bounds b and a, $\lim_{\gamma\downarrow 0} \gamma a(X/\gamma)$ and $\lim_{\gamma\downarrow 0} \gamma b(X/\gamma)$ may coincide and provide a unique price $\tilde{p}(X)$ suitable for small trades. For a general result on good deal bounds, see Staum (2004, Prop. 5.2). Under expected utility preferences, $\tilde{P}(W) = \mathrm{E}[u(W)]$, this suggestion corresponds to using the marginal utility u' to define a pricing measure $\mathbf{Q}$:

$$\tilde{p}(X) = \mathrm{E}_{\mathbf{P}}\big[u'(V)DX\big] = \mathrm{E}_{\mathbf{Q}}[DX], \tag{10}$$

where D is the discount factor and $\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P} = u'(V)/\mathrm{E}[u'(V)]$. That is, in the most straightforward case, marginal indifference pricing results in the selection of a single EMM $\mathbf{Q}$ whose likelihood ratio with respect to the statistical probability measure $\mathbf{P}$ is proportional to the marginal utility of terminal wealth provided by an optimal portfolio.

Marginal indifference pricing is based on the idea that a single trade is small and does not need to be hedged. This argument is appropriate for finding the equilibrium price of a security that is traded and infinitely divisible, but see Section 2.1. If a small trade has negligible impact on the whole portfolio's risk profile, e.g. it has little effect on marginal utility, that is an argument for using the unique marginal indifference price. This argument is not generally appropriate for OTC market-making. A single small trade might seem to be priced adequately by marginal indifference, but many small trades cumulatively can involve large risks. Ignoring the likely cumulation of risks can cause initial, myopic underpricing of OTC securities that are in high demand, followed by a concentration of related risks and thus the need to set high prices, at which fewer trades would be made (see Section 4.1.1). The contribution of a small trade to total risk depends on the opportunities for hedging, which should therefore affect pricing.

4.4 Minimum-distance pricing measures

Marginal indifference prices based on expected utility are an example of pricing with a minimum-distance measure. The expected utility is an expectation under a statistical probability measure $\mathbf{P}$. The marginal indifference price is an expected discounted payoff under a *minimax martingale measure* $\mathbf{Q} \in \mathcal{Q}$ that is "closest" to $\mathbf{P}$ in the sense of providing the least possible expected utility to an investor who could buy any payoff V for $\mathrm{E}_{\mathbf{Q}}[DV]$. That is, $\mathbf{Q}$ corresponds to the "least favorable market completion": in the fictitious complete market in which the price of any payoff V is $\mathrm{E}_{\mathbf{Q}}[DV]$, the utility derived from optimal investment is as low as possible (Skiadas, 2006). The minimax martingale measure $\mathbf{Q}$ is the solution to

$$\min_{\mathbf{Q} \in \mathcal{Q}} \max_{V} \bigl\{ \mathrm{E}_{\mathbf{P}}\bigl[u(V)\bigr] \mid \mathrm{E}_{\mathbf{Q}}[DV] \leqslant c \bigr\}. \tag{11}$$

For a version based on local utility, see Kallsen (2002b) and references therein. Particular choices of utility yield quadratic and exponential methods in Sections 6–7; the latter distance can also be described in terms of relative entropy. The same concepts appear in Section 9, featuring not just the minimum-distance measure but a set of EMMs having low distance to $\mathbf{P}$. For more on portfolio optimization and minimum-distance measures, see Goll and Rüschendorf (2001).

Somewhat different is the case of calibration (Section 11), which is not based on a statistical probability measure $\mathbf{P}$. Instead it starts from a parametric family $\mathcal{P}$, and selects the pricing measure $\hat{\mathbf{Q}} \in \mathcal{P}$ that is closest to $\mathcal{Q}$ in the sense of having the least error in replicating the prices of marketed derivative securities; an EMM in $\mathcal{Q}$ would yield zero replication error.

Figure 1 illustrates the structure of four schemes for selecting a probability measure for pricing in an incomplete market. It uses the very simple setting of a one-period model with three states and two marketed securities: a riskless bond paying \$1 in all states and with initial price of \$1, and a stock worth \$2 in state 1, \$1 in state 2, and \$0 in state 3 and having initial price \$0.80. To simplify matters even further for purposes of two-dimensional representation, we will assume that the bond must be repriced exactly, so the price assigned to a payoff X is $\mathrm{E}_{\mathbf{Q}}[DX] = \mathrm{E}_{\mathbf{Q}}[X]$ where the pricing measure $\mathbf{Q} = (q_1, q_2, q_3)$ is a true probability measure, such that the probabilities of the three states sum to one: $q_1 + q_2 + q_3 = 1$. Thus, $q_3 = 1 - (q_1 + q_2)$, so all possible pricing measures can be parametrized by the triangle in Fig. 1: $q_1 \geqslant 0$, $q_2 \geqslant 0$, $q_1 + q_2 \leqslant 1$. The diagonal line $2q_1 + q_2 = 0.8$ represents the constraint of repricing the stock, so its line segment in the interior of the triangle is the set $\mathcal{Q}$ of EMMs for any statistical probability measure $\mathbf{P}$ that assigns positive probability to all states. The vertical line segment inside the triangle and defined by $q_1 = 0.5$ represents a set $\mathcal{P}$ of models. Of course, this example is so simple that there is no need to restrict attention to a subset of the possible pricing measures that does not include any measures that reprice the stock; also, ordinarily models include underlying securities' initial prices as parameters, so all underlying securities, as opposed to derivative securities, are repriced exactly. The point of the setup in Fig. 1 is that the resulting structure is not only very simple, but also similar to that encountered in practice, in which one works with a parametric family of models that does not include an EMM.

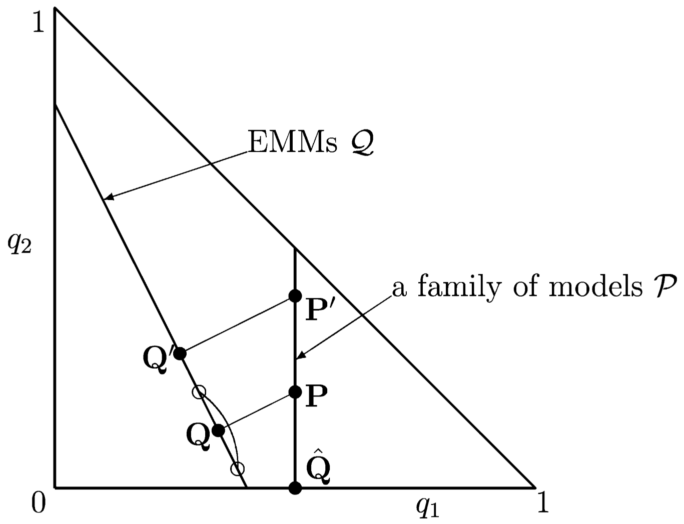

Fig. 1. Structures of schemes for selecting a pricing measure.

Calibrating the family of models $\mathcal{P}$ to the stock price selects $\hat{\mathbf{Q}} = (0.5, 0, 0.5)$ as the pricing measure, which minimizes the error in repricing the stock by assigning it the price \$1, the least possible within this family of models. Another scheme begins with a statistical probability measure $\mathbf{P}$, which may have been estimated within the family $\mathcal{P}$ by econometric inference, and then selects the EMM $\mathbf{Q}$ that is closest to $\mathbf{P}$. In Figure 1, $\mathbf{Q} = (0.34, 0.12, 0.44)$ minimizes Euclidean distance, but several distances have been proposed, relating e.g. to entropy or expected utility. Instead of selecting only $\mathbf{Q}$, which minimizes the distance to $\mathbf{P}$, to get a unique price, one may select a set of pricing measures having low distance to $\mathbf{P}$, and get an interval of prices: the empty dots connected by a curved arc around $\mathbf{Q}$ represent the extreme measures selected by this scheme. Where distance is a function of $\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P}$, this scheme includes some approaches based on pricing kernel restrictions (Section 9). The fourth scheme begins with multiple probability measures, here $\mathbf{P}$ and $\mathbf{P}'$, yielding robustness to ambiguity about the statistical probability measure (Section 10). Each of these yields a minimum-distance EMM, here $\mathbf{Q}$ and $\mathbf{Q}'$ respectively, and this resulting set of EMMs can be used to generate a unique price or a price interval.

5 Issues in pricing and expected utility examples

Our main example is adapted from Carr et al. (2001).

Example 5.1. Consider a single-period economy with five possible states and three assets: a riskless bond, a stock, and a straddle. The bond and stock are

Table 2.
Terminal Asset Values.

	State 1	State 2	State 3	State 4	State 5
Bond	$1	$1	$1	$1	$1
Stock	$80	$90	$100	$110	$120
Straddle	$20	$10	$0	$10	$20

Table 3.
Expected Utility Indifference Pricing of a Straddle.

Initial wealth	Initial portfolio		Transaction type	Portfolio adjustment		Indifference price
	Bond	Stock		Bond	Stock	
$100	77.6	0.334	buy	−25.1	0.147	$9.92
			sell	21.3	−0.082	$12.11
$1000	776	3.34	buy	−22.8	0.111	$10.94
			sell	22.5	−0.105	$11.17

marketed, with initial prices $0.9091 and $88.1899 respectively. The terminal values of the three assets are given in Table 2. The no-arbitrage bounds for the straddle price are $2.72 and $18.18. Consider the utility function $u(W) = -(W/100)^{-4}$ for $W > 0$, and suppose the states have equal probabilities.

For any level of initial wealth, the optimal portfolio in marketed securities has 70.55% of the wealth in the bond and 29.45% in the stock.[2] Pricing by marginal utility uses the probabilities $\mathbf{Q} = (26.55\%, 22.68\%, 19.46\%, 16.78\%, 14.53\%)$, yielding a unique price of $11.06 for the straddle. The bid and ask indifference prices when the initial wealth is $100 or $1000 allocated optimally are ($9.92, $12.11) and ($10.94, $11.17) respectively. Table 3 shows the corresponding portfolio adjustments providing optimal payoff Y^*.

5.1 Dependence on trading opportunities

The opportunities to trade in the market affect the indifference price. For example, suppose that the stock were not marketed, but the initial portfolio still had 29.45% of its wealth in the stock. Then the indifference prices based on initial wealth of $100 would be ($9.86, $12.14): with fewer opportunities to rebalance the portfolio, the price interval would become wider. The marginal

[2] Power and log utilities, having constant relative risk aversion, can lead to optimal portfolios whose allocation fractions do not depend on the initial wealth: see e.g. Karatzas and Shreve (1998, Examples 3.6.6–7).

indifference price given by Eq. (10) would *not* change: it involves an infinitesimal change in the portfolio, in the direction defined by the straddle payoff, and no portfolio rebalancing. On the other hand, there is a difference in the marginal prices derived from the indifference price (8) incorporating portfolio optimization, depending on whether the optimal portfolio is allowed to contain stocks and bonds, or only bonds. In the latter case, the optimal portfolio provides the same wealth in each state, so the marginal indifference price is \$12, based on $\mathbf{Q} = (20\%, 20\%, 20\%, 20\%, 20\%)$.

5.2 Dependence on current portfolio

The indifference prices and the optimal portfolio adjustments also depend on the random wealth V provided by the initial portfolio. This is intuitively reasonable, as a trader should be less eager to acquire a payoff that exacerbates unhedgeable risk in the current portfolio than one that cancels out such risks. As Rouge and El Karoui (2000) say, "it is unrealistic that agents with different endowments should have the same attitude toward risk." Indeed, OTC market makers describe an unhedgeable risk in their portfolios as an "axe," thinking of the expression "having an axe to grind." For example, suppose a trader is long OTC options on a stock with no marketed options. It would not be easy to hedge the risk of a decline in the implied volatility (and hence value) of these options, so this long position is an "axe" which the trader would like to "grind" by selling OTC options. The trader would set low ask and bid prices, to encourage sales of options, which decrease this risk, and get adequate compensation for purchases, which increase this risk.

Table 4 illustrates this point for Example 5.1. It shows that after a trader has bought a straddle and re-optimized the portfolio, as in Table 3, the bid and ask prices decrease. The new ask price is the same as the original bid price, which makes sense: together, the two transactions return the trader to the original portfolio, so a net cost of zero produces indifference. The marginal indifference price after buying a straddle and optimally rebalancing is \$8.77 for the case of \$100 initial wealth; this too decreases because the change in the portfolio has reduced marginal utility in most of the states in which the straddle pays off.

Table 4.
Effect of Initial Portfolio on Expected Utility Indifference Pricing.

Initial wealth	Initial portfolio		Transaction type	Hedge portfolio		Indifference price
	Bond	Stock		Bond	Stock	
\$100	52.5	0.481	buy	–33.5	0.258	\$7.66
	plus 1 straddle		sell	25.1	–0.147	\$9.92
\$1000	753	3.45	buy	–23.2	0.118	\$10.72
	plus 1 straddle		sell	22.8	–0.111	\$10.94

The dependence of indifference price on initial portfolio, illustrated in Table 4 for constant relative risk aversion, occurs even with constant absolute risk aversion (exponential utility, Section 7), which is often used to obtain separation of investment and hedging decisions. In Example 5.1, if the utility function is replaced by $u(W) = -\exp(-(0.0453)W)$, the optimal allocation of \$100 initial wealth remains very nearly the same, leading to a similar bid–ask spread of (\$9.89, \$12.10). After the trader has bought a straddle and re-optimized the portfolio, the bid–ask spread becomes (\$7.64, \$9.89).

5.2.1 Optimality as prerequisite for indifference pricing

Indifference prices should fall within the no-arbitrage bounds, so as to avoid arbitrage in OTC trades. To prevent the indifference price in Eq. (7) from violating no-arbitrage bounds, the initial portfolio V must be optimal, i.e. $V = V^* = W + Y^*$ where Y^* solves the portfolio optimization problem (6). If V is suboptimal, the indifference price may exceed the market price for a replicable payoff that increases the preference index $\tilde{P}$ in Problem (6). Likewise, the indifference price for a non-replicable payoff may exceed its upper no-arbitrage bound. Example 5.2 illustrates these effects.

Example 5.2. Continuing Example 5.3, suppose the initial portfolio delivers \$100 except in state 1, in which it delivers only \$60. Consider the payoff Y provided by a portfolio long 100 shares of the bond and short one share of the stock.

Based on marginal utility, the valuation probability of state 1 is 76.28%, giving a marginal indifference price for the put of \$6.93. Its indifference price is \$8.42. These both exceed the upper no-arbitrage bound of \$5.23. Acquiring Y increases expected utility. Its market price is \$2.72, but its marginal indifference price is \$9.49 and its indifference price is \$7.28.

The need to base indifference pricing on an optimal portfolio causes a grave difficulty in using expected utility. Because expected utility maximization is not robust to ambiguity about the statistical probability measure **P** (Section 4.2.3), it is not actually a good idea to adopt the supposedly optimal portfolio. Typically, the true expectation of V^* is lower than $E_{\mathbf{P}}[V^*]$, because the optimal portfolio overinvests in assets that are wrongly believed to have high expected returns. Consequently, the trader does not optimize his portfolio, $V \neq V^*$, but to avoid arbitrage, the indifference prices must be based on V^*. The result is that the trader is not indifferent between trading and not trading at these "indifference prices"; someone else with a different portfolio would be. The economic justification for expected utility indifference pricing evaporates.

5.3 Risk vs. preference

It is tempting to think of the optimal portfolio adjustment Y^* in Problem (7) as a hedge for the payoff X, but as we have seen, Y^* and the indifference price

$b(X)$ depend on the payoff V from the existing portfolio strategy as well as on the payoff X. Only in special cases such as neutralization of Greeks does hedging apply to payoffs without reference to a portfolio: delta-hedging each security in a portfolio produces the same net position as delta-hedging the whole portfolio. The hedge that minimizes a portfolio's risk does not generally coincide with the sum of such hedges for each security in the portfolio.

Another way in which the optimal Y^* in Problem (2) or (7) is not a hedge is that it need not reduce risk. To formulate a less risky alternative, suppose that market prices π are linear and there is a reference security (e.g. riskless bond) with payoff denoted $\mathbf{1}$. One can finance the purchase of a payoff X for $b(X)$ by acquiring Y^* or, more simply, by acquiring $-(b(X)/\pi(\mathbf{1}))\mathbf{1}$. By definition, it is *preferable* to acquire Y^*: $P(V+X+Y^*) \geqslant P(V+X-(b(X)/\pi(\mathbf{1}))\mathbf{1}$, but it need not be less *risky* to acquire Y^*. Unless the preference P and risk measure ρ are related as $P = -\rho$, it is possible that $\rho(V+X+Y^*) > \rho(V+X-(b(X)/\pi(\mathbf{1}))\mathbf{1}$. The following example illustrates this point.

Example 5.3. We extend Example 5.1 by including another non-traded asset, a put option on the stock with strike \$90. Its only nonzero payoff is \$10 in state 1, in which the stock is worth \$80. As a risk measure of a payoff W, we use the tail conditional expectation (see Artzner et al., 1999) of $W - 110$, the shortfall relative to investing \$100 in bonds.

The no-arbitrage bounds on the put's price are (\$0, \$5.23) and the bid and ask indifference prices when the initial wealth is \$100 allocated optimally are (\$2.22, \$2.60), with marginal indifference price \$2.41. Table 5 shows the state-by-state values of the original optimized portfolio V, of the portfolio $V + X - (b(X)/\pi(\mathbf{1}))\mathbf{1}$ after buying the put for \$2.22 by selling bonds, and of the re-optimized portfolio $V+X+Y^*$. Table 6 shows these portfolio's tail conditional expectations at several probability levels, corresponding to average values over the worst 1–5 states.

The last column, tail conditional expectation at the 100% level, is E[$110 - W$], which simply measures the portfolio's expected value. The other columns are more properly risk measurements. They each show that buying the put by selling bonds reduces risk, while re-optimizing the portfolio increases risk even beyond its original levels. Because the put adds extra wealth in state 1, the worst state for the original portfolio, it allows the re-optimized portfolio

Table 5.
Portfolio Values when Buying a Put.

Portfolio	State 1	State 2	State 3	State 4	State 5
Original Optimal	\$104.32	\$107.66	\$111.00	\$114.34	\$117.68
Buy Put, Sell Bonds	\$111.88	\$105.22	\$108.55	\$111.89	\$115.23
Re-optimized	\$107.97	\$103.61	\$109.24	\$114.88	\$120.51

Table 6.
Risk in Buying a Put.

Portfolio	Tail Conditional Expectation				
	20%	40%	60%	80%	100%
Original Optimal	5.68	4.01	2.34	0.67	−1.00
Buy Put, Sell Bonds	4.78	3.11	1.45	0.61	−0.56
Re-optimized	6.39	4.21	3.06	1.07	−1.24

to allocate a greater fraction of wealth to the stock. This maximizes expected utility, but it increases risk: for instance, the re-optimized portfolio has less wealth ($103.61) in its worst state than does the original portfolio ($104.32) in its worst state.

Even if preferences are risk-averse, preference and risk are not simply opposites, as the example shows, even though it is always preferable and less risky to have more wealth. To incorporate risk management concerns, one may add a risk constraint. We could reformulate the trader's portfolio optimization problem (6) as

$$\sup_{Y \in R} \{ \tilde{P}(W + Y) \mid \pi(Y) \leqslant c, \rho(W + Y) \leqslant r \}, \tag{12}$$

where internal or external regulators impose the risk measure ρ and the limit r on the risk of the trader's portfolio. Given this formulation, one might think of the solution to Problem (6) as an optimal portfolio adjustment and of the difference between the solutions to Problems (6) and (12) as a hedge. "Hedging" is a good description of neutralizing Greeks, which is solely risk minimization, with no other preference involved; when optimizing with preferences distinct from risk, portfolio re-optimization need not be hedging i.e. risk reduction.

6 Quadratics

Quadratic hedging is a much-studied, mathematically elegant approach to incomplete markets. Surveys include Pham (2000) and Schweizer (2001). The quadratic method is a special case of expected utility indifference pricing, with quadratic utility $u(x) = -x^2$. Because it is decreasing for $x > 0$, quadratic utility is not a realistic model of preferences, as has often been pointed out, e.g. by Dybvig (1992). Quadratic utility penalizes the gain due to a hedge's excess over the liability to be covered, as well as the loss due to shortfall with respect to the liability. The same charge has been leveled against mean–variance portfolio analysis. Markowitz (2002, pp. 155–156) responds:

> ... the problem was to reconcile the use of single-period mean–variance analysis by (or on behalf of) an investor who should maximize a many-period utility function. My answer lay in the observation that for many utility functions and for probability distributions of portfolio returns "like" those observed in fact,

> one can closely approximate expected value of the (Bellman 1957 "derived") utility function knowing only the mean and variance of the distribution.

For details, see the references Markowitz (2002) cites after this quote. It would be interesting to investigate how well the mean and variance can approximate the derived utility of hedged portfolios resulting from OTC market-making.

One might try to separate the problems of hedging, to be solved with a quadratic approach for tractability, and optimal investment, to be solved with an appropriate utility function. However, Dybvig (1992) provides a negative result for the case where incompleteness is due to nonmarket risks: this separation does not occur except with constant absolute risk aversion (exponential utility) and independence of the hedging residual and the marketed risks.

Föllmer and Schweizer (1991) developed a martingale decomposition theorem that yields a *locally risk-minimizing* hedging strategy for a payoff X, where risk is instantaneous or one-step variance. The solution relates to the *minimal martingale measure* $\hat{P}$. For senses in which $\hat{P}$ is minimal, relating both to quadratic and entropy criteria, see Schweizer (1999). In local risk minimization, it is standard to optimize over hedging strategies that need not be self-financing. A non-self-financing portfolio strategy has an associated cost process C, where $C(t)$ is the cumulative cash influx required to rebalance the portfolio over the time interval $[0, t]$. At each instant t, a locally risk-minimizing strategy minimizes $\mathrm{E}[(C(T) - C(t))^2|\mathcal{F}_t]$, the conditional expectation of the squared cumulative future costs, without regard to past costs. A locally risk-minimizing strategy is "mean-self-financing" in the sense that its cost process is a martingale (Schweizer, 2001, Lem. 2.3), so $C(t) = \mathrm{E}[C(T)|\mathcal{F}_t]$, and thus local risk minimization is equivalent to minimizing the conditional variance of the cumulative cost. In discrete time, a backward recursion shows that this is equivalent to choosing the portfolio weights at time t_i to minimize $\mathrm{Var}[(C(t_{i+1}) - C(t_i))^2|\mathcal{F}_{t_i}]$, the conditional variance of the cost incurred at time t_{i+1}. This method is local in the sense that it involves one-step optimizations, and in the sense that an infinitesimal perturbation of the locally risk-minimizing strategy must increase the variance of the cost over the next step or instant. The optimal cost process is orthogonal to the gains process of the locally risk-minimizing strategy, which is a projection of the $\hat{P}$-conditional expectation process of X (Pham, 2000, Thm. 4.2).

The *mean–variance optimal* self-financing hedging strategy minimizes $\mathrm{E}[(Y - X)^2]$, the variance of the hedging residual. This global quadratic criterion relates to the *variance-optimal martingale measure* $\tilde{P}$ (Schweizer, 1996), which is a minimum-distance measure (Section 4.4) based on L^2-distance. Bertsimas et al. (2001) provide a stochastic dynamic programming algorithm for computing the mean–variance optimal hedging strategy. This hedging problem can be studied by means of martingale measures or backward stochastic differential equations: for recent work on the latter, see Lim (2004) and references therein.

Heath et al. (2001) provide a theoretical and numerical comparison of the local and global quadratic approaches. The following example illustrates the difference between a local and global approach in the quadratic setting.

Example 6.1. Continuing Example 4.1, suppose that a trader wishes to hedge the sale of a contingent claim paying \$1 in state 1.

The locally risk-minimizing hedge is $\xi_1 = 0.1$, $\xi_2^{(0)} = \xi_2^{(-)} = 0$, and $\xi_2^{(+)} = 0.2$. The variance-optimal hedge is $\xi_1 = \xi_2^{(0)} = \xi_2^{(-)} = 0$ and $\xi_2^{(+)} = 0.33$. The cost processes associated with these hedges are given in Table 7. The total cost is the hedging residual. Its variance is minimized by the variance-optimal hedge, yielding a variance of 0.037, as opposed to 0.047 for the locally risk-minimizing hedge, which does not take into account the partial cancellation of costs incurred at different times in state 2. The conditional variances at time 1 of the cost incurred at time 2 are 0 when the risky asset's price is 0 or -1, under either hedging scheme, and 0.133 or 0.222 for the locally risk-minimizing and variance-optimal hedges respectively, when the risky asset's price is 1. The unconditional variance of the cost incurred at time 1 is 0.007 or 0.037 for the locally risk-minimizing and variance-optimal hedges respectively.

Suppose that the set R of replicable payoffs is a linear space. The quadratic criteria behave linearly in the sense that, if the hedge Y is optimal for a payoff X, then for any multiple $\gamma \in \mathbb{R}$, γY is optimal for γX. Consequently, the quadratic methods result in unique prices and select a single martingale measure $\hat{P}$ or $\tilde{P}$.

However, it is not appropriate to interpret an expected discounted payoff under $\hat{P}$ or $\tilde{P}$ as a price. As suggested earlier, because quadratic utility does not model preferences well, these prices may not be compatible with the trader's preferences (Bertsimas et al., 2001). Moreover, they may violate the no-arbitrage bounds. The measures $\hat{P}$ and $\tilde{P}$ may be *signed*, that is, they may assign negative values to some events. Pricing under a signed measure

Table 7.
Quadratic Hedging Cost Processes.

State	Probability	Locally risk-minimizing			Variance-optimal		
		Time 1	Time 2	Total	Time 1	Time 2	Total
1	1/9	\$0.1	\$0.4	\$0.5	\$0.33	\$0	\$0.33
2	1/6	\$0.1	−\$0.4	−\$0.3	\$0.33	−\$0.67	−\$0.33
3	1/18	\$0.1	\$0	\$0.1	\$0.33	\$0	\$0.33
4	1/6	\$0	\$0	\$0	\$0	\$0	\$0
5	1/6	\$0	\$0	\$0	\$0	\$0	\$0
6	1/6	\$0.1	\$0	\$0.1	\$0	\$0	\$0
7	1/6	\$0.1	\$0	\$0.1	\$0	\$0	\$0

would imply willingness to pay to give away a lottery ticket, i.e. Arrow–Debreu security, for such an event (Schweizer, 1995). The reason this happens is precisely that quadratic utility penalizes gains as well as losses, so its marginal utility may be negative. For examples of arbitrage resulting from quadratic pricing, see Schweizer (1995, §5) or Frittelli (2000b, p. 50). For similar reasons, jump processes in continuous time pose difficulties for the quadratic approach: there may be negative marginal utility for wealth in a state in which a jump in marketed asset prices causes the optimal portfolio's value to exceed the liability X. An example of what can go wrong occurs in Example 6.1, where $\hat{P}(\omega_1) = \tilde{P}(\omega_1) = 0$, so the optimal initial capital for local or global quadratic hedging of the Arrow–Debreu security for state 1 is zero.

According to Biagini and Pratelli (1999), in discrete time or with jumps, the results of local risk-minimization depend on the numéraire: the hedging strategy depends on whether the costs of the portfolio, which is not self-financing, are measured in units of cash, bonds, stocks, etc. One response to this is that the trader should simply choose the numéraire such that the variance of costs as measured in this numéraire best describes his preferences. However, this observation draws attention to a theoretical shortcoming of using strategies that are not self-financing: costs which are cashflows at different times are simply added, ignoring the time value of money. This may not be a significant issue unless long time spans or high interest rates are involved.

7 Entropy and exponential utility

Another special case of expected utility indifference pricing uses *exponential utility*, also known as *negative exponential utility*, which may be conveniently expressed as $u(x) = 1 - \exp(-\alpha x)$. It has the feature of constant absolute risk aversion, which can produce theoretically elegant results, such as separation of hedging and investment decisions, and independence of the indifference price in Eq. (8) of the initial budget c. Also interesting is the relationship between maximization of exponential utility and minimization of relative entropy $\mathrm{E}_{\mathbf{Q}}[\ln(\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P})]$. The marginal exponential utility indifference price is the expected discounted payoff under a minimum-distance measure (Section 4.4), the *minimal entropy martingale measure* (MEMM) having minimal relative entropy with respect to the statistical probability measure $\mathbf{P}$ (Frittelli, 2000b; Rouge and El Karoui, 2000). Relative entropy also appears in Section 10 as a way of quantifying ambiguity.

Delbaen et al. (2002) cover the topic of exponential utility maximization and valuation via the MEMM with special attention to the set of feasible portfolio strategies over which the optimization occurs. Becherer (2003) gives a general presentation and more explicit results in a special case in which the financial market is complete, but one must value payoffs that depend also on risks independent of the financial market. Mania et al. (2003) discuss special cases in which the MEMM can be constructed explicitly. Another explicit example, with intuition, and an algorithm for indifference pricing in a similar setting are in Musiela and Zariphopoulou (2004a, 2004b). Fujiwara and Miyahara (2003) discuss representation of the MEMM in terms of Esscher transforms when the underlying process is a geometric Lévy process, giving as examples Brownian motion plus a compound Poisson process, a stable process, and the variance gamma process.

Under some conditions, including restrictions on the form of the mean–variance ratio, the minimal martingale measure coincides with the MEMM (Mania et al., 2003, Prop. 3.2). The minimal martingale measure $\hat{P}$ (see Section 6) is the solution to the dual of the problem of maximizing exponential utility given an initial endowment equal to a multiple of the mean–variance ratio (Delbaen et al., 2002, Thm. 5.1). An alternative is to minimize the entropy-Hellinger process instead of relative entropy. Choulli and Stricker (2005) develop this approach and show that it corresponds to the neutral derivative prices of Kallsen (2002a), for which see Section 9.1, and that it selects the minimal martingale measure (Section 6) when the discounted price process is continuous. Choulli et al. (2006) provide an extension of this approach and a more general framework including it and other minimum-distance measures.

8 Loss, quantiles, and prediction

What unifies the ideas covered in this section is an emphasis on the *loss* or *shortfall* $(Y - X)^-$ associated with hedging the sale of the payoff X by acquiring the payoff Y. They ignore the positive part of the hedging residual, $(Y - X)^+$. Unfortunately, the nomenclature surrounding these methods is a bit confusing: they may also involve a *loss function* ℓ, which is another way of expressing utility: $\ell(x) = -u(-x)$. Minimizing the expected loss is then the same as maximizing expected utility, so pricing via expected loss minimization could be understood as a special case of expected utility indifference pricing. That is, the trader would be seeking the cheapest hedge Y such that $\mathrm{E}[\ell((V + Y - X - B)^-)] \leqslant \mathrm{E}[\ell((V - B)^-)]$, where V is the endowment and B is a benchmark relative to which losses are measured, possibly zero. (If $V = B = 0$, the resulting indifference prices are the no-arbitrage bounds, because gains are ignored and cannot make up for losses.) However, this is not the way that loss minimization has usually been treated.

This literature primarily addresses the problem of minimizing expected loss given a fixed initial budget with which to hedge a liability, without reference to an endowment payoff V. The focus is solely on the shortfall of an approximate hedge. This literature also addresses the very closely related problem of determining the minimal required initial budget to hedge so that expected loss does not exceed some prespecified threshold.

We will consider how the latter problem may apply to pricing in incomplete markets. The approach falls into the framework described in Section 4.2, with the acceptance set $A = \{Z \mid \mathrm{E}[\ell(Z^-)] \leqslant p\}$. Because the loss function ignores gains, if A is nontrivial, it must include a payoff $Z < 0$. Therefore, pricing a payoff X as the minimal initial budget required to hedge so that expected loss is less than p can result in giving the counterparty an arbitrage. For example, the minimal cost of a replicable payoff Y subject to the constraint $\mathrm{E}[\ell(Y^-)] \leqslant p$ for $p > 0$ may be negative. This method is not generally sound for OTC pricing, as illustrated in Example 8.1.

The two following subsections describe two particular choices of loss function, whose original application was for hedging given a capital constraint, and show that this expected loss methodology should not be transposed directly to the application of OTC pricing.

8.1 *Expected shortfall*

The choice $\ell(x) = x$ is minimization of expected shortfall. For theoretical results, see Cvitanić (2000), who discusses the form of the optimal hedge in a market that is incomplete due to stochastic volatility or trading constraints.

Example 8.1. Continuing Example 5.1, consider the minimal initial capital required to hedge the straddle given a constraint on expected shortfall.

Figure 2 shows this initial capital as a function of the level p of the constraint. The lower curve in Fig. 2 has negative values for large p because the optimal "hedge" has a negative value in some states, and its cost is negative. The upper curve gives the initial capital required to attain the expected shortfall constraint given the additional constraint that the hedge must be nonnegative; this constraint is appropriate only when treating nonnegative payoffs X such as the straddle. The result is that for $p = \$12$, which is the expected shortfall of the unhedged straddle, the required initial capital is \$0. This is still below

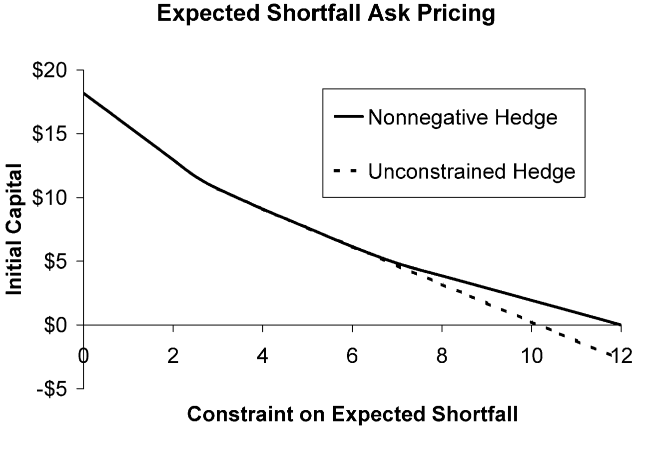

Fig. 2. Cost of hedging a straddle to achieve an expected shortfall constraint.

the lower no-arbitrage bound, which is \$2.72. For $p = 0$, the initial capital equals the no-arbitrage upper bound. The initial capital required for hedging a replicable payoff does not, in general, equal its market price. For example, the no-arbitrage price is \$0 for an equity swap replicated by a portfolio that is long 1 share of stock and short approximately 97 shares of the bond. However, with an expected shortfall constraint of $p = 0.25$, the price assigned is −\$0.26.

8.2 *Quantile hedging*

Another special case, $\ell(x) = \mathbf{1}\{x > 0\}$, is known as *quantile hedging* (Föllmer and Leukert, 1999). With this loss function, the hedger tries to minimize the probability of a positive shortfall, without regard to the magnitude of shortfall. Alternatively, one might try to apply quantile hedging to pricing by finding the minimal initial budget required to hedge so that the probability of a positive shortfall does not exceed p. The special case $p = 0$ results in superreplication, that is, any feasible hedge Y satisfies $Y \geqslant X$, corresponding to the no-arbitrage upper bound.

However, for $p > 0$, the method may not work: if there is an event F such that $\mathbf{P}[F] \leqslant p$ and a replicable payoff Y_F of negative price $\pi(Y_F)$ such that $Y_F \mathbf{1}_F \geqslant 0$, the optimization $\min_{Y \in R}\{\pi(Y) \mid \mathbf{P}[Y - X < 0] \leqslant p\}$ tends to be unbounded. For example, if the space of replicable payoffs R and market prices π are linear, and $Y^* \geqslant X$ is a superreplicating payoff, then $Y^* + \lambda Y_F$ is feasible for all $\lambda \in \mathbb{R}$, so the optimization is unbounded. Even if portfolio constraints and nonlinear market prices render the optimization bounded, the results are still likely to be unusable. The optimal solution will tend to involve, for any X, a large negative price to be paid to the buyer of X, funded by incurring large liabilities on some event F of sufficiently small probability. The more complete the market is, the worse this problem will be, as it becomes easier to concentrate liabilities on events of low probability but high state price. One way to ameliorate this problem is to restrict the hedge to be nonnegative (Föllmer and Leukert, 1999). This still leaves the methodology with the same deficiencies as for expected shortfall.

8.3 *Statistical prediction intervals*

Related to quantile hedging is a statistical approach based on prediction intervals for financial quantities such as cumulative interest rates and volatility over an option's life (Mykland, 2003a, 2003b). Quantile hedging looks for a hedge Y that covers the liability X on some event F_X of probability p, i.e. $\mathbf{1}_{F_X}(Y - X) \geqslant 0$ and $\mathbf{P}[F_X] = p$. By contrast, this statistical approach specifies a fixed event G, a prediction interval of probability p used for all payoffs X, and requires that Y satisfy $\mathbf{1}_G(Y - X) \geqslant 0$. This makes the bounds wider for this prediction interval approach than for quantile hedging at the same error level p, assuming the same statistical probability measure $\mathbf{P}$ in both cases. An advantage of the prediction interval approach is that it need not be based on a single probability measure **P**. Without assuming a specific model for stochastic volatility and interest rates, much less that its parameters are known, Mykland (2003a, 2003b) works out bounds for European option prices and the related hedging strategies in a diffusion setting, given prediction intervals for cumulative volatility $\int_0^T \sigma^2(t)\,\mathrm{d}t$, or for this and cumulative interest rates $\int_0^T r(t)\,\mathrm{d}t$ together.

Because it is similar to quantile hedging, this prediction interval approach has a similar drawback as a method for OTC pricing: it assigns zero value to payoffs that are zero inside the prediction interval but positive outside it, which allows arbitrage. The prediction interval approach may be most useful in risk management, for reducing model risk (Mykland, 2003b, §1) or in formulating a liquidation strategy for a trade (Mykland, 2003a, §6).

9 Pricing kernel restrictions

One way of expressing the problem of pricing in incomplete markets is that total ignorance about the pricing kernel Π allows the price to be anywhere within the no-arbitrage price bounds (Section 2.3). This suggests that one may apply a restriction to the pricing kernel to get price bounds. The main idea is that some of the pricing kernels that are possible, in the sense of repricing all marketed securities, are economically implausible. One basis for this is to assert that some pricing kernels make some of the replicable payoffs into objective good deals (Section 4.2.4), and it is implausible that such good deals should exist. That is, one may exclude pricing kernels that would result in too good a deal for a typical investor or most investors.

An early approach, not related to good deals, is to impose restrictions on the moments of asset prices under the pricing measure, rather than directly on the moments of the pricing kernel. If the statistical probability measure **P** is known, this restriction on the pricing measure **Q** is equivalent to a restriction on the pricing kernel $\Pi = D\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P}$. Lo (1987) applied restrictions on a stock's variance under **Q** to pricing an option on that stock. Further research has incorporated restrictions on higher moments and developed computational algorithms. It may seem advantageous that the bounds derived from **Q**-moment restrictions do not depend on the statistical probability measure **P**, and thus do not require a choice of statistical model – but from where does knowledge of **Q**-moments come? Lo (1987) showed that, for two simple models, the **Q**-variance can be computed from the **P**-variance under the statistical measure, and that method-of-moments estimation yields the same result for the two models. However, in general, the need to connect **Q**-variance to estimable quantities can introduce dependence on a model.

9.1 Low-distance pricing measures: pricing kernels and good deals

Pricing by minimum-distance measure (Section 4.4) selects the single pricing measure **Q** with the lowest distance from the statistical probability measure **P**, or equivalently, the corresponding pricing kernel or likelihood ratio $\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P}$ representing the lowest distance. For example, the relative entropy distance (Section 7) is a function of the likelihood ratio. A modification of this method is to select a set of pricing measures $\{\mathbf{Q} \mid d(\mathbf{P}, \mathbf{Q}) < \epsilon\}$ with low distance d from **P**. The distance constraint is equivalent to a restriction on the pricing kernel. It may be more convenient to consider restrictions directly in terms of the pricing kernel.

One approach is to place restrictions on the moments of the pricing kernel, which can be translated into restrictions on the assets' returns. Hansen and Jagannathan (1991) discussed relations between the mean and variance of the pricing kernel, connecting this to assets' Sharpe ratios. Cochrane and Saá-Requejo (2000) adapted these results to asset pricing and initiated the phrase "good-deal bounds" for their price bounds based on a ceiling for the variance of the pricing kernel. The point is to bound the prices of payoffs based on the assumption that they should not have Sharpe ratios that are too high. Here "too high" means more than some arbitrary multiple of the highest Sharpe ratio of any replicable payoff. A similar approach to establishing bounds is taken by all papers discussed in this section.

Bernardo and Ledoit (2000) have an approach very similar to that of Cochrane and Saá-Requejo (2000), restricting not the Sharpe ratio, but the gain–loss ratio $\mathrm{E}_{\mathbf{Q}}[X^+]/\mathrm{E}_{\mathbf{Q}}[X^-]$ of any payoff X replicable at zero cost. Here **Q** is a benchmark pricing measure: although one might not trust it to assign unique prices to all contingent claims, it can serve as the basis for assessing whether a deal is good in the sense that gains outweigh losses. Subjective considerations might be taken into account through the choice of benchmark pricing kernel. Bernardo and Ledoit (2000) relate the gain–loss ratio restriction to a restriction not on the pricing kernel's variance, but to bounds on the ratio between the pricing kernel and the benchmark pricing kernel. However, as Černý (2003, pp. 195–196) points out, it may not be possible to find any other pricing kernels that satisfy such a bound at any finite level. For example, in the Black–Scholes model, the ratio between pricing kernels is proportional to a power of the stock price, based on the equation $\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P} = \exp(-(\lambda^2/2)T - \lambda B(T))$, where λ is the market price of risk, B is Brownian motion under **P**, and T is the time horizon: this ratio is unbounded because $B(T)$ is unbounded.

A drawback of the Sharpe ratio approach is that the Sharpe ratio is a poor measure of preference, especially for derivative securities having nonlinear payoffs. As Bernardo and Ledoit (2000, p. 166) point out, this can cause the lower good-deal bound based on pricing kernel variance (Sharpe ratio) for an out-of-the-money call option to be zero, because the upside variance is too great. As Černý (2003, p. 193) illustrates, one payoff may stochastically dominate another, while the latter has a higher Sharpe ratio than the other. These problems relate to the defects of quadratic utility (Section 6).

Summaries of the main points and mathematical results of Bernardo and Ledoit (2000) and Cochrane and Saá-Requejo (2000) can be found in Geman and Madan (2004). For an example of good deal bounds, involving Sharpe ratios, applied to pricing European options on a stock following the Heston stochastic volatility model, see Bondarenko and Longarela (2004, §4.2). Björk and Slinko (2006) provide a solid mathematical foundation for good deal bounds based on Sharpe ratios in the case of a continuous-time underlying price process that has jumps.

Černý (2003) proposes to adapt the approach of Cochrane and Saá-Requejo (2000) by replacing the Sharpe ratio, with its connection to quadratic utility, with a *generalized Sharpe ratio* based on a more suitable utility function. For example, using exponential utility corresponds to a bound on relative entropy (Section 7), power utility corresponds to a bound on the expectation of a negative power of the pricing kernel, and log utility corresponds to a bound on the expected log of the pricing kernel. A 6-period example of a call option shows that the good deal bounds do not depend very much on the choice of utility function, i.e. of which generalized Sharpe ratio to use, but depend strongly on the level of the bound which defines the set A of good deals (Černý, 2003, §4.2). This makes sense, as changing the utility function changes the shape of A, while changing the bound changes the size of A.

In a dynamic model, one can implement the pricing kernel restrictions globally or locally (see Section 4.1.2). The local approach in continuous time rules out *instantaneous good deals*, forbidding any pricing kernel such that, if one could trade all claims frictionlessly at the prices it assigns, one could increase expected utility at too fast a rate at any instant. This relates to the local utility maximization of Kallsen (2002a), whose "neutral derivative pricing" assigns prices such that the opportunity to trade in derivatives does not allow for greater local utility than does trading in marketed securities alone. He derives price bounds by considering prices that are consistent with a limited nonzero position for the derivative security within an optimal portfolio. Exploring the local approach, Černý (2003, §5.1) concludes that, if the discounted gains processes from portfolio strategies are Itō processes, then ruling out instantaneous good deals imposes, for all utility functions having the same coefficient of absolute risk aversion, the same bound on the norm $\|\lambda(t)\|$ of the market price of risk vector stochastic process λ. As the minimal martingale measure of Section 6 corresponds to a pricing kernel $\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P} = \exp(-\int_0^T \|\lambda(t)\|/2\,\mathrm{d}t - \int_0^T \lambda(t)\,\mathrm{d}B(t))$ where $\|\lambda(t)\|$ is minimal for each t, this means the instantaneous good deal bounds always contain the value assigned by the minimal martingale measure (Černý, 2003, §7).

9.2 Equilibrium and stochastic dominance

Other pricing kernel restrictions are related to equilibrium among expected utility maximizers. Structural considerations can impose direct restrictions on the pricing kernel, and bounds for prices of nonreplicable payoffs can be constructed by comparison to replicable payoffs.

A structural feature of equilibrium among expected utility maximizers is that the pricing kernel should be decreasing in aggregate wealth (or consumption, depending on the economic model). Usually, the pricing kernel is decreasing in the price of an asset that is held in net positive supply, unless it has a negative association with aggregate wealth. For example, if there are two assets, a stock in net positive supply and a bond in net zero supply, the set of pricing kernels can be restricted to include only those that are decreasing in the stock price. Chazal and Jouini (2004) show that this restriction can significantly tighten the option pricing bounds when added to restrictions on the first two moments in the manner of Lo (1987). This approach goes back at least to Perrakis and Ryan (1984), who also initiated a literature on option price bounds based on comparisons among portfolios.

That method might be thought of as ruling out comparative good deals. Perrakis and Ryan (1984) used the CAPM pricing rule, in which the expected return of a portfolio is an affine function of the covariance between a representative investor's marginal utility of consumption and the portfolio's final value. Although this marginal utility and the distribution of portfolio value may not be known, the comparison of three portfolios allows Perrakis and Ryan (1984) to formulate bounds for the price of a European call option in a model with one stock and bond. The lower and upper bounds involve an expectation of a function of the terminal stock price, discounted at either the risk-free rate or the stock's expected return, respectively. To use the bounds, one need not know the statistical probability measure **P**, but one must know the **P**-expectations of some functions of the terminal stock price. Various extensions have been derived, involving intermediate trading, transaction costs, and puts. An apparent limitation of the methodology is the necessity, for each new security, to identify new comparison portfolios. For a review of subsequent literature related to Perrakis and Ryan (1984), see Constantinides and Perrakis (2002, §1).

Bizid and Jouini (2005) point out that an equilibrium among various agents in an incomplete market need not coincide with the equilibrium in a completion of that market. They demonstrate that the bounds imposed by very weak equilibrium conditions in an incomplete market, without assuming that the pricing kernel is a nonincreasing function of aggregate consumption, might be wider than those that result from considering any possible completion of that market. They view the reliance of Perrakis and Ryan (1984) on the CAPM as invoking a possible, but unknown, completion of the market. On the other hand, the CAPM might be justified not by market completeness, but by invoking the approximately quadratic preferences of well-diversified investors. Moreover, Constantinides and Perrakis (2002) rederive and extend results of

Perrakis and Ryan (1984) and related work, while relying on stochastic dominance rather than the CAPM.

They use stochastic dominance considerations to rule out option prices that allow trades in the option to increase expected utility, versus a portfolio only of marketed securities, under all increasing, convex utility functions. If a derivative security were offered for less than this lower bound, any expected utility maximizer would prefer to buy some of it than to keep all wealth invested in the market. This is very much in the spirit of bounding the Sharpe ratios, etc., of all payoffs to be no more than a certain multiple of the maximum Sharpe ratio of a replicable payoff. The differences are that the multiple is fixed at 1, and no particular measure such as a generalized Sharpe ratio is used, rather, a price is excluded only if it gives rise to an increase in any expected utility. The CAPM approach is even more similar, with the excess return "alpha" in the CAPM substituting for the Sharpe ratio.

For stochastic dominance constraints in optimization, which may be applied to portfolio optimization or pricing, see Dentcheva and Ruszczyński (2003).

10 Ambiguity and robustness

Risk, as something that can be quantified by means of a probability distribution, is to be distinguished from *ambiguity* or *Knightian uncertainty*, which represent a greater degree of ignorance. (Sometimes "uncertainty" is used more broadly to include both risk and ambiguity.) When we can assign a probability distribution and compute risk, we know something. For example, suppose we can assign to a potentially infinite sequence of repeatable experiments a probability measure such that the experiments are independent, and an event F has probability 30% of occurring in any repetition. Although we do not know whether F will occur in the next repetition, we do know, by the law of large numbers, that the fraction of experiments in which F occurs will eventually be between 29.99 and 30.01%. If we do not possess such knowledge, then we cannot assign a probability measure to this phenomenon, and we may require concepts such as that of imprecise probability (Walley, 1991). We may, for example, regard all probability measures in a set $\mathcal{P}$ as plausible, and all those not in $\mathcal{P}$ as implausible, and assign $\inf_{\mathbf{P}\in\mathcal{P}} \mathbf{P}[F]$ and $\sup_{\mathbf{P}\in\mathcal{P}} \mathbf{P}[F]$ as bounds for the probability of F. There is a substantial literature devoted to the Ellsberg (1961) experiment, which showed that such considerations affect willingness to gamble: subjects prefer to bet on unambiguous gambles rather than ambiguous ones, and these preferences are not consistent with maximizing any expected utility function. This may be because subjects have no faith that they can describe an ambiguous gamble with a single probability measure.

The unreliability of our stochastic models of financial markets suggests that ambiguity should be an important consideration in financial engineering. As discussed in Section 3.3, to be able to hedge all payoffs perfectly, the hedger must be in a complete market and know the stochastic process that marketed security prices follow. Ambiguity about this stochastic process is a source of effective incompleteness, as it becomes impossible to find the perfect hedge.

The theme of the application of ambiguity to incomplete markets is the decomposition of uncertainty about eventual outcomes into risk and ambiguity. A trader's aversions to delaying consumption (intertemporal substitution), to risk, and to ambiguity all determine the price at which he is willing to trade. Aversion to ambiguity is often described as a desire for robustness to misspecification of the stochastic model. A common approach in the financial literature (Chen and Epstein, 2002; Anderson et al., 2003; Maenhout, 2004) is to consider an equilibrium in which all traders have the same preferences, resulting in an analysis of assets' equilibrium expected returns in terms of market prices of risk and of ambiguity. Anderson et al. (2003) conclude, "Because mean returns are hard to estimate, ... there can still be sizable model uncertainty premia in security prices," and Maenhout (2004) concurs: "Empirically a 3% to 5% wedge is difficult to detect given the usual length of available time series. Given plausible values of risk aversion and uncertainty aversion, an equilibrium equity premium between 4% and 6% can then be sustained." Liu et al. (2005) proceed in similar fashion, but consider only ambiguity about rare jump events, not diffusion coefficients, and examine the impact on option prices. This main stream of financial research, discussed in Section 10.1, is an example of equilibrium marginal indifference pricing. A somewhat different, subjective approach occupies Section 10.2.

10.1 Complete preferences

In Section 4.2.3, expected utility maximization was criticized as a basis for portfolio optimization or derivative security pricing because it is too sensitive to unknown inputs, such as the probability measure. This defect has inspired work on *robust utility*. The literature looks back primarily to Gilboa and Schmeidler (1989), who considered portfolio optimization in which the expected utility $\mathrm{E}[u(V)]$ of random wealth V is replaced by

$$U(V) = \inf_{\mathbf{P}\in\mathcal{P}} \mathrm{E}_{\mathbf{P}}\big[u(V)\big], \tag{13}$$

where $\mathcal{P}$ is a set of plausible probability measures or *multiple priors*. Given $\mathcal{P}$, one can choose a portfolio to maximize the robust utility of Eq. (13), which as a preference function specifies complete preferences and is a foundation for indifference prices (Section 4.2.2). Talay and Zheng (2002) describe such an approach to derivative security pricing given considerations of model risk.

Although the form of Eq. (13) makes it look like a convex risk measure (Föllmer and Schied, 2002), not all convex risk measures have an interpretation in terms of ambiguity or robustness. For example, Schied (2004) describes the problem of maximizing robust utility functionals, equivalently, minimizing convex risk measures, subject to a capital constraint. He provides more explicit results for law-invariant risk measures ρ, where $\rho(X)$ depends only on the law of X under $\mathbf{P}_0$, a reference probability measure. This lacks the interpretation of ambiguity, in which the law of X under other measures also counts. An example of a law-invariant risk measure is expected shortfall, defined by $\mathcal{P} = \{\mathbf{P} \mid \mathrm{d}\mathbf{P}/\mathrm{d}\mathbf{P}_0 \leqslant r\}$, i.e. a pointwise (almost sure) constraint on the likelihood ratio. This considers only one probability measure $\mathbf{P}_0$, but all conditional probability measures $\mathbf{P}_0[\cdot|F]$, where F is an event such that $\mathbf{P}_0[F] < 1/r$.

The pointwise constraint on the likelihood ratio contrasts with a constraint on the relative entropy $\mathrm{E}_{\mathbf{P}}[\ln(\mathrm{d}\mathbf{P}/\mathrm{d}\mathbf{P}_0)]$. The set

$$\mathcal{P} = \big\{\mathbf{P} \mid \mathrm{E}_{\mathbf{P}}\big[\ln(\mathrm{d}\mathbf{P}/\mathrm{d}\mathbf{P}_0)\big] < \epsilon\big\} \tag{14}$$

can be interpreted as a set of probability measures that are plausible, given that econometric inference leaves $\mathbf{P}_0$ as the best estimate, but the econometrician remains uncertain as to the true probability measure. After estimation, some probability measures are more plausible, i.e. have a higher p-value or posterior likelihood, than others. An entropy criterion can be tractable, at least if one works with the intersection of $\mathcal{P}$ in Eq. (14) with a family of models, such as diffusions. However, entropy may not be a suitable way of describing which probability measures are plausible. It may be that different events have different levels of ambiguity, or that some aspects or parameters of the model are more ambiguous than others. For example, practitioners of financial engineering often have less confidence in their estimates of correlations or of means than of volatilities. Interesting effects arise when one considers that there may not simply be one correct entropy penalty or constraint for all traders to use in accounting for the ambiguity surrounding a probability measure estimated from commonly available data. Some assets may be more ambiguous than others, which can lead to under-diversification (Uppal and Wang, 2003) or cause negative skewness in short-term returns and premia for idiosyncratic volatility (Epstein and Schneider, 2005). Different traders may assign different levels of ambiguity to assets, which could explain the home-bias puzzle in investments (Epstein and Miao, 2003) and limited participation in the stock market (Cao et al., 2005).

Anderson et al. (2003) consider a portfolio-optimizing econometrician who wishes to construct a portfolio whose utility is robust with respect to the ambiguity about the true probability measure, that is, is high for all alternatives which remain plausible given the observed data. This leads them to an optimization including a penalty proportional to the relative entropy between each model under consideration and the best-fit model. Results can be computed using a worst-case model among those that are plausible. Maenhout (2004) considers a more tractable version of this methodology, in which the entropy penalty depends on wealth in a way that makes the optimal portfolio weights wealth-independent, and gives some more explicit results. An alternative to a penalty on relative entropy is a constraint on relative entropy, as in Eq. (14). Entropy penalty and entropy constraint model different preferences, but not only do they both result in the use of a worst-case model, entropy-penalty and entropy-constraint problems come in pairs sharing the same solution, i.e.

worst-case model and optimal portfolio (Hansen et al., 2006, §5). Thus, it is not possible to deduce whether a trader's portfolio is the result of solving a problem with an entropy penalty or constraint.

This issue is known as *observational equivalence*, and it appears frequently in the finance literature. Skiadas (2003) shows that, in a market driven by Brownian motion, the entropy-penalty value function coincides with that of stochastic differential utility (SDU): see also the discussion of source-dependent risk aversion in Skiadas (2006). Maenhout (2004) shows that his homothetic version of the entropy-penalty formulation is also observationally equivalent to SDU, but emphasizes that this observational equivalence is limited to portfolio choice and asset prices within a single model. The observational equivalence arises because the solution to the portfolio optimization with robust preferences reduces to the use of a worst-case model, which can then be mapped to a specific case of SDU. However, if market opportunities change, then the worst-case model will also change, becoming equivalent to a different case of SDU, so the observational equivalence breaks down in a broader context (Chen and Epstein, 2002, §1.2).

Moreover, from a financial engineering perspective, different methods that may yield the same answer given different inputs are different. As instrumental rather than descriptive devices, one method may be superior: it may be easier to specify good inputs and compute a useful result with one method than the other. For example, when preferences featuring risk aversion and ambiguity aversion are observationally equivalent to preferences featuring risk aversion only, the level of risk aversion is greater in the latter case. It would be easier to specify risk aversion and ambiguity aversion by introspection than to guess what level of risk aversion alone yields the same price. Indeed, Maenhout (2004) uses ambiguity aversion to explain the equity premium puzzle, which is that the level of risk aversion required to justify an expected return for equities matching the historical average is implausibly high when compared to the level of risk aversion that most subjects display when confronted with unambiguous gambles (Mehra, 2003; Mehra and Prescott, 2003). However, it may be that they display much greater risk aversion in financial markets, much of which is actually generated by aversion to these markets' ambiguity. Liu et al. (2005) find that aversion towards ambiguity about rare events involving jumps in the aggregate endowment can account for option pricing smirks. Routledge and Zin (2004) model fluctuations in the liquidity supplied by market-makers who have multiple priors, giving explicit, simple examples of OTC option trading, with the market-maker's optimal bid, ask, and hedges based on robust utility.

It is possible to construct a set $\mathcal{P}$ of multiple priors on principles other than entropy. One major motivation for not using the tractable entropy methodology is *dynamic consistency*. Various versions of dynamic consistency have been much discussed in the recent literature on risk measures: see Roorda and Schumacher (2005). Roughly speaking, dynamic consistency means that for payoffs X and Y occurring at time T, if X will always be preferred to Y at time $t < T$, then X must be preferred to Y at any time $s < t$. To do otherwise would create inconsistency between choices at different times. Whether such inconsistency is unacceptable depends on the application: for example, it is more troublesome in regulation than in pricing OTC securities. Such inconsistency might even be appropriate given certain kinds of beliefs incorporating ambiguity: see Epstein and Schneider (2003, §4) and Roorda et al. (2005, §4).

Dynamic consistency requires that the set $\mathcal{P}$ of multiple priors be *rectangular*. In a discrete-time model, this means that $\mathcal{P}$ has the following property: for any event F_i that involves only step i, any event $F_{<i}$ that involves only steps $1, \ldots, i-1$, and any pair $P_1, P_2 \in \mathcal{P}$, there must exist a $P_3 \in \mathcal{P}$ such that $P_3(F_i \cap F_{<i}) = P_1(F_i)P_2(F_{<i})$. That is, any one-step conditional probability for step i must appear in combination with all probability measures for steps $1, \ldots, i-1$. For example, if $\mathcal{P}$ contains two probability measures, both of which correspond to multi-period binomial models of a log stock price with independent increments, and one of them says that the probability of an up move is 40% and the other says that it is 60%, then $\mathcal{P}$ is not rectangular. It would also have to contain, among others, a probability measure under which the probability of an up move is 40% at step 1 and 60% at step 2.

Rectangularity leads to the preference structure known as "recursive multiple priors" (Chen and Epstein, 2002; Epstein and Schneider, 2003), which can be viewed as a combination of stochastic differential utility and the robust utility of Equation (13). The set $\mathcal{P}$ of multiple priors defined by an entropy constraint is not rectangular (Epstein and Schneider, 2003; Hansen et al., 2006). Advocates of dynamic consistency suggest enlarging a non-rectangular candidate set of multiple priors until it becomes rectangular, while others (e.g. Hansen et al., 2006, § 9) object that the resulting rectangular set is too large, depriving the modeler of the ability to impose interesting restrictions on probabilities. In terms of pricing in incomplete markets, the result is price bounds that are too wide.

10.2 Incomplete preferences

Using the robust utility of Eq. (13) to define complete preferences as in Gilboa and Schmeidler (1989) is suitable for the application of a one-time portfolio optimization, in which a portfolio strategy is chosen with a pessimistic attitude in the face of ambiguity about which of the probability measures in $\mathcal{P}$ is correct. The result is the selection of a worst-case model $\mathbf{P}^* \in \mathcal{P}$, similar to the least favorable completion mentioned in Section 4.4. The methods discussed in Section 10.1 price all payoffs under an equilibrium pricing measure $\mathbf{Q}^*$ derived from $\mathbf{P}^*$. Assigning this price to all payoffs at all times would reflect an ongoing concern with maximizing expected utility under the worst-case model, and no concern for expected utility under any other plausible model in $\mathcal{P}$.

To see what might be undesirable about this, consider the difference between optimizing over random total wealth and optimizing over a payoff, which is a change in wealth, discussed in Section 4.2.2. Also, whereas indifference pricing is based on complete preferences, no-arbitrage pricing and other good deal bounds are based on incomplete preferences. No-arbitrage pricing is based on the incomplete preference such that V is weakly preferable to W when $\operatorname{ess\,inf}(V - W) \geqslant 0$, i.e. $V \geqslant W$. When neither $V \geqslant W$ nor $W \geqslant V$, this preference structure expresses neither indifference nor preference between V and W, but rather cannot decide between them. A complete preference structure using the essential infimum as the preference function for total wealth evaluates portfolios based on the worst-case scenario: V is preferred to W if $\operatorname{ess\,inf} V > \operatorname{ess\,inf} W$. This is not a suitable preference structure for financial decisions. According to this preference function, it is better to get one cent for sure than to have a 99.99% chance of getting one million dollars and a 0.01% chance of getting nothing.

The same problem can occur with the Gilboa and Schmeidler (1989) robust utility. If the set $\mathcal{P}$ of plausible measures is large, reflecting a great degree of ambiguity, we may find that a change in the portfolio that increases expected utility under the worst-case measure decreases it under other plausible measures. Then we may lack confidence that this change is an improvement, or even suspect it of being a bad deal. In other words, the acceptance set

$$A_{GS} = \Big\{ Z \mid \inf_{\mathbf{P}\in\mathcal{P}} \mathrm{E}\big[u(V + Z)\big] \geqslant \inf_{\mathbf{P}\in\mathcal{P}} \mathrm{E}\big[u(V)\big] \Big\} \tag{15}$$

defined by robust utility for use in subjective good deal bounds (see Section 4.2) may not be suitable as a set of good deals.

An alternative is robust evaluation not of total wealth but of changes in it, or equivalently, incomplete preferences over portfolios, as in no-arbitrage price bounds. This corresponds to the incomplete preference scheme of Bewley (2002), in which the acceptance set is

$$A_B = \Big\{ Z \mid \inf_{\mathbf{P}\in\mathcal{P}} \mathrm{E}\big[u(V + Z)\big] - \mathrm{E}\big[u(V)\big] \geqslant 0 \Big\} \subseteq A_{GS}. \tag{16}$$

That is, a change is considered a good deal if it increases expected utility under *every* plausible probability measure, not if it merely increases expected utility under the worst-case measure. This smaller acceptance set is more conservative in that it recognizes fewer good deals and thus leads to wider good deal bounds. This Bewley (2002) approach also responds better to an error of wrongly including an implausible measure $\mathbf{P}_x$ in $\mathcal{P} = \mathcal{P}' \cup \{\mathbf{P}_x\}$, where $\mathcal{P}'$ is the correct set of plausible probability measures. Then there might be a payoff Z such that $\inf_{\mathbf{P}\in\mathcal{P}} \mathrm{E}[u(V + Z)] > \inf_{\mathbf{P}\in\mathcal{P}} \mathrm{E}[u(V)] \geqslant \inf_{\mathbf{P}\in\mathcal{P}'} \mathrm{E}[u(V)] > \inf_{\mathbf{P}\in\mathcal{P}'} \mathrm{E}[u(V + Z)]$, so that $\mathbf{P}_x$ is the worst-case model and the Gilboa and Schmeidler (1989) approach would have us erroneously switch from V to $V + Z$, which actually makes us worse off. The Bewley (2002) approach focusing on changes in portfolios would only cause us wrongly to reject some good deals, not wrongly accept bad deals.

The question is how aversion to ambiguity manifests itself in OTC market-making. Is one willing to pay high prices for "ambiguity hedges," that is, payoffs that reduce the ambiguity of one's expected utility? Or does one accept only unambiguously good deals, paying a low enough price so that it is implausible that they do not improve one's portfolio?

11 Calibration

It is standard practice to price with an incomplete-markets model much as described in Section 2.2 for complete-markets models, by calibrating **Q** to marketed securities' prices and assigning the expected discounted payoff $E_{\mathbf{Q}}[DX]$ as the price for a payoff X. If one calibrates to a family of complete-market models containing the true model, then **Q** must be the unique no-arbitrage pricing measure. However, if the market is incomplete, choosing **Q** such that $E_{\mathbf{Q}}[DS]$ is the market price for any payoff S of a marketed security does not guarantee that $E_{\mathbf{Q}}[DX]$ is an arbitrage-free price for any payoff X. This is merely a curve-fitting scheme. Arbitrage-free pricing requires that **Q** be equivalent to the statistical probability measure **P**. Moreover, it is characteristic of incomplete markets that more than one pricing measure **Q** yields arbitrage-free prices.

Researchers who propose new incomplete-markets models of underlying asset prices often provide a formula for a single "risk-neutral" price, which appeals to practitioners. A typical procedure is to posit a model for the statistical probability measure **P**, next to assume that one should look for an equivalent pricing measure **Q** of the same parametric form, and finally to relate the parameters under **P** and **Q**. The last step can be done by means of an unspecified market price of risk (e.g. Heston, 1993), or through construction of an equilibrium among expected utility maximizers, in which case unspecified parameters of the utility function are involved (e.g. Madan et al., 1998; Kou, 2002). This last step is not important in practice, because practitioners calibrate the parameters of **Q** to market prices without any regard to **P**.

What usually happens is that parsimonious models, with a small number of parameters, cannot exactly match the prices of all marketed securities: the models are not perfect.[3] Although multiple pricing measures **Q** are consistent with observed market prices, in practice, none of the probability measures within the family under consideration will be perfectly consistent. Calibration selects the member of the family that is most consistent. The rationale for using prices based on calibration in OTC trading is that because these prices are nearly consistent with market prices, they are likely to avoid arbitrage and to assign reasonable values to payoffs if market prices are reasonable.

That is, if market prices exclude arbitrage and good deals, then it seems likely that OTC prices calibrated to market prices should also exclude arbitrage and good deals. However, the plausibility of this conclusion depends on how

[3] Models with many parameters may fit the data exactly, but their calibrated parameters tend to change substantially over time, a sign that they are not perfect either; they tend to suffer from over-fitting.

similar payoffs of OTC securities are to those of marketed securities. If there are no marketed securities whose prices yield information through the model about events that are important to valuing the OTC securities, the scheme will be unreliable. An example is the calendar spread on swaption straddles discussed in Section 2.3. Good price quotes are available for swaptions expiring on one of a limited set of dates, all of which are much more than two days apart. Therefore, calibration to these swaptions' prices can only give information about the total volatility under **Q** of interest rates over the long periods between adjacent expiration dates, not about how volatility is spread between the expiration dates. Typical calibration schemes interpolate smoothly, assuming that there is no reason for volatility to be concentrated. However, interest rate volatility under **P** *is* concentrated around dates of scheduled major economic announcements. Therefore the prices at which JP Morgan sold the calendar spreads on swaption straddles, although consistent with market prices of swaptions, resulted in a good deal for the customers and a bad deal for JP Morgan. Having a better model could not solve this problem, because market prices do not contain the information required to calibrate the model. What is needed is a better method, one which is grounded in an assessment of statistical probabilities, allowing the trader to base pricing on such information as the concentration of volatility around economic announcements.

12 Conclusion

One might dream of a unified theory of contingent claim valuation in incomplete markets, covering not only derivative securities but also equilibrium pricing of underlying securities and corporate investment via the real options approach. Although these applications have much in common, we have focused entirely on making a market in OTC derivatives, in the belief that the practical settings of these applications differ so much that any valuation methodology should be evaluated differently, depending on the use to which it is to be put. For example, when making a market in derivatives, a trader is concerned that potential customers may possess superior information, while executives making corporate investment decisions are concerned that interested subordinates may have provided biased information about future cashflows; also, hedging is of paramount importance in trading derivatives, but of at most secondary importance in corporate investment. For the application of OTC derivatives market-making, we want a valuation methodology that is robust to misspecification of inputs that are hard to infer, that ensures that each trade made at the bid or ask prices is beneficial, and that is tractable, allowing rapid computation.

What is beneficial may be the subject of some debate, but a suitable valuation methodology should either have an appropriate economic grounding, as expected utility indifference pricing does, or be shown to give answers that agree with the results of a well-grounded method under specified circumstances. The economic grounding should involve the subjective situation of the market-maker who is considering a trade, including his current portfolio, the risk management framework in which he operates, and his future opportunities; again, an objective methodology that does not take account of the individual's situation would be appropriate only in circumstances in which it could be shown to yield results that are subjectively beneficial. For example, if some objective good deal bounds were wider than a trader's subjective good deal bounds and still narrow enough to be usable, they would be appropriate. It is more helpful to a decision-maker to identify a price at which trade benefits him than to identify a "fair price"; $300,000 might be a fair price for a luxury automobile, but if one cannot resell it, it might not be beneficial to buy it at or near this price.

We conclude by assessing the extent to which various methods achieve these desiderata. Along the way, we will point out some cases in which further research is needed for such an evaluation. We focus on a few major kinds of methods. These include, first, the standard practice of calibrating to market prices without reference to a statistical probability measure. Second are methods based on expected utility maximization and indifference, including marginal indifference pricing or minimum-distance measures, whether founded on local or global criteria. Third, there are methods of pricing kernel restriction founded on constraints, such as pricing with low-distance measures rather than minimum-distance measures. Finally, there are methods that account for ambiguity, whether they deal with it by using just a worst-case model or by considering all plausible models.

The most tractable method is calibration: it prices all payoffs by taking expectations under a single probability measure calculated by a single parametric optimization. Next best are other methods that also price using just one probability measure, such as any minimum-distance method or marginal indifference pricing, whether it is founded on expected utility or robust utility yielding a worst-case model. It appears that it takes more work to identify these single measures than calibration requires. Less tractable than these is non-marginal indifference pricing, which is not simply pricing under one measure: it requires a new optimization to price each payoff. Local variants are more tractable than global variants. Pricing kernel restrictions and robust methods that use optimization over multiple probability measures look most difficult of all. These optimizations may be non-parametric, e.g. requiring computation of a pricing kernel in all states.

Robustness is the aim of the methods founded on multiple statistical probability measures, but it remains to be confirmed by extensive empirical study that the resulting prices are indeed robust to statistical sampling error. Pricing kernel restrictions featuring low-distance measures also use a single statistical probability measure **P**, but use multiple pricing measures; their robustness too is an open question. Any method involving expected utility indifference or distance minimization, including the quadratic and exponential special cases, is not robust with respect to the statistical probability measure **P**.

Calibration is more robust to its observed inputs, because it is easier to observe market prices than to infer the parameters of econometric models; however, because price data may be out of date, erroneous, or have noise due to market microstructure and bid–ask spreads, robustness to this data is still an issue. The more parsimonious models tend to be more robust. Calibration is not robust to the choice of the family $\mathcal{P}$ of models within which calibration takes place. The resulting risk of trading losses is known as *model risk*. It would be interesting to know whether model risk can be mitigated by using multiple calibrated models $\hat{\mathbf{Q}}$ from different families $\mathcal{P}$ in the manner of the robust methods (Section 10), or by using all the models from a single family that have sufficiently low calibration error, instead of just the one with minimal error, in the manner of low-distance measures (Section 9).

To ensure that trades made at bid and ask prices are beneficial, it helps to use a method that produces price bounds that are suitable for use as bid and ask prices. When using a method that produces unsuitable price bounds, or a single price, a trader is reduced to intuition in setting bid and ask prices, making it difficult to tell whether trades include adequate compensation for unhedgeable risk.

Expected utility indifference pricing, based on the trader's optimized portfolio, is the paradigm of a method for generating price bounds that are beneficial, but this method's fatal flaw is its lack of robustness. Either the trader must optimize his portfolio according to an unreliable expected utility maximization procedure, or the indifference prices are suited to an imagined optimal portfolio, not his actual portfolio (Section 5.2.1).

The methods founded on marginal indifference pricing and minimum-distance measures have weaker economic grounding than expected utility indifference pricing. It remains to be seen whether and under what circumstances they yield results that are approximately the same as expected utility indifference pricing, despite the apparent flaws of various of these methods, such as producing a single price, using an inappropriate utility function, and an objective orientation that disregards the trader's portfolio. This last point also affects the methods based on low-distance measures. A major issue in using them is the question of how great a distance is "low."

Calibration provides no reason to believe that trading at the resulting price is beneficial. Its successes have much to do with traders' skillful use of their experience and intuition, the ability to hedge well in very competitive OTC markets, and large bid–ask spreads in OTC markets where hedging is harder. Its failures point to its limitations. As hedge funds know, the ability to price all OTC securities well must include an assessment of the statistical probability measure $\mathbf{P}$, which calibration avoids. A synthesis with econometrics is desirable. This returns us to the problem of robustness to statistical errors in specifying $\mathbf{P}$. Methods based on robustness to subjective ambiguity try to overcome this problem while retaining the justifiability of expected utility indifference prices, for instance, by using robust utility. However, it remains to show how to model and quantify the ambiguity left after econometric inference in a way that yields a useful method for OTC pricing.

A fundamental question is how we derive information from current market prices of derivative securities and from econometric study of underlying securities' price histories. In particular, how do we respond when derivative securities' current prices seem to be out of line with our beliefs about underlying securities' future prices, as expressed in the statistical probability measure **P**, making possible a good deal by trading in marketed securities? If we are not only making a market in OTC securities but also willing to speculate or invest in marketed securities, then this is an opportunity to trade against a perceived mispricing. This trade would generate enough risk for our indifference price bounds for marketed securities to adjust so that they contain the actual market prices. If we are not willing to speculate or invest in marketed securities, do we simply take account of their market prices when computing hedging costs, or do we infer something about **P** based on the belief that good deals should not exist? If the latter, inference about **P** might seek not just to maximize statistical likelihood, but to balance this objective with minimizing a distance to the set $\mathcal{Q}$ of EMMs.

Acknowledgements

The author is grateful for the support of the National Science Foundation under Grant No. DMS-0202958 and of the National Security Agency under Grant No. H98230-04-1-0047. He thanks Philippe Artzner, Aleš Černý, Dmitry Kramkov, and Costis Skiadas for valuable discussions and for suggesting references. He remains responsible for the views expressed and any errors.

Appendix A. Definition of incompleteness and fundamental theorems

We might like to define a complete market as one in which it is possible to replicate any cashflow. This raises several questions. What is the set C of cashflows that we hope to be able to replicate? What is the set Θ of possible portfolio strategies with which we hope to replicate them? What does it mean to replicate?

First, we must specify the set C of cashflows to be replicated. As usual, let us focus on cashflows that are simply random variables representing a payoff at a terminal time T. In assessing completeness of the market, it makes sense to consider replication only of payoffs that are functions of underlying financial variables observed over the time interval $[0, T]$. Even this set is too large for mathematical convenience, and the literature usually imposes a further restriction that the payoffs under consideration must be integrable or bounded. There are also economic reasons for a boundedness restriction.

The second question is of which portfolio strategies are allowed, and with limited credit, it is impossible to execute a portfolio strategy whose value is unbounded below. Along with this restriction of "admissibility" or "tameness," we also restrict attention to portfolio strategies that are self-financing (after any transaction costs), for the same reasons as in the study of no-arbitrage pricing and the first fundamental theorem of asset pricing. We must also consider only portfolio strategies whose initial cost is finite; this is a substantive restriction in models with an infinite number of marketed securities. We might also consider imposing other restrictions. For instance, we may consider only "stopping time simple" portfolio strategies, which (almost surely) include only a finite number of times at which the portfolio is rebalanced, because continuous-time hedging is impossible. We might also restrict the number of marketed securities in the portfolio to be finite, even if an infinite number are available in the model, for similar reasons. The result of defining the set Θ of possible portfolio strategies is a set of exactly replicable payoffs, $R := \{Y \mid \exists\theta \in \Theta \ni Y = \theta_T S_T\}$, where S is the stochastic process of marketed securities' prices and $\theta_T S_T$ is the terminal value of portfolio strategy θ.

Third, what does it mean to replicate? Exact replication led to the definition of R, and one candidate definition for completeness is $C = R$, all payoffs can be exactly replicated. Jarrow et al. (1999) refer to this property as *algebraic completeness*, saying, "This definition is too strong and would hardly ever be satisfied in practice." After a discussion of mathematics, we will argue that algebraic completeness is unnecessarily strong, and completeness should be defined differently.

The mathematical finance literature originally focused on algebraic completeness, but this created difficulties with the second fundamental theorem of asset pricing (FTAP), which relates market completeness to uniqueness of a pricing kernel. These difficulties were analogous to those that previously beset the first FTAP, which relates absence of arbitrage to existence of a pricing kernel. The difficulties with the first FTAP were solved by introducing a weaker notion than arbitrage, namely the *free lunch with vanishing risk* (Delbaen and Schachermayer, 1999; Protter, 2006). While an arbitrage is a portfolio strategy in Θ with nonpositive initial cost and terminal value that is nonnegative and nonzero, a free lunch with vanishing risk is a sequence of portfolio strategies in Θ with nonpositive initial cost and whose limiting terminal value is nonnegative and nonzero. Mathematically, the idea is to replace the no-arbitrage condition with a stronger one, which excludes even approximate arbitrages, such as a free lunch with vanishing risk. What constitutes an "approximate" arbitrage is determined by a topology on the space of payoffs i.e. terminal values: the concepts of closure and limit depend on this topology (Cherny, 2005; Staum, 2004). By analogy, for the second FTAP, it would make sense to replace algebraic completeness with a weaker, topological notion (Battig and Jarrow, 1999; Jarrow et al., 1999; Jarrow and Madan, 1999). The resulting second FTAPs connect uniqueness of a pricing kernel that is continuous with respect to some topology to approximate replicability of any payoff $Y \in C$ in the sense that, for any neighborhood U of Y, there is a portfolio strategy in Θ whose terminal value is in U.

That is, the more successful mathematical notion of completeness relates the target payoffs C and the replicable payoffs R by means of a topology specifying what approximate replication means. This is an eminently practical notion, because we need only concern ourselves with whether a payoff can be approximately replicated. If we can find a hedging scheme that results in an arbitrarily small hedging error, we will be satisfied. An incomplete market, then, is one in which there are target payoffs that cannot even be approximately replicated, so that we must find a methodology for dealing with the resulting non-negligible residual risks after hedging.

Appendix B. Financial perspectives on incompleteness

B.1 Descriptive analysis: Are markets incomplete? How much so?

Whereas a financial engineer might directly test financial time series for features that are known to cause incompleteness (see Section 3.1), tests of market incompleteness in the financial literature often look for evidence of incompleteness in consumption data. As Saito (1999, §II) says, "When markets are complete, the intertemporal rate of substitution is equalized among agents." If so, then a calibrated representative agent model would reflect aggregate preferences, and microeconomic data would show that households are capable of fully insuring themselves against idiosyncratic risks. The approach focusing on calibration to aggregate data typically finds that calibrated parameters reflect implausible aggregate preferences. For instance, this is one guise of the equity premium puzzle. One response is to conclude that markets must not be complete after all. However, attempts to explain away the equity premium puzzle on the basis of incomplete markets have not been universally accepted (Mehra, 2003; Mehra and Prescott, 2003). An alternative conclusion is that the models being calibrated are themselves wrong, so that these tests do not correctly assess market completeness. However, Hansen and Jagannathan (1991) devised a test which is based on fewer assumptions and not specific to a particular model, and once again one is led to the conclusion that there is a puzzle if markets are complete. Another approach to testing is to use microeconomic data to show that household consumption has not been fully insured against idiosyncratic risks. A tentative conclusion is that markets are significantly incomplete, but some doubts may remain. See Saito (1999, §II) for further references.

B.2 Normative analysis: What should we do about incompleteness?

Completing the market increases welfare, but increasing the attainable span in an incomplete market without completing it may increase or decrease welfare. See Huang (2000, §III.A) for a qualitative summary and Duffie and Rahi

(1995, §2.2) for a mathematical synopsis of one result. Even something as apparently straightforward as an increase in welfare does not have clear normative implications. One way in which incomplete markets can lower welfare is by inducing agents to engage in precautionary saving as a substitute for unavailable insurance against risks. The resulting investment exceeds the level consistent with maximal welfare of agents existing today and thus produces economic growth which is in this sense excessive (Saito, 1999, §IV.B), but which may lead to greater welfare for future generations.

It is also unclear how great the welfare loss due to incompleteness is. Many factors influence the welfare loss generated within a model of an incomplete-market equilibrium: how many goods there are, whether the model describes only exchange or also production, what assets are marketed, whether there is aggregate risk or only idiosyncratic risk, and whether the time horizon and the persistence of shocks are infinite, short, or long relative to agents' patience, which has to do, for instance, with whether one can find a new job after being laid off, and with the length of business cycles. Levine and Zame (2002) ask "Does market incompleteness matter?" They answer that it does not in a model of an exchange economy with a single perishable good, agents who are patient, i.e. have a low discount rate in their intertemporal utility functions, and have an infinite time horizon, shocks that are not persistent, and only idiosyncratic risk; incompleteness matters if it prevents insuring against aggregate risks or the relative prices of multiple goods. To their question, Kübler and Schmedders (2001) respond unequivocally that "incomplete markets matter for welfares," even if agents are patient. Kim et al. (2003) study a simple international model of two countries and report that welfare loss is negligible when agents are patient and shocks are transitory, but is considerable and highly sensitive to the model's parameters in the more realistic case of patience and persistent shocks.

Equilibria in incomplete markets may even be Pareto inefficient given the constraints about contingent claims that cannot be traded, because agents make decisions based on the current equilibrium prices, whereas everyone's welfare might be increased at a different price system and allocation: see Hens (1998, §4), Huang (2000, §III.B), and Duffie and Rahi (1995, §3.3). This raises the possibility that suitably crafted regulatory intervention might increase welfare (for a simple example, see Huang, 2000, Appendix), but such a suggestion needs to be treated with the utmost caution, as the relevant central planning problem would require a tremendous amount of information: see Huang (2000, §§IV–VI) and Herings and Polemarchakis (2005).

There is a connection between Pareto inefficiency and the topic of sunspot equilibria in incomplete markets. On sunspot equilibrium see e.g. Hens (1998, §9). A sunspot equilibrium is one in which allocations of goods depend on extrinsic events, such as sunspots, having nothing to do with preferences, endowments, and production possibilities; agents may have self-fulfilling expectations associated with these extrinsic events, generating volatility in excess of what is warranted by fundamentals (Prescott and Shell, 2002). According to Hens (1998, §9.2), "sunspots matter if and only if markets are incomplete." Sunspot equilibria generate excess uncertainty and are Pareto inefficient and dominated by nonsunspot equilibria in strictly convex economies (Prescott and Shell, 2002). Pareto efficiency of sunspot and non-sunspot equilibria remains a subject of active research, e.g. Pietra (2004). The existence of sunspot and non-sunspot equilibria has an interesting relation to options. Antinolfi and Keister (1998) report that the introduction into a market of a small number of options can render it "strongly sunspot immune," i.e. eliminate the possibility of sunspot equilibria no matter what extrinsic phenomenon constitutes the sunspots; this is in contrast to previous results they cite, asserting that options can have a destabilizing effect. For instance, according to Bowman and Faust (1997), it is possible for the addition of a market in options to introduce sunspot equilibria into an economy that previously did not have any, even if that economy's market was already complete! This has to do with the fact that options, as derivative securities, have payoffs related to underlying security prices and not directly to the state of the economy.

Public prices reveal private information and one may analyze how much private information a certain market structure reveals (Duffie and Rahi, 1995, §3.2); recent work on this topic includes Kübler et al. (2002). One is tempted to suppose that complete revelation is desirable because it increases market efficiency (in the sense of the efficient markets hypothesis, not Pareto efficiency) and thus promotes the allocation of resources to maximally productive uses. However, the normative issues surrounding information revelation are not simple. When private information is not revealed, uninformed investors may be hesitant to trade; this is the rationale behind the prohibition on insider trading, and the insight underlying an extensive economic literature spawned by the famous paper on lemons (Akerlof, 1970). Yet private information might be revealed in a way that resolves uncertainty about individuals' endowments so that they cannot well insure it. Hirshleifer (1971) describes how public information can disrupt a market's ability to provide insurance. For example, suppose that the only uncertainty about the price of corn at harvest comes from ignorance about the total number of acres planted. In this case farmers would not be able to insure themselves well against price risk by hedging in futures markets, because the very act of their attempting to hedge their crops would reveal the total crop size, thus resolving all uncertainty about the price. Marin and Rahi (2000) and Dow and Rahi (2003) study this tension; the magnitude of these opposing informational effects is unknown.

References

Akerlof, G.A. (1970). The market for lemons – Quality uncertainty and market mechanism. *Quarterly Journal of Economics* 84 (3), 488–500.

Andersen, T.G., Benzoni, L., Lund, J. (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance* 57 (3), 1239–1284.

Anderson, E.W., Hansen, L.P., Sargent, T.J. (2003). A quartet of semigroups for model specification, robustness, prices of risk, and model detection. *Journal of the European Economic Association* 1 (1), 68–123.

Antinolfi, G., Keister, T. (1998). Options and sunspots in a simple monetary economy. *Economic Theory* 11, 295–315.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* 9, 203–228.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., Ku, H. (2007). Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research* 152, 5–22.

Battig, R.J., Jarrow, R.A. (1999). The second fundamental theorem of asset pricing: A new approach. *Review of Financial Studies* 12 (5), 1219–1235.

Becherer, D. (2003). Rational hedging and valuation of integrated risks under constant absolute risk aversion. *Insurance Mathematics and Economics* 33, 1–28.

Bernardo, A.E., Ledoit, O. (2000). Gain, loss, and asset pricing. *Journal of Political Economy* 108 (1), 144–172.

Bertsimas, D., Kogan, L., Lo, A.W. (2001). Hedging derivative securities and incomplete markets: An ϵ-arbitrage approach. *Operations Research* 49 (3), 372–397.

Bewley, T.F. (2002). Knightian decision theory: Part I. *Decisions in Economics and Finance* 25, 79–110. First appeared in 1986 as Cowles Foundation Discussion Paper No. 807.

Biagini, F., Pratelli, M. (1999). Local risk minimization and numéraire. *Journal of Applied Probability* 36 (4), 1126–1139.

Bizid, A., Jouini, E. (2005). Equilibrium pricing in incomplete markets. *Journal of Financial and Quantitative Analysis* 40 (4).

Björk, T., Slinko, I. (2006). Towards a general theory of good-deal bounds. *Review of Finance* 10 (2), 221–260.

Bondarenko, O., Longarela, I.R. (2004). Benchmark good-deal bounds: An application to stochastic volatility models of option pricing. *Working paper*.

Bowman, D., Faust, J. (1997). Options, sunspots, and the creation of uncertainty. *Journal of Political Economy* 105 (5), 957–975.

Boyle, P., Wang, T. (2001). Pricing of new securities in an incomplete market: The Catch 22 of no-arbitrage pricing. *Mathematical Finance* 11 (3), 267–284.

Brandt, M.W. (2003). Hedging demands in hedging contingent claims. *Review of Economics and Statistics* 85 (1), 119–140.

Cao, H.H., Wang, T., Zhang, H.H. (2005). Model uncertainty, limited market participation and asset prices. *Review of Financial Studies* 18 (4), 1219–1251.

Carr, P. (2002). Frequently asked questions in option pricing theory. *Journal of Derivatives*, in press.

Carr, P., Geman, H., Madan, D.B. (2001). Pricing and hedging in incomplete markets. *Journal of Financial Economics* 62, 131–167.

Carr, P., Geman, H., Madan, D.B., Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *Journal of Business* 75 (2), 305–332.

Černý, A. (2003). Generalised Sharpe ratios and asset pricing in incomplete markets. *European Finance Review* 7, 191–233.

Černý, A., Hodges, S. (2002). The theory of good-deal pricing in financial markets. In: Geman, H., Madan, D., Pliska, S., Vorst, T. (Eds.), *Mathematical Finance – Bachelier Congress 2000*. Springer-Verlag, Berlin, pp. 175–202.

Chazal, M., Jouini, E. (2004). Equilibrium pricing bounds on option prices. *Working paper*.

Chen, Z., Epstein, L. (2002). Ambiguity, risk, and asset returns in continuous time. *Econometrica* 70 (4), 1403–1443.

Cherny, A.S. (2005). General arbitrage pricing model: probability approach. *Working paper*.

Choulli, T., Stricker, C. (2005). Minimal entropy-Hellinger martingale measure in incomplete markets. *Mathematical Finance* 15 (3), 465–490.

Choulli, T., Stricker, C., Li, J. (2006). Minimal Hellinger martingale measures of order q. *Working paper*.

Clewlow, L., Hodges, S. (1997). Optimal delta-hedging under transactions costs. *Journal of Economic Dynamics and Control* 21, 1353–1376.

Cochrane, J.H., Saá-Requejo, J. (2000). Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *Journal of Political Economy* 108 (1), 79–119.

Constantinides, G.M., Perrakis, S. (2002). Stochastic dominance bounds on derivatives prices in a multiperiod economy with proportional transaction costs. *Journal of Economic Dynamics and Control* 26, 1323–1352.

Cont, R., Tankov, P. (2004). *Financial Modelling with Jump Processes*. *Financial Mathematics Series*. Chapman & Hall/CRC, London.

Cvitanić, J. (2000). Minimizing expected loss of hedging in incomplete and constrained markets. *SIAM Journal on Control and Optimization* 38 (4), 1050–1066.

Davis, M.H.A. (2004a). Complete-market models of stochastic volatility. *Proceedings of the Royal Society of London (A)* 460, 11–26.

Davis, M.H.A. (2004b). Valuation, hedging and investment in incomplete financial markets. In: Hill, J.M., Moore, R. (Eds.), *Applied Mathematics Entering the 21st Century: Invited Talks from the ICIAM 2003 Congress*. In: *Proceedings in Applied Mathematics*, vol. 116. Society for Industrial and Applied Mathematics, Philadelphia. Chapter 4.

Delbaen, F., Schachermayer, W. (1999). Non-arbitrage and the fundamental theorem of asset pricing: Summary of main results. In: Heath, D.C., Swindle, G. (Eds.), *Introduction to Mathematical Finance*. In: *Proceedings of Symposia in Applied Mathematics*, vol. 57. American Mathematical Society, Providence, RI, pp. 49–58.

Delbaen, F., Grandits, P., Rheinländer, T., Samperi, D., Schweizer, M., Stricker, C. (2002). Exponential hedging and entropic penalties. *Mathematical Finance* 12 (2), 99–123.

Dentcheva, D., Ruszczyński, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization* 14 (2), 548–566.

Dow, J., Rahi, R. (2003). Informed trading, investment, and welfare. *Journal of Business* 76 (3), 439–454.

Dritschel, M., Protter, P. (1999). Complete markets with discontinuous security price. *Finance and Stochastics* 3 (2), 203–214.

Duffie, D., Rahi, R. (1995). Financial market innovation and security design: An introduction. *Journal of Economic Theory* 65, 1–42.

Duffie, D., Gârleanu, N., Pedersen, L.H. (2006). Valuation in over-the-counter markets. *Working paper*.

Dunbar, N. (2005). JP Morgan reorganises US operations after trading losses. *RISK Magazine* 0, 12–13. January.

Dybvig, P.H. (1992). Hedging non-traded wealth: When is there separation of hedging and investment? In: Hodges, S. (Ed.), *Options: recent advances in theory and practice, vol. 2*. Manchester Univ. Press, NY, pp. 13–24. Chapter 2.

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

El Karoui, N., Quenez, M.-C. (1995). Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM Journal on Control and Optimization* 33 (1), 29–66.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics* 75, 643–669.

Epstein, L.G., Miao, J. (2003). A two-person dynamic equilibrium under ambiguity. *Journal of Economic Dynamics and Control* 27, 1253–1288.

Epstein, L.G., Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory* 113, 1–31.

Epstein, L.G., Schneider, M. (2005). Ambiguity, information quality and asset pricing. *Working paper*.

Foldes, L. (2000). Valuation and martingale properties of shadow prices: An exposition. *Journal of Economic Dynamics and Control* 24, 1641–1701.

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

Föllmer, H., Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics* 6 (4), 429–447.

Föllmer, H., Schweizer, M. (1991). Hedging of contingent claims under incomplete information. In: Davis, M.H.A., Elliott, R.J. (Eds.), *Applied Stochastic Analysis*. In: *Stochastics Monographs*, vol. 5. Gordon and Breach, New York, pp. 389–414.

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

Frittelli, M. (2000b). The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance* 10, 39–52.

Fujiwara, T., Miyahara, Y. (2003). The minimal entropy martingale measures for geometric Lévy processes. *Finance and Stochastics* 7, 509–531.

Geman, H., Madan, D.B. (2004). Pricing in incomplete markets: From absence of good deals to acceptable risk. In: Szegö, G. (Ed.), *Risk Measures for the 21st Century*. Wiley, Hoboken, NJ, pp. 451–474. Chapter 21.

Gilboa, I., Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18, 141–153.

Goll, T., Rüschendorf, L. (2001). Minimax and minimal distance martingale measures and their relationship to portfolio optimization. *Finance and Stochastics* 5, 557–581.

Hansen, L.P., Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy* 99 (2), 225–262.

Hansen, L.P., Sargent, T.J., Turmuhambetova, G.A., Williams, N. (2006). Robust control and model misspecification. *Journal of Economic Theory* 128 (1), 45–90.

Heath, D., Platen, E., Schweizer, M. (2001). A comparison of two quadratic approaches to hedging in incomplete markets. *Mathematical Finance* 11 (4), 385–413.

Hens, T. (1998). Incomplete markets. In: Kirman, A. (Ed.), *Elements of General Equilibrium Analysis*. Blackwell, Oxford, pp. 139–210. Chapter 5.

Herings, P.J.-J., Polemarchakis, H. (2005). Pareto improving price regulation when the asset market is incomplete. *Economic Theory* 25, 135–154.

Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6 (2), 327–343.

Hirshleifer, J. (1971). Private and social value of information and reward to inventive activity. *American Economic Review* 61 (4), 561–574.

Hodges, S.D., Neuberger, A. (1989). Optimal replication of contingent claims under transactions costs. *Review of Futures Markets* 8, 222–239.

Huang, P.H. (2000). A normative analysis of new financially engineered derivatives. *Southern California Law Review* 73 (471), 471–521.

Hugonnier, J., Kramkov, D. (2004). Optimal investment with random endowments in incomplete markets. *Annals of Applied Probability* 14 (2), 845–864.

Hugonnier, J., Kramkov, D., Schachermayer, W. (2005). On utility-based pricing of contingent claims in incomplete markets. *Mathematical Finance* 15 (2), 203–212.

Jarrow, R.A., Madan, D.B. (1999). Hedging contingent claims on semimartingales. *Finance and Stochastics* 3, 111–134.

Jarrow, R.A., Purnanandam, A. (2004). The valuation of a firm's investment opportunities: A reduced form credit risk perspective. *Working paper*.

Jarrow, R.A., Jin, X., Madan, D.B. (1999). The second fundamental theorem of asset pricing. *Mathematical Finance* 9 (3), 255–273.

Jaschke, S., Küchler, U. (2001). Coherent risk measures and good-deal bounds. *Finance and Stochastics* 5, 181–200.

Jouini, E. (2001). Arbitrage and control problems in finance: A presentation. *Journal of Mathematical Economics* 35, 167–183.

Kabanov, Y.M., Stricker, C. (2002). On the optimal portfolio for the exponential utility maximization: Remarks to the six-author paper. *Mathematical Finance* 12 (2), 125–134.

Kallsen, J. (2002a). Derivative pricing based on local utility maximization. *Finance and Stochastics* 6, 115–140.

Kallsen, J. (2002b). Utility-based derivative pricing in incomplete markets. In: Geman, H., Madan, D., Pliska, S., Vorst, T. (Eds.), *Mathematical Finance – Bachelier Congress 2000*. Springer-Verlag, Berlin, pp. 313–338.

Karatzas, I., Shreve, S.E. (1998). *Methods of Mathematical Finance*. *Applications of Mathematics*, vol. 39. Springer-Verlag, New York.

Karatzas, I., Žitković, G. (2003). Optimal consumption from investment and random endowment in incomplete semimartingale markets. *Annals of Probability* 31 (4), 1821–1858.

Kim, J., Kim, S.H., Levin, A. (2003). Patience, persistence, and welfare costs of incomplete markets in open economies. *Journal of International Economics* 61, 385–396.

Kou, S.G. (2002). A jump-diffusion model for option pricing. *Management Science* 48 (8), 1086–1101.

Kübler, F., Schmedders, K. (2001). Incomplete markets, transitory shocks, and welfare. *Review of Economic Dynamics* 4, 747–766.

Kübler, F., Chiappori, P.-A., Ekeland, I., Polemarchakis, H.M. (2002). The identification of preferences from equilibrium prices under uncertainty. *Journal of Economic Theory* 102, 403–420.

Larsen, K., Pirvu, T.A., Shreve, S.E., Tütüncü, R. (2005). Satisfying convex risk limits by trading. *Finance and Stochastics* 9 (2), 177–195.

Levine, D.K., Zame, W.R. (2002). Does market incompleteness matter? *Econometrica* 70 (5), 1805–1839.

Lim, A.E.B. (2004). Quadratic hedging and mean–variance portfolio selection with random parameters in an incomplete market. *Mathematical Methods of Operations Research* 29 (1), 132–161.

Liu, J., Pan, J., Wang, T. (2005). An equilibrium model of rare-event premia and its implication for option smirks. *Review of Financial Studies* 18 (1), 131–164.

Lo, A. (1987). Semiparametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics* 19 (2), 373–388.

Madan, D.B., Carr, P.P., Chang, E.C. (1998). The variance gamma process and option pricing. *European Finance Review* 2, 79–105.

Maenhout, P.J. (2004). Robust portfolio rules and asset pricing. *Review of Financial Studies* 17 (4), 951–983.

Magill, M., Quinzii, M. (1996). *Theory of Incomplete Markets, vol. 1*. MIT Press, Cambridge, MA.

Mania, M., Santacroce, M., Tevzadze, R. (2003). A semimartingale BSDE related to the minimal entropy martingale measure. *Finance and Stochastics* 7, 385–402.

Marin, J.M., Rahi, R. (2000). Information revelation and market incompleteness. *Review of Economic Studies* 67, 455–481.

Markowitz, H.M. (2002). Efficient portfolios, sparse matrices, and entities: A retrospective. *Operations Research* 50 (1), 154–160.

Mehra, R. (2003). The equity premium: Why is it a puzzle? *Financial Analysts Journal* 59 (1), 54–69.

Mehra, R., Prescott, E.C. (2003). The equity premium puzzle in retrospect. In: Constantinides, G., Harris, M., Stulz, R. (Eds.), *Handbook of the Economics of Finance*. Elsevier, Amsterdam, pp. 887–936. Chapter 14.

Musiela, M., Zariphopoulou, T. (2004a). An example of indifference prices under exponential preferences. *Finance and Stochastics* 8, 229–239.

Musiela, M., Zariphopoulou, T. (2004b). A valuation algorithm for indifference prices in incomplete markets. *Finance and Stochastics* 8, 399–414.

Mykland, P.A. (2003a). Financial options and statistical prediction intervals. *Annals of Statistics* 31, 1413–1438.

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

Perrakis, S., Ryan, P.J. (1984). Option pricing bounds in discrete time. *Journal of Finance* 39 (2), 519–525.

Pham, H. (2000). On quadratic hedging in continuous time. *Mathematical Methods of Operations Research* 51, 315–339.

Pietra, T. (2004). Sunspots, indeterminacy and Pareto inefficiency in economies with incomplete markets. *Economic Theory* 24, 687–699.

Prescott, E.C., Shell, K. (2002). Introduction to sunspots and lotteries. *Journal of Economic Theory* 107, 1–10.

Protter, P. (2006). A partial introduction to financial asset pricing theory. In: Birge, J.R., Linetsky, V. (Eds.), *Financial Engineering*. In: *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam.

Roorda, B., Schumacher, H. (2005). Time consistency conditions for acceptability measures, with an application to tail value at risk. *Working paper*.

Roorda, B., Schumacher, H., Engwerda, J. (2005). Coherent acceptability measures in multiperiod models. *Mathematical Finance* 15 (4), 589–612.

Rouge, R., El Karoui, N. (2000). Pricing via utility maximization and entropy. *Mathematical Finance* 10 (2), 259–276.

Routledge, B.R., Zin, S.E. (2004). Model uncertainty and liquidity. *Working paper*.

Ruszczyński, A., Shapiro, A. (2004). Optimization of convex risk functions. *Working paper*.

Saito, M. (1999). Dynamic allocation and pricing in incomplete markets: A survey. *Monetary and Economic Studies* 17 (1), 45–75.

Schachermayer, W. (2002). Optimal investment in incomplete financial markets. In: Geman, H., Madan, D., Pliska, S., Vorst, T. (Eds.), *Mathematical Finance – Bachelier Congress 2000*. Springer-Verlag, Berlin, pp. 427–462.

Schachermayer, W. (2003). A super-martingale property of the optimal portfolio process. *Finance and Stochastics* 7, 433–456.

Schied, A. (2004). On the Neyman–Pearson problem for law-invariant risk measures and robust utility functionals. *Annals of Applied Probability* 14 (3), 1398–1423.

Schweizer, M. (1995). Variance-optimal hedging in discrete time. *Mathematics of Operations Research* 20 (1), 1–32.

Schweizer, M. (1996). Approximation pricing and the variance-optimal martingale measure. *Annals of Probability* 24, 206–236.

Schweizer, M. (1999). A minimality property of the minimal martingale measure. *Statistics and Probability Letters* 42, 27–31.

Schweizer, M. (2001). A guided tour through quadratic hedging approaches. In: Jouini, E., Cvitanić, J., Musiela, M. (Eds.), *Option Pricing, Interest Rates and Risk Management*. In: *Handbooks in Mathematical Finance*. Cambridge University Press, Cambridge, pp. 538–574. Chapter 15.

Skiadas, C. (2003). Robust control and recursive utility. *Finance and Stochastics* 7, 475–489.

Skiadas, C. (2006). Dynamic portfolio theory. In: Birge, J.R., Linetsky, V. (Eds.), *Financial Engineering*. In: *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam.

Staum, J. (2004). Fundamental theorems of asset pricing for good deal bounds. *Mathematical Finance* 14 (2), 141–161.

Talay, D., Zheng, Z. (2002). Worst case model risk management. *Finance and Stochastics* 6, 517–537.

Uppal, R., Wang, T. (2003). Model misspecification and underdiversification. *Journal of Finance* 58 (6), 2465–2486.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, New York.

Zhang, L., Mykland, P.A., Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100 (472), 1394–1411.