COST-EFFECTIVE VARIANCE MINIMIZATION: 
A GENERAL STRATEGY FOR PRICING AND HEDGING 
IN INCOMPLETE MARKETS

TIMOTHY R. KLASSEN*
Department of Physics, Pupin Hall
Columbia University
New York, NYC 10027, USA

ABSTRACT

It is well-known that real-world stock price processes are far from the (log-) Brownian motion assumed in the Black-Scholes approach to option valuation. However, this and other assumptions, like continuous hedging, are crucial in constructing a risk-free replicating portfolio consisting of the stock and a bond. There is no consensus on how to price and hedge under more realistic assumptions about the price process, where risk cannot be completely eliminated (in so-called incomplete markets). One proposal is variance minimization: if we cannot completely eliminate the randomness in the cost of hedging, let us minimize it, and fix prices by a fair-game condition. Doubts have been cast on the consistency of this approach, since in some cases it leads to negative premiums on options with large strikes, an apparent absurdity. We show that such doubts are unfounded. It is not variance minimization that breaks down in such cases, but rather that the fair-game condition becomes inappropriate. However, a more detailed study reveals that variance minimization can suffer from a different problem: By minimizing the variance of not just losses but also (welcome!) gains, the fair-game price it leads to can be unreasonably high. We propose a simple strategy, cost-effective variance minimization, to better harness the hedging gains. It allows a market maker to sell the option at a lower price or make a larger profit. We illustrate this and various other aspects of pricing and hedging in incomplete markets with examples from the toy world of multinomial models. In particular, we discuss implied volatilities and their dependence on the underlying stock price distribution.

Keywords: option pricing, dynamic hedging, variance minimization, incomplete markets, statistical arbitrage, multinomial models, implied volatility

1. Introduction

The modern financial industry is based on the seminal work of Black and Scholes, and Merton [1] on option pricing. Their main insight was to realize that under certain assumptions it is possible to completely eliminate the risk of selling a stock option, despite the fact that the stock price evolves randomly. This is possible using dynamic hedging, where one buys or sells appropriate amounts of the stock (roughly speaking, the writer of the option has to buy when the stock goes up, sell when it goes down). Eliminating the risk for the seller of the option determines a rational price for the option via a fair-game condition. The point is that with
riskless hedging there would be *arbitrage* opportunities (the possibility of making a *guaranteed* profit by suitably buying or selling options and stock) unless the price is fixed so that neither side makes money in the deal.\footnote{As usual in such discussions, we ignore the small surcharge allowing the seller of the option to make a living.}

It was clear from the beginning that the Black-Scholes model is an idealization; none of its assumptions actually holds in the real world. Ignoring practical/technical issues that can be be “fixed” more or less easily, we concentrate on two points: First of all, one can only hedge at *discrete* times, not continuously (due to transaction costs, if nothing else). Secondly, and this is the main motivation for this paper, numerous studies have shown that stock prices and foreign exchange rates do not follow a (log-)Gaussian process, technically known as (log-)Brownian motion. A general finding is that the distribution of price changes is much more *leptokurtic* or *fat-tailed* than Gaussian-type distributions.

A further deviation from Brownian motion is that successive stock price changes are not iid (independent and identically distributed), in a rather subtle way: Empirical studies show that even though the auto-correlation function of price changes themselves decay extremely rapidly, their *magnitude* is (i) not constant, as phases of high and low variability alternate, and (ii) correlated over quite large time scales, often days or weeks, even years [2,3,4].\footnote{This is, strictly speaking, a violation of the *efficient market hypothesis*, which states that all relevant information about an asset at a given time is reflected in its price. In particular, price changes from one moment to the next should be independent. Note, however, that knowing a trend for the *magnitude* of price changes is, in general, not sufficient to allow for arbitrage opportunities.} In other words, successive price changes are (linearly) uncorrelated, but not independent.

In the language of finance, the variance of price changes is referred to as *volatility*; the fact that it is not constant (and seems to have a random component) in time is known as *stochastic volatility*; attempts to model its persistence properties come under acronyms such as ARCH and GARCH, to name just two of the simpler ones; see [5] for a review and references.

The problem of pricing and hedging options under real world conditions is a difficult one. Considering that trillions of dollars are traded each day in the financial markets, it is, however, imperative to continue tackling these issues. In particular, risk management requires realistic estimates of the probability of large losses. The problem falls into two parts:

- To determine an accurate model for the stock price process. Such a model can be purely phenomenological (since we are far away from a “microscopic” understanding of the price process there is not much of a choice in this matter at this point).

- Understand how to price and hedge options given such a model.

Concerning the first point, it has been known since the pioneering work of Mandelbrot [6] (and Fama [7]) that the fat-tailed price distributions observed in the real world can be described by so-called stable Levy processes, which have power-like
tails. Levy processes are generalizations of Gaussians, in that they emerge as limiting distributions for processes with infinite variance. They have definite scaling properties under convolution in accord with empirical price distributions for small and moderate time scales. Despite these attractive features, Levy processes did not gain long-term popularity in the financial community (although lately there has been renewed interest in these ideas, partially due to a stronger emphasis on proper risk management). One reason is presumably their infinite variance, which, among other things, is incompatible with option pricing in a Black-Scholes type framework. With the availability of large sets of high-frequency data it has become clear empirically that the far tails actually fall quicker than for standard Levy processes, so that the variance (and at least one more higher moment) is finite \[8,9\]. A model that incorporates all the empirical properties of the price distribution is provided by truncated Levy processes \[8,10,11\], which are Levy-like except in the far tails, where they have some kind of cut-off. These seem to be the only models that are compatible with the observed scaling of price distributions under change of time horizon (note, in particular, that all stochastic volatility models tested so far give incorrect scaling, cf. \[12,13\]).

The more subtle correlation properties of the price process have received detailed scrutiny only more recently \[4,14,3,15,16,9\]. Although no completely satisfactory model of all aspects of the price process is currently known, it would appear to be only a question of time before one will emerge. We think it will be necessary to use models that, one way or another, incorporate heavy-tailed distributions on short time scales, embedded in GARCH or stochastic volatility type models to reproduce the longer term correlation properties of the magnitude of price changes (and other details such as the leverage effect, the negative correlation of volatility and price increments). Very recently Mandelbrot, Fisher and Calvet proposed a new model of asset returns that incorporates both fat tails and long-term correlations. Their so-called multi-fractal model \[17\] looks promising but requires further study.

The empirical properties of the price process serve as motivation, but will not be investigated in this paper. Instead we would like to address the largely conceptual issues related to the second point raised above. Since any realistic model of the price process can not be expected to have particular, simplifying mathematical features, we should aim to understand pricing and hedging for basically arbitrary price processes. There is no generally accepted principle of how to do so. In fact, a significant part of the beautiful mathematical framework (Ito calculus, equivalent Martingale measures, etc) developed in the wake of the Black-Scholes model has Brownian motion and risk-free hedging hardwired into it. This is perhaps part of the reason that the two points above have received less attention than they deserve, in our opinion. Another reason might be the adherence to the concept of arbitrage as a strictly riskless profit opportunity. In the real world such a concept is somewhat academic; it is necessary to allow for risk, which not just a potential arbitrageur, but all market players have to weigh against the expected profit.

Given the well-known limitations of the Black-Scholes model, how do market
makers determine the price and hedging strategies for their products? To outline the answer it is useful to recall the most straightforward empirical proof that the Black-Scholes model does not reflect reality. In this model the price process is (log- ) Brownian motion of constant variance. As it turns out — and this is not true in general — the average return of the asset underlying an option drops out of the Black-Scholes price. Besides the observable parameters relevant for an option (exercise price, initial stock value, maturity, risk-free interest rate, etc) the Black-Scholes price only depends on the volatility (which is only indirectly observable). One can therefore use the Black-Scholes formula in reverse and obtain for each maturity and exercise price an estimated volatility from the observed prices of traded options. If the Black-Scholes model were correct, all such estimates should agree for a given underlying asset. Instead, one finds that the “implied volatility” depends on both maturity and exercise price. This is known as the “term structure” and “smile”, respectively, of volatility. The latter term refers to the roughly smile-like form of a plot of implied volatility versus exercise price (although after the '87 crash it often looks more like a smirk).

In practice the following procedure is used. The prices for liquid vanilla are taken from the market and used to obtain the maturity and exercise price dependence of the implied volatility. It, in turn, is then used, perhaps after some interpolation or smoothing, with less liquid and/or more exotic options.

However, even if we assume that one is using the “right” price in this approach (perhaps not, see below), there is still the question of hedging. There is not much point in knowing the “right” price of an option, if one does not know how to hedge it properly. Most of the literature on alternative option valuation theories focuses on pricing. The delta hedging strategy which can be derived for Gaussian processes is often also used for such alternative models, where it is purely ad hoc, and, in particular, not risk-free. Presumably one will usually not lose a lot of money this way, since the market price will reflect any additional loss one might otherwise incur due to the incorrect hedging (since everybody uses roughly the same, slightly wrong strategy?). However, one might seriously misjudge the risk of exceptional, large losses, especially for exotic or new options. Furthermore, a market maker with a proper understanding of the price process and hedging might be able to offer the same product at a lower price while assuming no more or even less risk.

There are various attempts to avoid the problem of incomplete markets. One is to assume that risk can be diversified away. In view of the experience of traders that the volatilities of different assets tend to move together, this assumption seems implausible. Other studies assume the price process to have a definite and known number of sources of randomness (usually Brownian motion). In some but certainly not all [18] cases the market can then be completed by including more than one instrument in the hedge. Although using several hedging instruments can certainly be useful in various situations, we are far away from a “microscopic” understanding of financial price processes, so that such studies appear rather optimistic. Another approach amounts to making some indirect and rather ad hoc assumptions about
the market price of risk, allowing one to obtain a Martingale measure for pricing. However, the question of hedging is often left open (or requires further assumptions).

We think the right approach is to face the problem of risk head-on. Instead of living with the fiction that risk does not exist, one should study what the random cost of hedging is in realistic markets. This is the attitude taken by proponents of the *variance minimization* strategy, where, as the name suggests, the idea is to minimize the variance of the random cost of hedging. Given the resulting hedging strategy, the price can then be determined by a fair-game condition. This approach was first suggested in the mathematical finance literature by Föllmer, Sondermann, Schweizer and others [19,20,21]. Independently, and in a slightly different form, these ideas were developed by the physicists Bouchaud and Sornette [22,23], whose work was taken up by a number of authors [24,25,26,27,28,29]. Very recently the two schools have begun to converge, cf. [30]. Variance minimization is a natural generalization of the standard approach taken in complete markets, to which it reduces in the case where risk can be completely eliminated.

Although seemingly a perfectly sensible approach, some doubts have been raised as to its consistency (cf. [24,25,26,27,28,31]). Namely, it turns out that in some cases option prices deduced by this method are negative for large strikes. This indeed is an apparent absurdity. However, it is not an indication that the variance minimization approach breaks down (instead, we will see, it is the fair-game condition that becomes inappropriate). Although this is not a fundamental problem, other examples do reveal limitations of the variance minimization approach. The basic problem is that variance is not truly a measure of risk, as it also includes welcome random *gains*, in addition to the losses one would like to minimize.

The outline of this paper is as follows. In sect. 2 we review variance minimization in the incomplete market provided by the discrete-time, discrete-state world of *multinomial models*, deriving explicit formulas for the cost and hedging of European call options. We also discuss some subtleties related to *local* versus *global* variance minimization. Sect. 3 addresses the issue of negative premiums with a number of clarifying examples. In sect. 4 we expand on the limitations of variance minimization that were partially revealed in sect. 3. We present a simple strategy, which we dub *cost-effective variance minimization* (CVM), that alleviates these problems. More precisely, CVM presents a family of hedging and pricing strategies, which are designed to map out the “efficient frontier” of the premium versus risk choice that the seller of an option faces in an incomplete market. Sect. 5 is devoted to the study of further conceptual and practical questions in the context of multinomial examples. In particular, we discuss how much a mistake we make by assuming the world to be risk-free, and hedge accordingly, when the actual price process does not allow for a complete elimination of risk. We study this in a number of examples with varying amounts of skew and kurtosis in the true price distribution. We also investigate the effect of these and other variables on the “smile” of the implied volatility. We summarize and conclude with an outlook on future directions in sect. 6.
2. Variance Minimization in a Discrete World

2.1. The setup

The issues we are concerned with arise for any option, so to be concrete we will consider the simplest of them all, the European call option for a single asset, referred to as stock in the following. Such an option gives the buyer the right, but not the obligation, to buy at some predetermined future time, the maturity date \( T \), one share (say) of the stock at the exercise or strike price \( K \). The initial time will always be \( t = 0 \), and the stock price at time \( t \) will be denoted by \( S_t \). The buyer will only exercise his option if \( S_t > K \), so the payoff of this option is \((S_T - K)_+ \equiv \max(0, S_T - K)\) at time \( t = T \). We assume a frictionless market, i.e. we neglect transaction costs and taxes.

As the buyer has a right and no obligation (and no risk), whereas the seller has an obligation and a risk, this option has a certain cost (or premium or price), \( C_0 \), the seller will charge up front. The question we want to address is: What is a fair price, and how should the seller hedge to minimize potential losses from adverse movements of the stock?

Our playground for these questions will consist of the multinomial models, where we assume that hedging can only take place at discrete times, \( t = 0, 1, 2, \ldots \) and the stock price change from time \( t \) to \( t + 1 \) can assume one of \( N \) possible values \( u_i > 0 \) with probability \( p_{ti} \):

\[
S_{t+1} = u_t S_t \quad \text{with probability} \quad p_{ti}, \quad i = 1, \ldots, N. \tag{2.1}
\]

We will often assemble the \( u_t \) and \( p_t \) into vectors \( \hat{u}_t, \hat{p}_t \). Here and elsewhere, when thinking of a quantity as a random variable, instead of a concrete realization, it will be ‘hatted’. The above equation becomes \( \hat{S}_{t+1} = \hat{u}_t \hat{S}_t \), for example.

To complete the setup we assume the existence of a riskless bond into which the seller can deposit the initial premium or from which he can borrow money to finance his hedge. We take the rate of return of the bond to be deterministic and given by the multiplicative factor \( r_t \) between times \( t \) and \( t + 1 \). For simplicity we assume that the distribution of the \( \hat{u}_t \) (and \( r_t \)) is independent of \( t \) and will often drop it as a subscript. In most formulas below the more general case is obtained by simply adding suitable subscripts to various quantities.

Before proceeding we should comment on our use of multinomial models throughout this paper. First of all, this work should be regarded as preparatory for later studies with more realistic models for the price process \[32\], to which the methods discussed can be generalized. Our aim here is to discuss the conceptual problems arising in incomplete markets in the simplest possible and explicit manner. We should also remark, though, that discrete models are not as unrealistic as they might appear at first sight. Their discrete-time nature is actually realistic, since hedging can only be performed discretely, of course. The discrete nature of the state space could in principle be overcome by making \( N \) sufficiently large. (Of course,
this would not be efficient from a numerical point of view; though, if \( N \) has many prime factors, FFT techniques can be used, cf. [33].) Finally, it is amusing to note that in at least one case, the Warsaw stock exchange, a trinomial model provides an excellent representation of the price evolution, due to regulations restricting the relative price changes [25].

### 2.2. The problem: an example

A nice, simple example to illustrate the issues arising when trying to price and hedge options in incomplete markets was presented in [34], which we will borrow.

Consider the trinomial model, \( N = 3 \), with the stock price process defined by \( u = (1.4, 0.9, 0.6) \) and \( p = (0.8, 0.1, 0.1) \). Assume that we want to sell a European call option with maturity \( T = 1 \) and strike price \( K = S_0 = $100 \), cf. figure 1. We set \( r = 1 \) (we can always use discounted prices, so this is no loss of generality). Our hedging strategy will be to buy \( \phi \) shares of the stock per option at time \( t = 0 \). At time \( t = T = 1 \), we then face a cost of $\$(40 - 40\phi, 10\phi, 40\phi) \$ for the three values that the stock can take at that time. This cost is the sum of the hedging losses (or gains) and the payoff if the stock ends up above $100. It is easy to see that we will never have to face a cost of more than $20 if we choose \( \phi = 0.5 \). In other words, $20 is an upper bound for the premium we could charge for this option; if we charged more we could make a riskless profit by choosing \( \phi = 0.5 \).

On the other hand, consider the point of view of a buyer of such an option. He can “go short”\(^3\) on the stock, by \( \phi' \) shares per option, so that his wealth changes by $\$(40 - 40\phi', 10\phi', 40\phi') \$ between time 0 and 1. By choosing \( \phi' = 0.8 \) he can arrange to gain at least $8 in all cases. Hence, if he paid less than $8 for the option he could make a riskless profit.

What the above shows is that arbitrage considerations — with arbitrage defined

\(^3\)Going short on a stock refers to the following procedure, for which the financial markets provide a mechanism using a broker as intermediary: We sell stock, not currently in our possession, at today’s price and buy it back at some later time at the price for which it is traded at that time. Formally it corresponds to \( \phi < 0 \). We make a profit if the stock falls; we loose if it goes up.
as a strictly riskless profit opportunity — only restrict the price of the option to lie in the range $8 – $20. This is a rather wide range.

Clearly, it should be possible to narrow this down to a more reasonable, much smaller range for the price. Just as clearly, this will require us to go beyond the strict definition of arbitrage and explicitly weigh profit and/or market share (coming from the ability to offer a low price for the option) versus non-zero risk. Since in the real world nothing is riskless, we think this step should be taken in any case. This is a point which we will emphasize repeatedly.

It is also clear that it will not be possible to narrow down a fair price for this option with infinite precision. When it comes to fine-tuning the price, the risk preferences of buyers and sellers, specific market conditions like liquidity, etc, will come into play. Nevertheless, from the sellers point of view, say (which we will usually take), it should be possible to design hedging strategies that strike a good balance between minimizing risk and allowing him to offer the option at a competitive price. In fact, one might wish to map out an “efficient frontier” of risk versus premium.

At least for the discrete world we are considering here, we are aware of three competitive proposals for pricing and hedging options in incomplete markets (another approach, super-replication [35] seems to be too expensive to be practical). These are the variance minimization approach, the risk minimization approach of Sommer [34] (which is specific to the trinomial model) and a hedging strategy [31] based upon the ideas of Kelly [36], otherwise used for maximizing the average growth rate of a portfolio. We will come back to all three approaches and compare them with our proposal. We start with the well-known (local) variance minimization strategy, as it provides a starting point for the framework we will develop in sect. 4.

2.3. The variance minimal hedge and price

We would like to determine the optimal hedge that will minimize the variance of the random cost the seller will face. To be precise, we would like to minimize the global variance over the complete hedging history from 0 to $T$. In general a global minimum does not exist [37], and even if it does, the corresponding hedging strategy can be very complicated. Therefore we will follow common practice and adopt a much simpler local variance minimization strategy. We will come back to the difference between local and global minimization later. As in [26,27,28] we will assume the following: We allow options to be continuously bought and sold before they reach maturity. This means, in particular, that options have well-defined values $C_t$ for intermediate times, $0 \leq t \leq T$ (given $S_t$). In accord with this, we also assume that the hedge ratio $\phi_t$ (see below) only depends on $S_t$, not on any of the earlier stock prices. In a loose sense, we consider only “Markovian” hedging strategies. We shall return to the effects of this restriction later.

We know that at maturity $C_T = (S_T - K)_+$ is the payoff, and would like to derive the premium $C_0$. We do so recursively, starting at maturity and working our way backwards. Let us consider the wealth balance of the seller between $T - 1$ and $T$. At time $t = T - 1$ the option has a definite (yet to be determined) value $C_{T-1}$,
and the stock is at the known value $S_{T-1}$. The random wealth balance is

$$\Delta \hat{W}_{T-1} = rC_{T-1} - \hat{C}_T + \phi_{T-1} S_{T-1} (\hat{u}_{T-1} - r).$$  \hspace{1cm} (2.2)

Here $\phi_{T-1}$ is the hedge ratio, the (to be determined) amount of the stock to hold in the time interval $[T-1, T]$. The first term in the above is the premium received at $T-1$, in units of the money’s worth at time $T$. The next represents the random payoff the seller will face at maturity; the last two terms are the difference between what the seller will gain (or loose) by borrowing money to buy stock versus leaving it in the bank.

The variance of the wealth balance is

$$R^2_{T-1} \equiv \langle \Delta \hat{W}_{T-1}, \Delta \hat{W}_{T-1} | S_{T-1} \rangle$$

$$= R^2_{T-1} (\phi_{T-1} \equiv 0) - 2 \phi_{T-1} S_{T-1} \langle \langle \hat{C}_T (\hat{u}_{T-1} - r) | S_{T-1} \rangle \rangle + \phi^2_{T-1} S^2_{T-1} \sigma^2$$  \hspace{1cm} (2.3)

where we introduced the volatility $\sigma^2$ of the stock,

$$\sigma^2 := \langle \langle \hat{u}_t - r, \hat{u}_t - r \rangle \rangle.$$  \hspace{1cm} (2.4)

The extremum condition for the hedge, $\partial R^2_{T-1} / \partial \phi_{T-1} = 0$, gives

$$\phi_{T-1} = \frac{1}{\sigma^2 S_{T-1}} \langle \langle \hat{C}_T (\hat{u}_{T-1} - r) | S_{T-1} \rangle \rangle = \frac{1}{\sigma^2 S_{T-1}} \langle \langle \hat{C}_T (\hat{u}_{T-1} - \langle \hat{u} \rangle) | S_{T-1} \rangle \rangle.$$  \hspace{1cm} (2.5)

The fair-game condition $\langle \Delta \hat{W}_{T-1} | S_{T-1} \rangle = 0$ implies

$$r C_{T-1} = \langle \hat{C}_T | S_{T-1} \rangle - \phi_{T-1} S_{T-1} \mu,$$  \hspace{1cm} (2.6)

where

$$\mu := \langle \hat{u}_t - r \rangle$$  \hspace{1cm} (2.7)

is the mean excess rate of return of the stock over the bond (to be sure, despite the terminology we do allow $\mu < 0$ too, of course). Using the optimal hedge ratio finally gives the price as

$$r C_{T-1} = \langle \hat{C}_T \hat{\mu}_{T-1} | S_{T-1} \rangle,$$  \hspace{1cm} (2.8)

with

$$\hat{\mu} := \frac{1}{\sigma^2} \langle \langle \hat{u}_t - \langle \hat{u} \rangle \rangle \rangle.$$  \hspace{1cm} (2.9)

Eq. (2.8) gives the price at $T - 1$ contingent on the value of $S_{T-1}$. In the obvious manner we can therefore promote $C_{T-1}$ to a random variable, which, by a slight
abuse of notation, we will denote as \( \hat{C}_{T-1} \). Considering now the interval \( T - 2 \) to \( T - 1 \) we are exactly in the same situation as before and obtain

\[
r^2 C_{T-2} = \langle r \hat{C}_{T-1} \hat{p}_{T-2} | S_{T-2} \rangle = \langle \hat{C}_T \hat{p}_{T-1} \hat{p}_{T-2} | S_{T-2} \rangle.
\]  

(2.10)

Note that the first average in the above is over all values of \( \hat{u}_{T-2} \), whereas the second is over all values of \( \hat{u}_{T-2} \) and \( \hat{u}_{T-1} \). It is now clear that for general \( t \), including \( t = 0 \), we can write

\[
r^{T-t} C_t = \langle r \hat{C}_{t+1} \hat{p}_t | S_t \rangle = \langle \hat{C}_T \hat{p}_{T-1} \hat{p}_{T-2} \cdots \hat{p}_t | S_t \rangle
\]  

(2.11)

for the premium and

\[
\phi_t = \frac{1}{\sigma^2 S_t} \langle \hat{C}_{t+1} (\hat{u}_t - \langle \hat{u} \rangle) | S_t \rangle = \frac{r^{-\left(T-t-1\right)}}{\sigma^2 S_t} \langle \hat{C}_T \hat{p}_{T-1} \hat{p}_{T-2} \cdots \hat{p}_{t+1} (\hat{u}_t - \langle \hat{u} \rangle) | S_t \rangle
\]  

(2.12)

for the optimal hedge ratio. This concludes the derivation of the (locally) variance-minimal hedging strategy and the corresponding fair-game price. Note that the explicit form of the European call payoff was nowhere used; all we need is that it depends only on the final stock price.

We now comment on the form of the solution for the optimal hedge and price. First of all, note that the forward price \( r^{T-t} C_t \) (in units of the money’s worth at time \( T \)) can be written as the expectation value of the payoff in terms of a new “pricing measure” \( q \),

\[
r^{T-t} C_t = \langle \hat{C}_T | S_t \rangle_q.
\]  

(2.13)

Here \( \langle \ldots \rangle_q \) denotes an expectation value in this measure, which is defined by

\[
q_{ti} := p_{ti} \rho_{ti} = p_{ti} \left[ 1 - \frac{\mu}{\sigma^2} (u_{ti} - \langle \hat{u} \rangle) \right]
\]  

(2.14)

and the same \( u_{ti} \) as for the original, or “historical”, measure. The above defines the measure only for one time slice, but in equations like (2.13) should be extended to paths in the obvious manner (by multiplication).

Note that indeed \( \langle 1 \rangle_q = 1 \), as required for a probability. Eq. (2.13) means that the discounted option price process is a martingale with respect to the (signed) measure \( q \) (modulo the technicality that the definition of a martingale refers to a filtration, a set of increasing \( \sigma \)-algebras; but with our assumptions we can always replace \( S_t \) by the full \( \sigma \)-algebra of “information available at time \( t \)” generated by \( S_t, S_{t-1}, \ldots, S_0 \)). We also have \( \langle u \rangle_q = r \), an analog of a well-known property of the Black-Scholes model, and the reason \( q \) is often referred to as the “risk-neutral” measure in that context.

Note that we can rewrite the hedge ratio at time \( t \) as

\[
\sigma^2 S_t \phi_t (S_t) = \sum_{i=1}^N p_{ti} (u_{ti} - \langle \hat{u} \rangle) \langle \hat{C}_T | S_{t+1} = u_{ti} S_t \rangle_q r^{-\left(T-t-1\right)}
\]
\[ = \sum_{i=1}^{N} p_{ti} (u_{ti} - \langle u \rangle) \, C_{t+1}(S_{t+1} = u_{ti} S_i) , \]  
(2.15)

where we have explicitly indicated the dependence on the stock price in \( C_{t+1} \) and \( \phi_t \).

Note that the case \( \mu = 0 \), where the stock has the same mean return as the bond, is special. In this case \( \rho \equiv 1 \), so that the pricing measure \( q \) is identical to the historical measure \( p \).

To complete the theoretical description of our toy world, let us write down the value of a European call option more explicitly for a multinomial model. Since the value of such an option is path-independent, in that the payoff depends only on the final value of the stock, we can reduce the \( N^T \) paths to a much smaller number of terms to sum over, since many give the same contribution (here we explicitly assume the \( \hat{u}_t, \hat{p}_t \) to be iid). Namely, let \( n_i \geq 0, i = 1, \ldots, N \) be "occupation numbers", \( n_i \) representing the number of times that a stock price change by a factor of \( u_i \) occurs. We then have

\[ r^T C_0(S_0, K, T) = \sum_{n_1 + \ldots + n_N = T} \frac{T!}{n_1! \ldots n_N!} q_1^{n_1} \ldots q_N^{n_N} (u_1^{n_1} \ldots u_N^{n_N} S_0 - K)_+ \]  
(2.16)

It is easy to see that the number of terms in this sum is given by the binomial \((T+N-1)\) which grows only like a power \( T^{N-1} \) for large \( T \), instead of exponentially fast.

Note that the special case \( N = 2 \), the binomial (or Cox-Ross-Rubinstein [38]) model, is non-generic in several respects. Using \( \sigma^2 = p_1 p_2 (u_1 - u_2)^2 \) one finds that the pricing measure is given by

\[ q_i = \begin{cases} \frac{r-u_0}{u_1-u_0} & i = 1 \\ \frac{u_1-r}{u_1-u_2} & i = 2 , \end{cases} \]  
(2.17)

which is independent of the initial probabilities \( (p_1, p_2 = 1-p_1) \). This is not true for higher \( N \). Also, since there are only two possible values the stock can move to, and the amount we loose or gain during hedging is linear in the hedge ratio \( \phi_t \) (with different coefficient for the two cases), we can always adjust \( \phi_t \) to loose or gain exactly the same amount in the two cases. In other words, hedging risk can be completely eliminated. This argument also makes clear why the probabilities \( p_i \) drop out of the picture. For \( N > 2 \), the same argument shows that risk and the dependence on the \( p_i \) can not be eliminated, generically.

2.4. Subtleties

As emphasized earlier, the hedging strategies we allowed in the above variance minimization are not the most general ones. In [27] it was conjectured that they, nevertheless, lead not just to a local, but even the global minimum of the variance of the wealth balance. The explicit proposal for the variance-minimal measure (2.14)
given in [27] was supported by some numerical checks [26]. A claim of a proof of the conjecture is to be found in [28].

Alas, it was very recently pointed out by Schweizer that the proof is incorrect and the conjecture wrong in general [30]. He gave an explicit example, $p = (0.25, 0.25, 0.25, 0.25)$, $u = (4, 2, 1, 0.5)$ with $T = 3$ in our notation, where the conjecture fails. This example has $q_i < 0$, which will lead to negative premiums for large strikes, as discussed in detail in sect. 3. However, this is not crucial, examples with all $q_i > 0$ can easily be found. In fact, Schweizer argued that the conjecture must be expected to fail generically when $N > T > 2$, and we have confirmed this numerically.

To clarify the difference between the local variance minimization performed in sect. 2.3 and the most general global one, we remark the following. The most general hedging strategy in the present context can be labelled by the $1 + N + \ldots + N^{T-1}$ numbers $\phi_0, \phi_1(S_1), \ldots, \phi_{T-1}(S_{T-1})$. We stress that $S_t$ here is meant to be a label for the $N^t$ numbers $\phi_t(S_t)$; we do not mean that as a function $\phi_t$ only depends on $S_t$. In other words, no Markovian or any other assumption on the functional dependence of $\phi_t$ is made; we consider the $\phi_t(S_t)$ simply as variational parameters.

On the other hand, for the minimization performed in sect. 2.3 we have, effectively, only one (not $N^t$) variational parameter for each time interval. This is despite the fact that the $\phi_t(S_t)$ will numerically be different for different $S_t$. They are not all independent variational parameters; there is really only one degree of freedom involved in the minimization over a given time interval $[t, t + 1]$. It is important to understand this difference, as it is easily overlooked on a first encounter with the approach of sect. 2.3.

It now may come as a surprise that, as shown by Schäl [21] and Schweizer [20,30], the truly variance-minimal fair-game price is nevertheless given by eq. (2.13) with the measure $q$ defined in (2.14). The minimal variance will however not be attained with the hedge (2.12), in general.

For multinomial, and somewhat more general models Schäl and Schweizer have given explicit formulas for the variance-minimal hedging strategy. Due to their recursive character they are somewhat cumbersome to evaluate, even in relatively simple cases. In realistic situations the cost of evaluating them might make the global variance minimization approach impractical. More importantly, since options are usually traded freely during their lifetime, hedging strategies depending on the complete history of a stock appear incongruous with market practice.

It is, nevertheless, important and interesting to study how big the difference between the global and local minimal variances is. Even in rather extreme examples, like that of Schweizer mentioned above, we have found that the difference is relatively small. The issue deserves further study, but is beyond the scope of this work, as the main point of this paper is to argue that a variance minimization strategy, either local or global, is not always the best way to proceed. We will argue that the

---

5Extreme in that $\mu = O(1)$, for example. Note that all the subtleties we are talking about disappear as $\mu \to 0$, cf. [21,30].
variance-minimal price can be too large, and propose hedging strategies that allow one to lower the price without a significant increase in risk. As the local and global variance-minimal fair-game price are identical, this issue arises for either approach; cf. sect. 4.

3. The Signed Measure \( q \) and Negative Premiums

We have seen that the value of a European call option can be written as an expectation value with respect to a new pricing measure \( q \), which is expressed in terms of the parameters defining the historical measure \( p \) and the risk-free interest rate \( r \).

Actually, \( q \) does not in general define a measure, but only a \emph{pseudo-measure} (or \emph{signed} measure) that can take negative values. To see this, note that we can write the factor \( p_i \) by which \( q_i \) and \( p_i \) differ as

\[
\rho_i = \frac{1}{r} \left( \langle \tilde{u}_i^2 \rangle - \langle \tilde{u} \rangle \tilde{u}_i \right),
\]

in terms of the variables \( \tilde{u}_i := u_i - r \). Note that the above quantity is positive \emph{on average}. The question is if it can become negative in some cases for some \( i \):

\[
\langle \tilde{u}_i^2 \rangle - \langle \tilde{u} \rangle \tilde{u}_i < 0 \quad \text{for some } i?
\]

To investigate this let us order the \( u_i \) so that \( u_1 > u_2 > \ldots > u_N \) (without loss of generality we can assume that the \( u_i \) are distinct; we can also assume that all \( p_i > 0 \)). Note that \( \rho_i \) is invariant under the reflection \( \tilde{u}_i \rightarrow -\tilde{u}_i \). For each “bullish” case with some \( \rho_i < 0 \) reflection gives a “bearish” case where the same is true, and vice versa. We will always consider the “bullish” case, say, and assume that half or more of the \( \tilde{u}_i \) are non-negative.

Amusingly, \( \rho_1 < 0 \) can even happen in the binomial model. This is easily seen to be true if and only if both \( \tilde{u}_1, \tilde{u}_2 \) are strictly positive (or in the reflected case when both are strictly negative). This can lead to negative prices at large strikes, especially for large times to maturity. An example is shown in figure 2.

Now, this example might be considered irrelevant, since here the stock price \emph{always} goes up, i.e. there is arbitrage between the stock and the bond. Such a situation — or the corresponding bearish one, where the stock always goes down — could at best hold for short periods. Nevertheless, this situation in the binomial model can be continuously connected to one in the trinomial model, where no strict arbitrage is possible: All we have to do is to add a third possible value for the stock price in the next time interval with a small probability and \( \tilde{u}_3 < 0 \). By continuity it is clear that negative premiums will still occur.

The fact that models where arbitrage is possible can be continuously transformed into ones where no strict arbitrage (but only \emph{statistical} arbitrage) is possible, reveals, in our mind, another facet of the limitations of the concept of strict arbitrage: In addition to being artificial from a practical point of view, it is also \emph{not stable} under small deformations.
Returning to the condition (3.19), consider the trinomial model. There are two ways in which \( q_i < 0 \) can occur. First, there is the "trivial" case where \( \tilde{u}_1 > \tilde{u}_2 > \tilde{u}_3 \geq 0 \). It is the analog of the binomial example above. In this case \( q_1 < 0 \) will always hold. \( q_2 < 0 \) can also occur, namely iff
\[
p_2 \tilde{u}_2 (\tilde{u}_2 - \tilde{u}_3) > p_1 \tilde{u}_1 (\tilde{u}_1 - \tilde{u}_2).
\]

The more interesting, second case is for \( \tilde{u}_3 < 0 \) (remember that in the "bullish" case we are considering we then have \( \tilde{u}_1 > \tilde{u}_2 \geq 0 \)). Now at most \( q_1 < 0 \) is possible, iff
\[
p_2 \tilde{u}_2 (\tilde{u}_1 - \tilde{u}_2) > p_3 [\tilde{u}_3 (\tilde{u}_1 - \tilde{u}_3)].
\]
Fair-game premiums from variance minimization are shown in figure 3 for such a case.

For higher \( N \) the analysis becomes more tedious and not very illuminating. Save to say that it always possible to find \( q_i < 0 \), but only under quite unrealistic conditions on the relation between excess return in state \( i \) and the variance: If we write \( \mu = \alpha \sigma \), then (3.19) requires
\[
u_i - r > (a + 1/a) \sigma \quad \text{for} \quad \mu > 0 \quad \text{(the sign of the lhs is reversed if} \quad \mu < 0). \]
Remember that in practice usually \( |\mu| \ll \sigma \), at least for short time scales.

From these considerations we conclude the following. It is not the idea of variance minimization that is breaking down and leads to a signed pricing measure \( q \). Rather, the market is so bullish or bearish — in a specific manner, not any bullish or bearish market with \( \mu \neq 0 \) will do, as we have seen — that the seller of an option will make money during the variance-minimal hedge. Enforcing a fair-game condition under such conditions inevitably leads to negative premiums. If such market conditions were ever to occur, nobody would adhere to the fair-game condition; there would be enough money to go around for everyone. Of course, it also means

*Unless otherwise indicated, the phrase 'variance minimization' here and elsewhere in the rest of the paper refers to minimization over the restricted set of hedging strategies considered in sect. 2.3.*
that writing options is not the most profitable business to be in then; one should simply invest all one's money in the stock market!

Note that in practice model parameters extracted from observed data have considerable uncertainty. It is easy to see that in situations where negative premiums can occur the dependence of premiums on strike price and model parameters is quite erratic (this is already suggested by figures 2 and 3). Therefore, any reasonably prudent strategy of taking the parameter uncertainty into account might well be sufficient, in practice, to wipe out negative premiums, even if the seller were otherwise willing to charge a variance-minimal fair-game price.

In preparation for the next section, let us point out one aspect of figures 2 and 3 that is still perhaps a bit puzzling. Namely, how can the premium, after being negative for some strike $K$, suddenly be strictly positive again for some larger $K$? It turns out that in such cases one could also charge a negative fair-game price — and in other cases where the premium is already negative one could charge an even more negative price — if one were to choose a better hedging strategy. We must conclude that variance minimization is in some cases simply pathological. We remark, for example, that some of the hedge ratios $\phi_t$ leading to the results in figures 2 and 3 are negative. Given the market conditions this is not a sensible hedging strategy; it simply arises from a slavish adherence to the dictate of variance minimization. The exploitation of this insight will be the subject of the next section.

4. Cost-Effective Variance Minimization

4.1. The problem
When discussing the issue of negative premiums in the previous section we noticed that the variance-minimal fair-game premiums were too large, in some cases even the negative ones. To show that this has nothing to do with the exceptional cases where negative premiums can occur, let us consider a simple binomial example, where negative prices can not occur with riskless hedging. Namely, let \( u = (1.4, 0.9) \), with \( p = (0.9, 0.1) \). Consider a European call option with \( K = S_0 = \$100 \), \( r = 1 \), and \( T = 1 \). If we sell such an option, the riskless strategy is to choose the hedge ratio \( \phi_0 = 0.8 \). Combining hedging and the payoff at the end we loose \$8, independent of whether the stock goes up or down (and independent of \( p \)). We should therefore charge a premium of \$8 for this option.

On the other hand, let us consider the hedging strategy \( \phi_0 = 1 \). Considering hedging and the payoff we loose nothing if the stock goes up, and \$10 if it goes down. On average — and \( p \) matters here now — we loose \$1. With this strategy, we would not loose money, in the long run, if we just charged \$1 for the option! Of course, we incure some risk for which we would like to be compensated. More to the point, according to conventional wisdom, we should prevent arbitrage opportunities, which requires us to also charge \$8 for the option (if we would charge less, a buyer of our option could set up a riskless profit scheme by performing the hedge of the previous paragraph in reverse, i.e. short-selling the stock, \( \phi_0 = -0.8 \)). \(^7\) We are happy to charge \$8, of course. We will very quickly become rich, as we make \$7 on each option we sell, on average. Anyone, except the most risk-paranoid, would regard this as a more than sufficient compensation for the small risk we are assuming in this strategy (occasionally loosing \$2 on a sale).

4.2. The remedy: Cost-effective variance minimization

The simple example above contains the germ of much that is to follow in this section. The main difference will be that there is now absolutely no reason to charge the analog of (as much as) \$8 in the above example, as no opportunities for riskless profit arise in an incomplete market when we charge somewhat less.

Our example shows that trying to minimize the variance is sometimes perhaps not such a good idea, as the equalization of the combined cost of hedging and payoff it attempts can lead to much larger losses, on average, than allowing some risk. After adding the fair-game premium received up front, we see that the situation can be described as follows. If our price is a fair-game price, the variance of the wealth balance has contributions from both losses and gains. Minimizing the variance will suppress not just losses but also gains. This will force us, generically, to charge a higher premium than if we could somehow harness hedging gains.

One could attempt to directly minimize just the contribution from losses to the

\(^7\)Actually, we think one could argue that it does not matter if someone else derives a riskless profit opportunity from our actions, as long as we also make a sufficient amount of money (if we charge somewhat less than \$8, we can always arrange to make much more money, on average, than the potential arbitrageur). At least if we are a small player in the market we see no logical problem with this approach, in particular, if we are perhaps the only player who has performed a careful risk versus profit analysis.
variance. This is a measure of what might properly be called risk, and therefore sounds attractive. However, this approach does not lead to a simple and elegant formalism. Even in simple cases one obtains only implicit equations for the optimal hedge ratios, which must be solved numerically.

A simpler approach is obtained by remembering the ultimate goal: We would like to be able to sell the option at a lower price, without a large increase in risk, if possible. We therefore consider minimizing a linear combination of the value of the option (squared) and local variance:

\[
\text{Minimize } \lambda (r C_t)^2 + R_t^2. \tag{4.20}
\]

For \( \lambda > 0 \) this will have the effect of suppressing the cost, at the expense of some increase in the variance (paradoxically at first sight, but in some cases the variance will also decrease, cf. the end of this section). It is easy to see that this approach is equivalent to minimizing the variance for fixed cost and then minimizing \( \lambda (r C_t)^2 + R_t^2 (C_t) \) with respect to the cost.

Note that the desired effect will not occur in situations where the fair-game price is negative. However, as discussed earlier, negative premiums occur only in quite unrealistic situations. Our interest in such cases arose solely from the worry that they might indicate a fundamental, conceptual problem with variance minimizations. Having shown, in sect. 3, that this is not the case, we feel free to ignore such cases whenever convenient from now on.

As in the pure variance minimization approach of sect. 2, corresponding to \( \lambda = 0 \), we minimize locally, starting at maturity, and work our way backwards. Using the fair-game constraint to eliminate the premium \( C_t \) in terms of the random cost \( \hat{C}_{t+1} \) (cf. eq. (2.6)), the optimal hedge is obtained by basically the same simple calculation as in sect. 2, as

\[
\phi_t = \frac{1}{S_{T-1}} \frac{\langle \hat{C}_t (\hat{u}_t - \langle \hat{u} \rangle + \lambda \mu) | S_{T-1} \rangle}{\sigma^2 + \lambda \mu^2}. \tag{4.21}
\]

The fair-game price can be written as

\[
C_t = \langle \hat{C}_T \hat{\rho}_{T-1} \hat{\rho}_{T-2} \ldots \hat{\rho}_t | S_t \rangle, \quad \hat{\rho}_t := \frac{\sigma^2 - \mu (\hat{u}_t - \langle \hat{u} \rangle)}{\sigma^2 + \lambda \mu^2}. \tag{4.22}
\]

Note that \( \hat{\rho} \) here differs by an overall factor from that of sect. 2,

\[
\hat{\rho}_t = \frac{\sigma^2}{\sigma^2 + \lambda \mu^2} \hat{\rho}_t | \lambda = 0 \tag{4.23}
\]

(except for \( \mu = 0 \), when they are both equal to 1). This form of \( \hat{\rho} \) has several interesting consequences:

- \( \rho_t < 0 \) can still occur; in fact, it occurs under the same conditions for all \( \lambda \). This is in accord with our conclusion that negative prices are related to unrealistic features of the market, making the fair-game condition inappropriate, not our specific hedging strategy.
• We can still define, \( q_k := p_k \hat{\rho}_k \), and all formulas of the \( \lambda = 0 \) case also hold here when expressed in terms of \( q \) or \( \hat{\rho} \) (besides eq. (2.13) for the premium, note that we can write the hedge as \( \mu S_0 \phi_t = (\hat{C}_{t+1} - \hat{\rho}_t)S_t \) for all \( \lambda \)). However, note that \( q \) does not even define a signed measure now, as

\[
\sum_i q_i = \frac{\sigma^2}{\sigma^2 + \lambda \mu^2} < 1.
\]

• The case \( \mu = 0 \) is again special, and everything becomes independent of \( \lambda = 0 \), i.e. we recover the usual (local) variance minimization approach.

• As long as \( \mu \neq 0 \) we can drive the cost \( C_0 \) of the option to 0 by choosing \( \lambda \) sufficiently large. This might at first sight appear illogical, but actually makes perfect sense: If \( \mu > 0 \) we can, with a suitable strategy, gain money during the hedge, enabling us to lower the premium. We can make the price arbitrarily small if we are allowed to use hedge ratios \( \phi_t > 1 \). Similarly, for \( \mu < 0 \) we can make money by short-selling stock, \( \phi_t < 0 \). It turns out that the maximal \( \phi_t \) (or minimal one for \( \mu < 0 \)) we need to drive the price to 0 does not even diverge: for given \( u, p \) it is uniformly bounded in \( T, K, S_0 \), and \( \lambda \). It does diverge, of course, as \( |\mu| \to 0 \).

• Naturally, there is a disadvantage in lowering the fair-game price by using a large \( \lambda \): the risk will usually go up.

We will refer to the above one-parameter family of strategies as cost-effective variance minimization, CVM(\( \lambda \)).

As seen above, for \( \mu \neq 0 \) we can, in principle, drive the cost of the option to zero by choosing \( \lambda \) sufficiently large. In practice this would not be advisable, as the risk will eventually become much larger than the price. To decide on a strategy and price we could, for example, add a risk premium to the naive price (at least if we go beyond the minimal risk of \( \lambda = 0 \)), and then find the value of \( \lambda \) that minimizes the risk adjusted price. As long as the price is much larger than the risk, one would presumably add at most a very small fraction of the risk, but as the risk becomes comparable to the naive price that fraction should grow to order 1. One might, for example, choose the risk-adjusted price to be,

\[
C_R(\lambda, \alpha) = \sqrt{C_0(\lambda)^2 + \alpha(R(\lambda)^2 - R(0)^2)},
\]

where \( \alpha \) is a number of \( O(1) \) and \( R(\lambda)^2 \) the variance of the wealth balance between times 0 and \( T \) under the hedging strategy CVM(\( \lambda \))\(^8\) (one could also choose \( R^2 \) to only count the contribution to the variance from losses, but for our present illustrative purposes the above is good enough). Note that by definition of the

\(^8\)We have chosen a form where \( C_R(0, \alpha) = C_0(0) \) independent of \( \alpha \). We are not claiming that one should not add a risk premium for the minimal variance hedge, but choose this form simply for a later comparison of different \( \alpha \) to be “fair”, i.e. not prejudiced against larger values of \( \alpha \).
CVM(\(\lambda\)) strategy the risk-adjusted price \(C_R(\lambda, \alpha)\) is minimal, for \(T = 1\) and fixed \(\alpha\), when \(\lambda = 1/(\alpha^2)\).

As mentioned already in sect. 2.2, it is of course not possible to pin down a price with arbitrary precision by such general considerations. Details of the market and its players will be relevant in practice. Nevertheless, the family of strategies CVM(\(\lambda\)) should be of some help in determining optimal, in some objective sense, prices and hedges under various circumstances.

4.3. The trinomial example

Let us now study the example from sect. 2.2 with our strategies. In figure 4 we show the naive and risk-adjusted prices, according to eq. (4.25), as a function of \(\lambda\) for various values of \(\alpha\) and two values of the maturity. Although there is some unresolvable ambiguity in the optimal hedging, and correspondingly in the price, it is quite clear that choosing some \(\lambda > 0\) in a suitable range (that will depend on \(T, K,\) etc) is of definite advantage: All prices fall rapidly for small \(\lambda\), till they reach a more or less broad plateau, where the optimal \(\lambda\) are to be found.

The above behavior is quite typical. The intuitive reason behind this behavior was mentioned earlier. By minimizing the variance, including those of gains, one misses the chance to harness some of the gains to lower the price. To check this interpretation we calculated the fraction of the variance that comes from losses, which is shown in figure 5. We see that for \(\lambda = 0\) only a relatively small fraction of the variance comes from losses (29\% and 25\%, respectively); all the rest comes from random gains. This is clearly an inefficient strategy. As we increase \(\lambda\) the fraction from losses shoots up, before reaching a plateau. This is a clear confirmation of the scenario we sketched for the problem of too high premiums and its remedy. Note that the fractional variance from losses and risk-adjusted prices reach their plateaux in roughly the same range of \(\lambda\). For completeness we show hedge ratios for this example in figure 6.

In figures 4 – 6 we also show where the strategy of Sommer [34] lies relative to ours. Results from his approach are shown at the value of \(\lambda\) where his fair-game price matches that of the CVM(\(\lambda\)) strategy (in the scale of the figure one can just barely see the difference between the fair-game and risk-adjusted premiums for his strategy). Sommer's strategy uses a measure of risk that is linear in fluctuations of the wealth balance, not quadratic like ours. This prevents a simple general formalism like the one presented here; his strategy is specific to the trinomial model.

For \(T = 1\) his strategy is exactly equivalent to ours for some value of \(\lambda\), which varies between 0 for sufficiently small strike \(K/S_0\) and, in this case, 1/15 for large \(K/S_0\). For higher \(T\) his strategy is, in general, not exactly equivalent to any \(\lambda\), but we have found empirically that there is still always a \(\lambda\) to which it is quite close, in price, variance and hedge ratios (cf. the right panel of the figures). Based on the criteria of optimality used above his strategy is not quite optimal, but it is better than strict variance minimization.

These and many other examples were also compared with another pricing and
Fig. 4. Fair-game and risk-adjusted premiums, eq. (4.25), for a European call option in a trinomial model using the CVM($\lambda$) strategy, as function of $\lambda$, for $T = 1$ (left) and $T = 3$ (right). The model is defined by $u = (1, 4, 0.9, 0.6)$, $p = (0.8, 0.1, 0.1)$, $r = 1$, $K = S_0 = 100$. The dot-dashed line presents the standard deviation $R(\lambda)$ of the wealth balance with respect to the fair-game price. The latter is given by the lower solid line. The upper three lines give $C_R(\lambda, \alpha)$ for $\alpha = 0.7, 1.0, 1.5$, respectively.

Fig. 5. Fractional contribution of losses (relative to the fair-game price) to the variance of the wealth balance for the model and parameters of figure 4, with $T = 1$ (left) and $T = 3$ (right).
Fig. 6. Hedge ratios $\phi_t(S_t)$ for the model and parameters of figure 4, with $T = 1$ (left) and $T = 3$ (right). For the latter case we only show the hedge ratios for $t = 0, 1$, as there would be nine curves to show for $t = 2$, corresponding to the nine possible values of $S_{t-2}$.

hedging strategy presented in [31]. The strategy is inspired by the ideas of Kelly [36] about maximizing the average growth rate of one’s wealth (in game theory, portfolio optimization, etc). This approach involves static hedging, in a sense.\footnote{Namely, in that one is supposed to invest a constant fraction of one's wealth in stocks. However, it is not a “lazy” strategy: as the stock goes up one must sell, if it goes down one has to buy to maintain the constant fraction.} We find that the prices of this approach are always higher than those of our CVM($\lambda$) strategy, for any $\lambda$. Also, the risk in this approach always seems to be higher. The seller does however often make a big profit; the price is not determined by a fair-game condition. In other words, this approach behaves somewhat like an investment strategy. In view of the underlying ideas this is perhaps not surprising.

Referring back to our discussion in sect. 2.4, we should remark that often the CVM($\lambda$) strategies with small $\lambda > 0$ have not just smaller (risk-adjusted) prices but even a smaller variance than for $\lambda = 0$! We have found this in many examples, including the above trinomial one (the value of $\lambda$ that minimizes the CVM($\lambda$) variance is however very small, and the corresponding variance only smaller by a tiny amount than for $\lambda = 0$; this fact is present but not visible in figure 4). This is another explicit proof that the local variance minimization performed in sect 2.3 does in general not find the global minimal variance when all possible hedging strategies are considered, as noted by Schweizer [30].

5. Implied Volatility, Skew and Kurtosis

We mentioned two other hedging and pricing strategies [34,31] in the previous section. We found that they are either less efficient than any CVM($\lambda$) strategy,
or roughly equivalent to CVM(\(\lambda\)) for some \(\lambda\) (which will generally not be optimal, depending on what criteria one adopts).

We have not yet discussed the most common hedging strategy used in practice, based on implied volatilities. For a vanilla option, as the European call under consideration here, one forgoes pricing the option and instead takes the Black-Scholes assumptions to hold, which allows one to back out an "implied volatility" \(\sigma_{\text{imp}}\) from the observed market price at the given strike and maturity. One then hedges as if the market were complete, i.e. as if there were no risk.

5.1. Price and risk with binomial hedging

Let us now explore the analog of this strategy in our multinomial toy world. We will assume that some \(N > 2\) model provides the real, non-risk-free world which sets the prices. For a given strike and maturity we determine the binomial distribution that reproduces the "real world" price. To make this problem well-defined we assume the binomial distribution to be skew-less, an assumption also made in the continuum Black-Scholes case. We also choose the binomial distribution to reproduce the mean rate of return of the stock in the multinomial world.\(^{10}\)

Once we know the binomial distribution we can determine the hedge ratios along any given price path (using the formulas of sect. 2). If the world really were binomial, this would be a riskless hedge. Since the true price evolution is not binomial, the hedge will not be riskless; in fact, the variance of the hedging gains and losses along the different paths must be at least as big as for the risk-minimized CVM(0) (assuming that the binomial hedge does not find a minimum lower than that from local variance minimization, which we have never seen happen). Furthermore, there is no reason that the total loss during hedging has to equal the price anymore; the price is only a fair-game price for the multinomial hedge! In the following we will refer to the loss during binomial hedging, together with the loss from the payoff, as the "binomial price", since this is what one should charge to reach a fair-game situation with the binomial hedge.\(^{11}\)

We would like to know how the binomial price and hedge, including the risk, compare to the price, hedge and risk obtained with the "real world" multinomial model using the CVM(\(\lambda\)) strategies.

To explore these questions we designed discrete distributions to have some of the properties that are known to exist for real world distributions, namely, non-zero skewness and kurtosis. These are dimensionless numbers, defined as the third and fourth order cumulants, respectively, measured in units of the standard deviation

\(^{10}\)Note that the discrete binomial distribution has three free parameters, whereas its continuum analog in the present context, the Gaussian, has only two, as it is automatically skewless. Although in the binomial case \(p\) does not matter for risk-less hedging and pricing, once we set \(p = (\frac{1}{2}, \frac{1}{2})\) the skew automatically vanishes and we still need two conditions (on the variance and the average) to fix \(u_1, u_2\).

\(^{11}\)If all market makers use the Black-Scholes approach, one wonders, of course, how exactly market prices come to be what they are in the real world. This interesting, important and difficult question is beyond the scope of this work.
of the distribution. Both vanish for a Gaussian distribution. A positive (negative) skewness means, as the name suggests, that the distribution is skewed towards larger (smaller) values of the \( u_i \). A positive (negative) kurtosis means that the distribution is overall more (less) fat-tailed than a Gaussian.

It is quite easy to design discrete distributions with any desired skewness and kurtosis. One can take a standard continuum distribution with suitable shape parameters, discretize it, and tune the parameters to get the desired values of skewness and kurtosis. The variance and average can always be adjusted later by rescaling and shifting the set of \( u_i \). In this manner we have studied a number of multinomial models, which allowed us to extract several generic features of pricing and hedging in this world. We have checked that these features are quite stable under changes of the details of the distribution; they therefore should also apply to continuum distributions with options priced and hedged along the lines discussed in this paper.

Before proceeding, we remark that for purposes of backing out \( \sigma_{\text{imp}} \) we take the “real-world” price from the multinomial model to be the variance-minimal, non-risk-adjusted CVM(0) price. This choice is mainly for definiteness; other choices of \( \lambda \) or the risk premium give very similar results. This is fortunate, since, as remarked in the previous footnote, a detailed quantitative understanding of how option prices are formed in the real world does not currently exist (even if we knew the underlying stock price process precisely).

The first question is, how do the binomial price and risk compare to those from the CVM(\( \lambda \)) strategies? We find that the variance of the hedging gains and losses is always larger than the minimal one achievable with pure variance minimization CVM(0). Correspondingly, the fair-game binomial price is often, though not always, somewhat lower than the CVM(0) one. Furthermore, the binomial price and variance are often (but not for positive skewness, see below) quite close to those of CVM(\( \lambda \)) for some \( \lambda \) that depends on the precise model, the strike, and the maturity.

We think this is very interesting. Remembering that the other strategy for pricing and hedging in incomplete markets, that of Sommer [34] considered in the previous section, was also usually very close to CVM(\( \lambda \)) for some \( \lambda \), one might begin to speculate that the CVM(\( \lambda \)) strategies exhaust, at least in an approximate numerical sense, the set of all “reasonable” strategies for incomplete markets. In other words, the simple CVM(\( \lambda \)) strategies might define an approximate “efficient frontier” of premium versus risk.

Returning to our qualitative findings concerning the binomial price and risk, we summarize them as follows:

- For a given “real world” distribution, the binomial price and risk correspond approximately to some CVM(\( \lambda \)) as long as the skewness is small or negative and as long as the strike is not too far out of the money. In the latter case there is a tendency for the binomial values to be worse than any CVM(\( \lambda \)). In other words, if we match the binomial price or risk to that of some CVM(\( \lambda \)), than the risk or price, respectively, of the latter is lower.
• This tendency at large strikes is enhanced for distributions with positive skewness, where typically the binomial price is higher than that of any CVM(λ), and the risk is also higher than that of CVM(0).

• A higher kurtosis generally moves the binomial hedge further away from the approximate efficient frontier defined by the CVM(λ) strategies.

Empirically it seems that stock indices tend to have negative skewness, whereas individual stocks exhibit positive or small values (cf. chapter 1.4 of [39]). When hedging options on the latter we expect the risk-free hedging approach to be more seriously wrong than for the former.

Let us illustrate some of the above remarks with an example. We consider two $N = 5$ models designed to have a kurtosis of 1 and a skew of either $-0.5$ or $+0.5$, with the same mean rate of return and volatility. With

\[
p = (0.181959, 0.629529, 0.167940, 0.018799, 0.001772)
\]

\[
u = (1.134060, 1.013484, 0.892908, 0.772333, 0.651757)
\]

we have (for $r = 1$) $\mu = 0.01$, $\sigma = 0.08$, and a skew of $-0.5$. To obtain the distribution with positive skew 0.5 we just have to reflect the $u$: $u \rightarrow 2 \cdot 1.01 - u$. The values of return, variance, skew and kurtosis are not atypical for the distribution of stock price changes on a time scale of a few weeks or a month (to be sure, when quoting dimensionful quantities like $\mu$ and $\sigma$ they are always measured in units of this time scale, not in terms of a fixed physical time scale like years).

In figure 7 we show the fair-game and risk-adjusted prices obtained with the CVM(λ) strategies for various maturities and strikes together with the binomial prices obtained as described above. They clearly show that the binomial hedge can be quite far away from the (approximate) efficient frontier of premium versus risk; and more so in the positively skewed case.

5.2. Implied volatilities

Let us now look at the implied volatilities themselves. Do they exhibit the familiar patterns of “smiles” and “smirks” observed in the real world? More precisely, under what conditions on the underlying price process do such patterns appear? How do they change as we vary skew and kurtosis? Do they depend on other features, such as the mean rate of return of the stock?

We have explored these questions for a wide variety of distributions. We find that the patterns observed in the real world can be reproduced by suitably varying

\footnote{To include the binomial hedge results in the same figure we show them at the value of $\lambda$ where they match the CVM(λ) fair-game price. The risk-adjusted price is then calculated according to eq. (4.25) using for $R(\lambda)^2$ the average variance of losses and gains accumulated during binomial hedging along the various price paths. In cases where there is no $\lambda$ such that the binomial price matches the CVM(λ) one, we show the binomial results at $\lambda = 0$.}

\footnote{We again used the fair-game CVM(0) price for purposes of backing out the implied volatility. We found that using other $\lambda$ gives qualitatively, and largely even quantitatively, similar results.}
Fig. 7. Fair-game and risk-adjusted prices for a European call option obtained with the CVM(\(\lambda\)) strategies in two \(N = 5\) models with same kurtosis 1 and negative (left), respectively, positive skew (right) \(\pm 0.5\). The precise models are defined in eq. (5.26). The solid and dotted lines are as in figure 4. We also show the fair-game price and the \(\alpha = 1\) risk-adjusted price resulting from binomial hedging when the multinomial CVM(0) price is used to obtain an implied volatility. The option parameters are given in the figures, with \(S_0 = 100\) in all cases. For details see the main text.
the characteristics of the multinomial distribution. Finding a smile- or smirk-like pattern for the implied volatility $\sigma_{\text{impl}}$ as a function of strike $K$ is quite generic. We can summarize the qualitative dependence of the implied volatility structure on the distribution as follows:

- For an excess rate of return $\mu > 0$ (recall $\mu \equiv \langle \hat{u} \rangle - r$) the implied volatility $\sigma_{\text{impl}}(K)$ tends to be a decreasing function. A larger $\mu$ tends to shift the characteristic features of any pattern (e.g., the minimum of the smile) to larger $K$, relative to $S_0$. Furthermore, all else being equal, a larger $\mu$ tends to raise $\sigma_{\text{impl}}$.

- For $\mu < 0$ these tendencies are reversed.

- A negative skew tends to lift the left end of the smile $\sigma_{\text{impl}}(K)$ and shifts the minimum of the smile to larger $K$.

- A positive skew has the opposite effects.

- Increasing the kurtosis enhances the smile (i.e., deepens it). It also tends to shift the minimum of the smile to larger $K$.

- As the time to maturity increases all patterns are flattened.

Note that the dependence on the excess rate of return $\mu$ is perhaps a bit surprising from a Black-Scholes point of view, since in that framework $\mu$ drops out of the picture.

When considering a small deviation from a Gaussian distribution in a continuum formalism, one can easily derive an analytic formula for the dependence of the implied volatility on kurtosis and strike, at least if $\mu$ and the skewness vanish [40]. It predicts a parabolic dependence of $\sigma_{\text{impl}}$ on $K$; the parabola being either a “smile” or a “frown” depending on the sign of the kurtosis. It also predicts that the extremum of the parabola is at strikes $K > S_0$. This is in accord with our above observations. We should also note the elementary fact that the kurtosis can be made arbitrarily large but is bounded from below by $-2$; this follows easily from its definition. This makes it harder to see a “frown”, even in a toy-world like ours where we can custom-design the price distribution (in fact, we never found a really convincing frown).

Illustrations of the stylized facts summarized above can be found in figures 8 – 10. The first two illustrate the dependence on the excess mean return and the kurtosis. The following figure illustrates the effect of skewness, and the decay of the volatility structure as the maturity increases.

The oscillatory fine structure in these plots is of course due to the discrete nature of the multinomial models. A continuous model with the same gross features as the discrete one would, roughly speaking, exhibit an implied volatility that is a smeared-out version of such a plot. These plots also indicate that from a numerical point of view the implied volatility is a somewhat dangerous quantity (like a numerical derivative), as it amplifies any (potentially irrelevant) fine-structure of the underlying price process.
Fig. 8. Implied volatility in $N = 5$ multinomial models with volatility $\sigma = 0.08$, vanishing skewness, negligible kurtosis 0.01 and mean excess rate of return $\mu = 0.02$ (left), respectively, $\mu = -0.02$ (right), for a maturity of $T = 10$ and an initial stock price of $S_0 = 100$. In both cases we use $p = (0.049159, 0.210287, 0.481107, 0.210287, 0.049159)$. Choosing $u = (0.842643, 0.931322, 1.02, 1.10668, 1.19736)$ gives $\mu = 0.02$. To obtain $\mu = -0.02$ just shift $u_i \rightarrow u_i - 0.04$.

Fig. 9. Implied volatility in $N = 5$ multinomial models with volatility $\sigma = 0.08$, vanishing skewness, kurtosis 2, and mean excess rate of return $\mu = 0.02$ (left), respectively, $\mu = -0.02$ (right), for a maturity of $T = 10$ and $S_0 = 100$. In both cases $p = (0.0308357, 0.12541, 0.687508, 0.12541, 0.0308357)$. Choosing $u = (0.793159, 0.90668, 1.02, 1.13342, 1.24684)$ gives $\mu = 0.02$. To obtain $\mu = -0.02$ shift $u_i \rightarrow u_i - 0.04$. 
Fig. 10. Implied volatility in $N = 5$ models defined by eq. (5.26) with kurtosis 1 and negative (left), respectively, positive skewness (right) $\neq 0.5$, for maturities $T = 1, 4, 10, 20$, from top to bottom.
One question that we have so far not discussed is whether, or to what extent, the features discussed in this paper, and this section in particular, survive the continuum limit. Actually, we are not really interested in the continuum limit, but rather in a horizon of a day, say, corresponding to the hedging frequency commonly used in practice. Even so, the examples we discussed correspond to a unit of time of weeks rather than days. When accessing the effect of hedging frequency on performance one must take into account the scaling of the price distribution with time horizon (it is known, for example, that the kurtosis rapidly increases on short time scales, with an anomalous scaling exponent \([4,40,9]\)). We prefer to postpone a discussion of hedging with higher frequencies to future work where it can be combined with more realistic models for the price process, including stochastic volatility and/or GARCH-like features \([32]\).

6. Conclusion and Outlook

We have studied option hedging and pricing strategies in incomplete markets. Our aim is to relax some of the unrealistic assumptions of the Black-Scholes framework concerning the price process. This requires one to go beyond no-arbitrage ideas and the implied volatility cum delta-hedging approach usually employed, which, in the real world, is rather ad hoc. Another potential drawback of this approach is that one has to take prices from the market, leaving the question of possible mispricings open.

As repeatedly emphasized, the concept of no-arbitrage seems rather artificial in the real world, where there is always risk. The mathematical pendant of this is that the concept of a complete market is not stable under generic small perturbations of the price process. Rather than pretending that there is no risk, as in the Black-Scholes world, it seems to us that explicitly considering and dealing with risk is the right way to proceed (note that the risk is quite large: it has been estimated to be 25% of the option price for typical one-month options on liquid markets \([41]\), and can presumably be significantly larger for illiquid markets). This view seems to be gaining support in the finance community; in the insurance world it is the canonical view, of course.

One popular approach for pricing and hedging in incomplete markets is provided by the variance minimization ideas of Föllmer, Schweizer, Bouchaud, Sornette et al. It is a natural generalization of the no-arbitrage approach to cases where risk cannot be completely eliminated and is free of subjective preferences (which is a two-sided sword, cf. below). We have seen that a potential conceptual worry often held against this approach is unjustified. Namely, the negative premiums that can arise in certain cases when one imposes a fair-game condition to fix the price, are not to be blamed on variance minimization. Rather, they occur because the market allows one, in such cases, to make money from the stock investments performed during variance-minimal hedging. A fair-game condition naturally leads to negative premiums then. In other words, the fair-game condition becomes inappropriate (in practice, parameter uncertainties might also be sufficient to prevent negative
premiums from arising, cf. sect. 3).

On the other hand, various examples support the intuitive notion that variance minimization can be non-optimal because it tries to minimize not just random hedging losses but also welcome hedging gains. We have proposed a simple method, \textit{cost-effective (local) variance minimization}, CVM, that allows one to partially capture these gains. We obtained explicit and simple formulas for the hedging strategy and the price of the option. CVM leads to more reasonable hedging strategies in many cases, allowing one to sell an option at a lower price (or make a larger profit) without a significant increase in risk.

More precisely, CVM is really a one-parameter family of hedging and pricing strategies, which allows a tradeoff between the risk the seller of an option is willing to take on and at how low a price he wants to offer his product on the market. We have presented some evidence that the CVM family of strategies might, at least in an approximate sense, map out the efficient frontier of premium versus risk. The evidence involves the comparison with other proposed hedging strategies, like that of Sommer [34] and the implied volatility approach often used by practitioners (suitably translated into our discrete world). We find that other hedging strategies are either less efficient or lie close to some point on the efficient frontier provided by the CVM strategies. Depending on one’s risk preferences and on the specific market conditions, one may or, generically, may not be satisfied with the price and risk these other hedging strategies lead to. The CVM family therefore provides a flexible tool that can be adapted to specific market conditions and risk preferences.

As repeatedly emphasized, we aim for strategies that are as free of subjective preferences as possible (but no more!). Being a hedging \textit{and} pricing tool, CVM serves this purpose in that it can be used to detect and exploit potential mispricings of the market. If such mispricings are detected one can choose to either make a larger profit or try to gain market share by lowering the premium. Ideally, this would be the only subjective choice the seller of an option would have to make. Of course, in reality there are other ways in which the (partially) subjective opinions of market players enter. For example, changes in option prices representing the collective wisdom of the market might indeed correctly anticipate future market developments. It might be advisable to incorporate such indications in one’s model of the price process, instead of just relying on historical price data.

We have studied and developed our framework in the incomplete toy world provided by the discrete-time, discrete-space multinomial models. In such models we have compared the CVM strategies in some detail with the implied volatility approach (which is only a hedging and not a pricing tool; the price has to be taken from the market, which in our toy world is a multinomial model). Besides the inefficiency of this approach mentioned above, we have also studied how implied volatility surfaces depend on the various features (average excess return, kurtosis, skewness) of the underlying “real-world” price process. All patterns observed in the real world can be reproduced in our toy universe by using suitable non-gaussian (i.e. non-binomial in our world) return distributions.
We regard this study as preparatory [32] for an investigation of more realistic price processes including continuum distributions, stochastic volatility and/or GARCH-like features. The extension of our ideas to such cases is in many respects straightforward; recall that the variance minimization approach is well-developed for both discrete and continuum processes. As far as discrete-time, continuous-state models are concerned, most of our results in fact hold without modifications, cf. [30,32]. Other extensions might also be of interest, e.g. the inclusion of transaction costs or the use of measures of risk that are not based on variance, but rather related to properties of the tails of the distribution (cf. [22,23]).

Acknowledgements

I am grateful to Mark Broadie and Jaksa Cvitanić for discussions and comments on the manuscript. This work was supported in part by the US Department of Energy.

References


