Estimating state price densities by Hermite polynomials: theory and application to Italian derivatives market

Paolo Guasoni *
Dipartimento di Matematica
Università di Pisa
Via Buonarroti, 2
56127 Pisa
guasoni@dm.unipi.it

September 26, 2001

Abstract

This paper explores the possibility to extract the risk-neutral probability of an asset implied in market prices of options.

Adapting a model of Madan and Milne to a multiple expiration setting, we present an estimation method for the risk-neutral probability at a moving horizon of fixed length, which exploits the prices of options with different expirations. With the exception of volatility, all model parameters can be estimated by linear regression, and their number can be chosen arbitrarily, depending on the size of the dataset.

We discuss empirical issues related to the application of this model to real data, and show results on listed options on the Italian MIB30 equity index.

*The first draft was completed while the author was affiliated to the Research Department of the Bank of Italy. This author wishes to thank Gur Huberman, Roberto Violi and Giuseppe Grande for stimulating discussions, and for carefully reading earlier versions of this paper. Financial support from the Bank of Italy and the University of Pisa is gratefully acknowledged. The views expressed herein are those of the author and not necessarily those of the Bank of Italy.
1 Introduction

In modern finance theory, usual financial instruments are seen as combinations of elementary Arrow-Debreu securities, and asset prices can be obtained as expected values of their future payoff under a state-price density $Q$, also known as risk-neutral density.

In general, $Q$ is unique if and only if the market is complete, and in this case option prices are exactly determined by the no-arbitrage condition. On the contrary, in a market with incomplete information there are infinitely many risk-neutral probabilities, each of them reflecting a particular attitude for risk.

The knowledge of $Q$ can be relevant for a number of applications, ranging from the pricing of unlisted derivatives (as OTC contracts) to risk management. From the point of view of a regulator, the knowledge of both the risk-neutral probability $Q$ is especially useful in combination with that of the physical probability $P$, as it allows to monitor the time change of risk aversion in the market. In view of these applications, the natural question is whether we can recover the marginal distribution of an underlying asset $S$ under $Q$ from the observation of option prices.

Several methods have been proposed in the literature for this purpose: a first approach consists in modeling the dynamics of the underlying asset under $Q$, so that risk-neutral densities can be written in parametric form. This case encompasses the stochastic volatility models of Heston [3] and Stein and Stein [7], as well as several others with deterministic volatility. In a few simple cases, this method is particularly flexible and easy to implement, but in general it poses a number of issues:

- it is heavily model-dependent, since it requires an *a priori* specification of a stochastic process for the asset price;

- for complex processes, the risk-neutral densities do not admit closed-form expressions, and numerical solutions of PDEs or simulation algorithms must be employed;

- when several multiple parameters appear in a joint minimization problem, it is necessary to devise an estimation algorithm which avoids local minima.

A second approach directly prescribes a parametric form for risk-neutral densities, without specific assumptions on the underlying process under $Q$. This method includes all various parametrizations of the “smile”, along the lines of Shimko [6], Rosenberg and Engle [2] and several others. Although
it has a clear edge for its simplicity, and it circumvents the first two issues above, the third problem remains, and others arise:

- the choice of a particular functional form is often arbitrary, and may pose specification problems;
- in the attempt to span a wide range of densities, several parameters might be necessary, leading to the risk of overfitting.

Some authors take up a Bayesian approach, solving for the risk-neutral density which is closest to a given prior, under the constraint of pricing correctly all options observed. For example, this was done by Rubinstein [5], to calibrate an \textit{implied binomial tree} from option prices. While this approach is general enough to allow virtually any density form, it is not completely clear what distance criteria should be preferred, and what is the impact of the prior on the final result.

The last approach, proposed by Aït-Sahalia and Lo [1] is essentially non-parametric: first the pricing function $C(S, K, r, \delta, \tau)$ is estimated with the kernel regression technique, then the risk-neutral density is obtained via the well-known identity due to Breeden and Litzemberger:

$$\frac{\partial^2 C}{\partial K^2}\bigg|_{K=x} = e^{-r(T-t)}q(x)$$

where $q(x)$ denotes the marginal density of $S_T$ at $x$. While this method is the most general, and can capture virtually any feature displayed by the data, it works best when a semiparametric variant is used, and it generally requires the aggregation of data across different time observations. This means that it is best suited for large-sample studies, where a single risk-neutral density is assumed to explain prices for a certain period of time.

In this paper, we adopt a model suggested by Madan and Milne [4], which is parametric in its implementation, while it allows the representation of any risk-neutral density satisfying reasonable integrability conditions. More precisely, the density of the underlying logarithm is expanded in Hermite series, after scaling by a normalization factor, which plays the role of volatility. When all other parameters are equal to zero, the model boils down to the standard Black-Scholes case.

Developing further the analysis of this model, we translate in terms of parameter constraints the conditions that the density represents indeed a probability (i.e. it integrates to one), and that it is risk-neutral. This reduces the scope for inconsistency and overfitting. We also show how the densities corresponding to two different expirations can be used to estimate the risk-neutral density at a fixed-length horizon. This involves the calculation of
the Hermite expansion of a convolution of two densities, and the assumption
that the underlying process has independent increments.

In this model, option prices are calculated in closed form as the scalar
product between the vector of Hermite coefficients and a vector of explicit
formulas, depending only on the volatility parameter: we show an efficient
method to obtain these formulas recursively, in symbolic form.

For a given value of volatility, the model is linear and can be easily solved
with ordinary least squares. The full nonlinear model can also be solved
with standard nonlinear regression algorithms, and convergence to the global
minimum is guaranteed by the convexity of the functional.

Finally, we show that for a particular two-parameter choice we have a
one-to-one correspondence between the Hermite coefficients and skewness
and kurtosis. Indeed, the two Hermite coefficients become constant multiples
respectively of skewness and excess kurtosis, thereby providing a consistent
framework for the estimation of these quantities.

The paper is organized as follows: in section 2 we describe the model
details, and show how the components of option prices can be computed
recursively. Then we exploit the same calculations to write the risk-neutrality
condition in terms of parameter values, and see that one parameter can be
eliminated if the density has to integrate to one. The aggregation of data
across expirations is covered in Section 3, where we present a method to
estimate the risk-neutral measure at intermediate horizons.

Empirical issues, as well as an application to real data from the Italian
Derivatives Market, are the subject of Section 4. We discuss the choice of the
set of parameters, which is intimately related to the moments of risk-neutral
densities, and show numerical results from our dataset, which consists of
intraday data on prices and volumes of all transactions on MIB30 index
options during 1998. As a matter of fact, the period under consideration
has shown a wide range of market conditions, which provide a challenging
stress test for the model. In the last section we briefly comment our results,
discussing the benefits and the limits of this methodology.

2 The Model

Throughout the paper, \( S_t \) denotes the price of the underlying asset at time
\( t \), \( T \) the expiration date of an option, \( K \) its strike price, \( r \) the interest rate,
and \( \delta \) the dividend yield.

We represent the random variable \( S_T \) as:

\[
S_T = S_t e^{(r-\delta-\frac{\sigma^2}{2})(T-t)-\sigma\sqrt{T-t}\psi}
\]
where $\sigma$ is an arbitrary positive constant. Notice that this representation does not involve any assumption on the asset price dynamics, but only establishes a one-to-one mapping between a positive random variable $S_T$ and a real-valued random variable $\psi$. In particular, if $S_T$ is a lognormal, then $\psi \sim N(0,1)$ under any risk-neutral measure $Q$. Denoting by $q(x)$ the probability density of $\psi$ under $Q$, $\tau = T - t$, and $d_2 = (\log \frac{S_t}{K} + (r - \frac{\sigma^2}{2})\tau)/(\sigma\sqrt{\tau})$, we can rewrite the call price as:

$$C(S_t, K, r, \tau, \sigma) = e^{-\tau r} \int_{-d_2}^{\infty} (S_t e^{(r - \frac{\sigma^2}{2})\tau + \sigma \sqrt{\tau} x} - K) q(x) dx = e^{-\delta \tau} S_t \int_{-d_2}^{\infty} e^{-\frac{\sigma^2}{2} \tau + \sigma \sqrt{\tau} x} \phi(x) dx - e^{-\tau r} K \int_{-d_2}^{\infty} q(x) dx$$

We denote the Hermite expansion of $q(x)$ as:

$$q(x) = \phi(x) \sum_{n=0}^{\infty} \theta_n H_n(x) = \phi(x) \sum_{n=0}^{\infty} \theta_n H_n(x)$$

where $\phi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ is the standard normal density, and $H_n(x) = \frac{1}{\phi \ f \ dx} \mid_{x}$ are the Hermite polynomials. We recall their properties in the following:

**Proposition 1.** We have that:

- $\int_{-\infty}^{+\infty} H_i(x) H_j(x) \phi(x) dx = 0$ for all $i \neq j$.
- $\int_{-\infty}^{+\infty} H_n(x)^2 \phi(x) dx = n!$ for all $n$.
- If $f \in L^2(\mathbb{R}, N(0,1))$, then $f(x) = \sum_{n=0}^{\infty} \zeta_n H_n(x)$ for some set $\{\zeta_n\}_{n \in \mathbb{N}}$.

In other words, the set $\{H_n\}_{n \in \mathbb{N}}$ is an orthogonal basis of the Hilbert space $L^2(\mathbb{R}, N(0,1))$. We shall assume that $f \in L^2(\mathbb{R}, N(0,1))$, so that convergence holds. The price of a call option can be calculated as:

$$C(S_t, K, r, \tau, \sigma) = e^{-\delta \tau} S_t \int_{-d_2}^{\infty} e^{-\frac{\sigma^2}{2} \tau + \sigma \sqrt{\tau} x} \phi(x) \sum_{n=0}^{\infty} \theta_n H_n(x) dx - e^{-\tau r} K \int_{-d_2}^{\infty} q(x) dx = e^{-\delta \tau} S_t \int_{-d_2}^{\infty} e^{-\frac{\sigma^2}{2} \tau + \sigma \sqrt{\tau} x} \phi(x) \sum_{n=0}^{\infty} \theta_n H_n(x) dx$$

$$= e^{-\delta \tau} S_t \int_{-d_2}^{\infty} e^{-\frac{\sigma^2}{2} \tau + \sigma \sqrt{\tau} x} \phi(x) H_n(x) dx - e^{-\tau r} K \int_{-d_2}^{\infty} \phi(x) H_n(x) dx$$

$$= \sum_{n=0}^{\infty} \theta_n \left( e^{-\delta \tau} S_t \int_{-d_2}^{\infty} e^{-\frac{\sigma^2}{2} \tau + \sigma \sqrt{\tau} x} \phi(x) H_n(x) dx - e^{-\tau r} K \int_{-d_2}^{\infty} \phi(x) H_n(x) dx \right)$$

(1)

To proceed further we need the following lemma, which is the key to most calculations in this section:
Lemma 1. Let us define:

\[ Y_n(y, \gamma) = \int_y^{+\infty} e^{-\frac{x^2}{2} + \gamma x} \phi(x) H_n(x) dx \]

Then the following relations hold:

\[
\begin{align*}
Y_n(y, \gamma) &= -\gamma Y_{n-1}(y, \gamma) - \phi(y - \gamma) H_{n-1}(y) \\
Y_0(y, \gamma) &= 1 - \Phi(y - \gamma)
\end{align*}
\]

In particular:

\[ \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2} + \gamma x} \phi(x) H_n(x) dx = (-\gamma)^n \]

Proof. We prove the Lemma by induction. By definition of \( H_n \), and integrating by parts:

\[ Y_n(y, \gamma) = \int_y^{+\infty} e^{-\frac{x^2}{2} + \gamma x} \phi(x) H_n(x) dx = \int_y^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + \gamma x} \frac{d^n \phi}{dx^n} dx = \]

\[ = \int_y^{+\infty} (-\gamma) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + \gamma x} \frac{d^{n-1} \phi}{dx^{n-1}} dx - \left. \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + \gamma x} \frac{d^{n-1} \phi}{dx^{n-1}} \right|_y = \]

\[ = -\gamma Y_{n-1}(y, \gamma) - \phi(y - \gamma) H_{n-1}(y) \]

Since for \( n = 0 \) the calculation is trivial, the proof is complete. \( \square \)

The integrals

\[ A_n(y) = \int_y^{\infty} H_n(x) \phi(x) dx \quad \text{and} \quad B_n(y, \sigma, \tau) = \int_y^{\infty} e^{-\frac{x^2}{2} + \gamma x} \phi(x) H_n(x) dx \]

can be computed in closed-form by an application of Lemma 1, with \( \gamma = 0 \) and \( \gamma = \sigma \sqrt{\tau} \) respectively. It follows that \( C(S_t, K, r, \tau, \sigma) \) admits an explicit formula in series form:

\[ C(S_t, K, r, \tau, \sigma) = \sum_{n=0}^{\infty} \theta_n C_n(S_t, K, r, \tau, \sigma) \]

where

\[ C_n(S_t, K, r, \tau, \sigma) = e^{-\delta \tau} S_t B_n(-d_2, \sigma, \tau) - e^{-r \tau} K A_n(-d_2) \]

The first terms of \( C_n \) are shown in the appendix: \( C_0(S_t, K, r, \tau, \sigma) \) is simply the Black-Scholes formula. This is not surprising, since choosing \( \theta_0 = 1 \) and \( \theta_n = 0 \) for \( n > 0 \), \( q(x) \) is a standard normal density.

Let us now see how the natural restrictions on \( Q \) translate in terms of the coefficients \( \theta_n \). In fact, we have two conditions on \( q(x) \):
• $q(x)$ is a probability density;
• $Q$ is risk-neutral.

Leaving aside the positivity of $q(x)$, these properties are characterized by the following:

**Proposition 2.** Let $q(x) \in L^2(\mathbb{R}, N(0,1))$ be a positive function. Then we have:

• $q(x)$ is a probability density if and only if $\theta_0 = 1$.
• $q(x)$ is risk-neutral if and only if $\sum_{n=0}^{\infty} \theta_n (-\sigma \sqrt{\tau})^n = 1$.

**Proof.** From the first condition, we simply get:

$$1 = \int_{-\infty}^{+\infty} q(x)dx = \int_{-\infty}^{+\infty} \phi(x) \sum_{n=0}^{\infty} \theta_n H_n(x)dx = \sum_{n=0}^{\infty} \theta_n \int_{-\infty}^{+\infty} \phi(x)H_n(x)dx = \theta_0$$

where the last equality follows from the observation that $\int_{-\infty}^{+\infty} \phi(x)H_n(x)dx = 0$ for all $n > 0$. Hence we simply set $\theta_0 = 1$.

The second condition is:

$$E_Q[S_T] = S_t e^{(r-\delta)\tau}$$

In other words:

$$\int_{-\infty}^{+\infty} S_t e^{(r-\delta-\frac{\sigma^2}{2}\tau+\sigma\sqrt{\tau}x)}q(x)dx = S_t e^{(r-\delta)\tau}$$

and hence:

$$\int_{-\infty}^{+\infty} e^{-\frac{\sigma^2}{2}\tau+\sigma\sqrt{\tau}x}q(x)dx = 1$$

Observe that:

$$\int_{-\infty}^{+\infty} e^{-\frac{\sigma^2}{2}\tau+\sigma\sqrt{\tau}x}q(x)dx = \int_{-\infty}^{+\infty} e^{-\frac{\sigma^2}{2}\tau+\sigma\sqrt{\tau}x}\phi(x)dx \sum_{n=0}^{\infty} \theta_n H_n(x)dx =$$

$$= \sum_{n=0}^{\infty} \theta_n \int_{-\infty}^{+\infty} e^{-\frac{\sigma^2}{2}\tau+\sigma\sqrt{\tau}x}\phi(x)H_n(x)dx = \sum_{n=0}^{\infty} \theta_n (-\sigma \sqrt{\tau})^n$$

where the last equality follows from Lemma 1. □
3 Multiple expirations

The previous section outlines a method for extracting the risk-neutral density implied by a cross-section of option prices with the same expiration and different strikes. Since listed options are available for multiple expirations, this procedure can be separately applied to each of them, obtaining densities for different horizons, which approach from day to day. In contrast, risk management practice requires to look at a time window of fixed length, which generally does not coincide with the expiration date of an option.

In this section we show how the information obtained on the risk-neutral densities on two successive expirations $T_1$ and $T_2$ can be used to estimate the density at a certain time $T$ between them. Of course, such an estimation requires some assumptions on the process of the underlying: here we assume that the increment $S_{T_2} - S_{T_1}$ is independent of $S_{T_1} - S_t$, and that the random variable $S_T - S_{T_1}$ has the same distribution as $(S_{T_2} - S_{T_1}) \sqrt{\frac{T - T_1}{T_2 - T_1}}$. This allows us to write:

$$S_T = S_{T_1} + S_T - S_{T_1} \sim S_{T_1} + (S_{T_2} - S_{T_1}) \sqrt{\frac{T - T_1}{T_2 - T_1}}$$

and, by the independence assumption, the density of $S_T$ is obtained as the convolution of those of $S_{T_1}$ and $S_{T_2} - S_{T_1}$. This reduces the problem to the computation of the density of $S_{T_2} - S_{T_1}$, in terms of those of $S_{T_1}$ and $S_{T_2}$. In a similar fashion as the previous section, we can write $S_{T_2}$ as:

$$S_{T_2} = S_t e^{(r + \delta)(T_2 - t) - \sigma_1^2 (T_1 - t) + \sigma_1 \sqrt{T_1 - t} \psi_1 - \sigma_2^2 (T_2 - T_1) + \sigma_2 \sqrt{T_2 - T_1} \psi_2}$$

where the random variables $\psi_1$ and $\psi_2$ represent the normalized returns of the underlying respectively in the $(t, T_1)$ and $(T_1, T_2)$ intervals. We denote by $q_1(x)$ and $q_2(x)$ respectively the densities of $\psi_1$ and $\psi_2$ under $Q$. Again, we expand $q_1$ and $q_2$ in Hermite series:

$$q_1(x) = \phi(x) \sum_{n=0}^{\infty} \theta_1^n H_n(x) \quad q_2(x) = \phi(x) \sum_{n=0}^{\infty} \theta_2^n H_n(x)$$

The next proposition shows the relation between the Hermite decompositions of $q_1(x)$, $q_2(x)$ and a normalized linear combination $q(x)$. In particular, the relation is linear, and given the coefficients of two of them, those of the third are uniquely determined.

**Proposition 3.** Let $q_1(x)$ and $q_2(x)$ be the densities of $\psi_1$ and $\psi_2$ as above, and denote by $q(x)$ the density of the random variable $(\gamma_1 \psi_1 + \gamma_2 \psi_2) / \sqrt{\gamma_1^2 + \gamma_2^2}$, where the random variables $\psi_1$ and $\psi_2$ represent the normalized returns of the underlying respectively in the $(t, T_1)$ and $(T_1, T_2)$ intervals. We denote by $q_1(x)$ and $q_2(x)$ respectively the densities of $\psi_1$ and $\psi_2$ under $Q$. Again, we expand $q_1$ and $q_2$ in Hermite series:

$$q_1(x) = \phi(x) \sum_{n=0}^{\infty} \theta_1^n H_n(x) \quad q_2(x) = \phi(x) \sum_{n=0}^{\infty} \theta_2^n H_n(x)$$

The next proposition shows the relation between the Hermite decompositions of $q_1(x)$, $q_2(x)$ and a normalized linear combination $q(x)$. In particular, the relation is linear, and given the coefficients of two of them, those of the third are uniquely determined.
where $\gamma_1, \gamma_2 > 0$. Assuming that $\psi_1$ and $\psi_2$ are independent, the coefficients $\theta_n$ in the Hermite expansion

$$q(x) = \phi(x) \sum_{n=0}^{\infty} \theta_n H_n(x)$$

are given by:

$$\theta_n = \left(\frac{\gamma_1^2 + \gamma_2^2}{\gamma_2}\right)^{-\frac{n}{2}} \sum_{k=0}^{n} \gamma_1^{n-k} \gamma_2^{k} \theta_{n-k}^1 \theta_k^2$$

**Proof.** Denoting $\beta = \sqrt{\frac{\gamma_1^2 + \gamma_2^2}{\gamma_2}}$, we have:

$$q(y) = \beta \int_{-\infty}^{+\infty} q_1(x)q_2 \left(\frac{\beta y - \gamma_1}{\gamma_2} x \right) dx = \beta \int_{-\infty}^{+\infty} \theta_1^1 \theta_2^2 \int_{-\infty}^{+\infty} H_i(x)H_j \left(\frac{\beta y - \gamma_1}{\gamma_2} x \right) \phi(x) \phi \left(\frac{\beta y - \gamma_1}{\gamma_2} x \right) dx$$

Integrating by parts, we get:

$$\int_{-\infty}^{+\infty} H_i(x)H_j \left(\frac{\beta y - \gamma_1}{\gamma_2} x \right) \phi(x) \phi \left(\frac{\beta y - \gamma_1}{\gamma_2} x \right) dx = \int_{-\infty}^{+\infty} \frac{d^i \phi}{dx^i} \left|dx^i\right| \frac{d^j \phi}{dx^j} \left|dx^j\right| \phi \left(\frac{\beta y - \gamma_1}{\gamma_2} x \right) dx$$

Finally:

$$\int_{-\infty}^{+\infty} \frac{d^n \phi}{dx^n} \left|dx^n\right| \phi \left(\frac{\beta y - \gamma_1}{\gamma_2} x \right) dx = \frac{1}{\beta} \phi(y) H_n(y) \left(\frac{\gamma_1}{\sqrt{\gamma_1^2 + \gamma_2^2}}\right)^n$$

And the proof is complete. \qed

The above proposition shows how to compute the density of $S_{T_2} - S_{T_1}$, in terms of the densities of $S_{T_1}$ and $S_{T_2}$. In fact it is sufficient to substitute $\gamma_1 = \sigma_1 \sqrt{T_1 - t}$ and $\gamma_2 = \sigma_2 \sqrt{T_2 - T_1}$.

For estimation purposes, Hermite polynomials are truncated to a finite number of terms, hence $q_1$ and $q_2$ are typically partial sums of degrees $n_1$ and $n_2$ respectively. From the proposition above, it is immediately seen that the degree of $q$ cannot exceed $n_1 + n_2$, since for higher order terms all the products $\theta_{n-k}^1 \theta_k^2$ vanish.

A further observation, which may be useful in applications, is that the $k$-th moment of a density $q(x)$ depends only on the first $k$ terms of the Hermite expansion. In practice, this means that truncation can be based on the number of moments that are considered relevant.
4 Application to Italian Derivatives Market

4.1 Data

We now turn to the estimation of the model to market data. Our dataset consists of prices and volumes of all transactions on listed options and futures contracts on the MIB30 index during the year 1998. These options are traded on the IDEM (Italian Derivatives Market), a section of the Italian Exchange, which kindly provided the dataset. For the interest rates, we used the three-month LIBOR on the Italian lira.

Before we enter into empirical issues, let us spend a few words on the institutional features of the market: the MIB30 is a capitalization-weighted index based on a fixed basket of the 30 most liquid and highly capitalized common stocks on the Italian Exchange. Since options and futures are traded simultaneously on the same exchange, there is no time lag between the reporting of option and underlying prices, unlike in the S&P 500 options market.

Options with expiration in the quarterly cycle of March, June, September and December are listed at any time. In addition, the expirations corresponding to the two nearest months outside the cycle are made available. In practice there are sufficiently many liquid contracts only for the two nearest months, therefore our analysis is constrained to this time horizon. Liquidity tends to decrease near expiration dates, as trading shifts from one contract to the next.

4.2 Methodology

Estimation of the risk-neutral density requires the simultaneous observation a cross section of options with the same expiration but different strike. In a very liquid market, this is achieved considering the last quote on each contract before a fixed time of the day. This procedure is infeasible with our dataset, which does not include quotes; even if applicable, it would not exploit all the information embedded in transaction prices, as many options (usually those in-the-money) are thinly traded, and bid-ask spreads are very wide.

For each trading day, we record the last transaction before noon for each option contract, as well as the underlying value at the time of each recorded transaction. This ensures that illiquid contracts, which may be traded few times in an hour, are not associated with the value of the underlying at noon, which may be significantly different from the time of the last transaction.

A critical point is usually the measurement of the underlying value, as many authors have pointed out that it is often unreliable, either due to a
time lag in reporting, or to the unobservability of dividends, or both. As mentioned before, the first issue does not arise in our case, while the latter remains.

A possible solution is suggested by Aït-Sahalia and Lo, observing that option prices depend from the underlying price $S_T$ and the dividend yield $\delta$ only through the forward price $F_T = S_t e^{(r-\delta)\tau}$

which can be estimated using the model-independent call-put parity:

$$C_t(K) - P_t(K) = e^{-\tau r}(F_T - K)$$

Attempts to apply this idea to our dataset gave disappointing results, as asynchronous transactions on calls and puts substantially compromise accuracy. In fact, the estimated underlying price varies widely even in short periods of time, due to the relative illiquidity of the option market.

However, $F_T$ can be estimated from the future price, which is also part of our dataset, and is reported synchronously with option prices. When the expiration of the future contract coincides with that of the option, the estimation error reduces to the difference between the future and the forward prices and to the uncertainty in expected dividends. As the two expirations may differ at most for two months (since future contracts follow the quarterly cycle), the forward price is obtained discounting the future (we use the three-month libor), but expected dividends in the expiration lag cannot be eliminated, and add up to the estimation error.

Summing up, for each trading day we observe the cross sections of those options with the two nearest expiration months. For each expiration, we have a certain number of strikes (typically from 10 to 20), for which a call or a put option, or both, are available. We keep the contract with higher trading volume, which generally coincides with the one out-of-the-money (i.e. calls for high strikes, and puts for low strikes).

Denoting by $\mathcal{K}$ the set of strikes, for each $K \in \mathcal{K}$ we have an option price $P_K$, the corresponding underlying value $S_K$, and a dummy variable $F_K$, which is equal to 0 for a call option, and to 1 for a put option. With this notation, we can write the theoretical price $\Pi_K$ of a call or put option in the single formula:

$$\Pi_K = C(S_K, K, r, \tau, \sigma, \theta) - F_K(e^{-\delta \tau}S_K - e^{-\tau r}K)$$

which is more convenient, for estimation purposes, than a conditional statement. Then we specify the model as:

$$P_K = \Pi_K + \varepsilon_K$$
where $\{\varepsilon_K\}_{K \in K}$ are IID random variables. The parameters $\sigma$ and $\theta$ can then be estimated by the least-squares method:

$$
\chi^2(\sigma, \theta) = \sum_{K \in K} (P_K - \Pi_K(F_K, S_K, K, r, \tau, \sigma, \theta))^2
$$

$$(\hat{\sigma}, \hat{\theta}) = \arg \min_{\theta, \sigma} \chi^2(\sigma, \theta)
$$

As mentioned before, the problem above is nonlinear, but only in the parameter $\sigma$. This means that it can be solved easily even without nonlinear regression software. In fact, one can define:

$$
\phi(\sigma) = \min_{\theta} \chi^2(\sigma, \theta)
$$

and minimize $\phi$ with a standard one-dimensional minimization algorithm (the golden search, for instance), while $\phi(\sigma)$ can be computed explicitly. A natural starting guess for $\sigma$ is the implied volatility of the at-the-money option.

If nonlinear regression software is available, all the parameters can be estimated simultaneously. Since the sum of squares $\chi^2$ is quadratic in $\{\theta_i\}$, the convergence is faster with the Levenberg-Marquardt algorithm, than with the ordinary steepest descent method.

At this point, it remains to select an appropriate set of $\theta_i$. Since each cross section consists roughly of 10 to 20 prices, it is clear that precision can only be achieved if a very small number of $\theta$ is used.

As remarked before, the choice of an appropriate set of $\theta_i$ can be guided by moment considerations on the risk-neutral density, as the first $n$ terms of the Hermite expansion uniquely determine its first $n$ moments. Since we are constrained by the dataset to a small number of parameters, we choose to restrict our attention to the first four moments. This still leaves a total of five parameters, namely $\sigma$, $\theta_1$, $\theta_2$, $\theta_3$ and $\theta_4$. Not surprisingly, the simultaneous estimation of all parameters leads to unsatisfactory results, as the size of the data is quite limited. Other attempts showed that the elimination of only one parameter would not produce a significant improvement, therefore we opted to leave only three parameters free. As two of them must be chosen out of $\theta_1$, $\theta_2$, $\theta_3$ and $\theta_4$, there are six possible combinations.

The first combination to be ruled out is $(\theta_2, \theta_4)$, as it can represent only symmetric distributions. The combinations $(\theta_1, \theta_2)$ and $(\theta_2, \theta_3)$ can also be dropped, since they force negative excess kurtosis in a neighborhood of $(0, 0)$ (which is the typical domain of these estimators). We are thus left with the three combinations: $(\theta_1, \theta_3)$, $(\theta_1, \theta_4)$, $(\theta_3, \theta_4)$. While all of them are acceptable with respect to the above considerations, there are a few differences which
are worth noting. In fact, expressing skewness and kurtosis with respect to the two parameters, and expanding in a neighborhood of \((0, 0)\), we obtain the following results:

Table 1: Series expansion of skewness and kurtosis

<table>
<thead>
<tr>
<th>((x, y))</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>((\theta_1, \theta_3))</td>
<td>(-6y + O(x^2y + x^3))</td>
<td>(3 - 24xy + O(x^3y))</td>
</tr>
<tr>
<td>((\theta_1, \theta_4))</td>
<td>(-2x^3 + O(x^4))</td>
<td>(3 + 24y + O(x^2y))</td>
</tr>
<tr>
<td>((\theta_3, \theta_4))</td>
<td>(-6x)</td>
<td>(3 + 24y)</td>
</tr>
</tbody>
</table>

The table above shows that the parametrization \((\theta_3, \theta_4)\) has two clear advantages over the others. First, skewness and kurtosis not only are linear on the parameters, but depend separately on each of them. Contrast this with the other cases, where either skewness or kurtosis depend on higher order terms, and boil down to zero for typical parameter values.

A further advantage of the last parametrization lies in the separation between the roles of \(\sigma\) and \(\theta_i\). In fact, while most of the variance is generally captured by \(\sigma\), the parameters \(\theta_1\) and \(\theta_2\) may still explain a part of it. When both of them are set equal to zero, all the variance must be explained by \(\sigma\), resulting in a higher accuracy for this estimator.

Finally, if two random variables \(\psi_1\) and \(\psi_2\) have densities \(q_1\) and \(q_2\) with Hermite coefficients \(\{1, 0, 0, \theta_1^3, \theta_1^4\}\) and \(\{1, 0, 0, \theta_2^3, \theta_2^4\}\), Proposition 3 implies that the random variable \((\gamma_1 \psi_1 + \gamma_2 \psi_2)/\sqrt{\gamma_1^2 + \gamma_2^2}\) has the following Hermite expansion:

\[
\{1, 0, 0, \frac{\theta_1^1 \gamma_1^3 + \theta_2^2 \gamma_2^3}{\gamma^3}, \frac{\theta_1^4 \gamma_4^1 + \theta_2^4 \gamma_2^4}{\gamma^4}, 0,\]
\[
\frac{\theta_1^3 \theta_3^1 \gamma_3^1 \gamma_2^3}{\gamma^6}, \frac{\theta_2^1 \theta_2^4 \gamma_1^4 \gamma_2^3 + \theta_3^1 \theta_3^4 \gamma_3^3 \gamma_2^4}{\gamma^7}, \frac{\theta_2^2 \theta_1^1 \gamma_1^4 \gamma_2^4}{\gamma^8}\}
\]

where \(\gamma = \sqrt{\gamma_1^2 + \gamma_2^2}\). In other words, we obtain that the fifth coefficient is null, regardless of the parameter values. Therefore, the error which results from neglecting the terms after the fourth involves only the coefficients from six to eight.
4.3 Numerical Results

As a matter of fact, the choice of \((\theta_3, \theta_4)\) performs better in empirical tests, as it is shown by the following table:

Table 2: Estimated standard deviation of option prices (index points)

<table>
<thead>
<tr>
<th></th>
<th>First expiration</th>
<th>Second expiration</th>
</tr>
</thead>
<tbody>
<tr>
<td>((\theta_1, \theta_3))</td>
<td>34.41</td>
<td>115.47</td>
</tr>
<tr>
<td>((\theta_1, \theta_4))</td>
<td>32.55</td>
<td>130.14</td>
</tr>
<tr>
<td>((\theta_3, \theta_4))</td>
<td>31.16</td>
<td>94.48</td>
</tr>
</tbody>
</table>

The table also shows that the standard error is over three times larger for the second expiration, which is less liquid.

Event at this stage, parameters still exhibit some instability, due to the small number of strikes available. In practice, the functional \(\chi^2\) exhibits large flat regions, where a small perturbation in the data causes a wide fluctuation in the minimizers. This problem can be circumvented adding to the \(\chi^2\) functional a small stability term which discourages large fluctuations from the previous value. While this term is generally negligible with respect to the sum of squared errors, it becomes significant in a flat region, leading the estimators to move as little as possible. A convenient choice can be:

\[
\tilde{\chi}^2(\sigma, \theta) = \chi^2(\sigma, \theta) + (\alpha|\sigma - \tilde{\sigma}|^2 + \beta|\theta - \tilde{\theta}|^2)^2
\]

for suitable values of the parameters \(\alpha\) and \(\beta\). Here \(\tilde{\sigma}\) and \(\tilde{\theta}\) are the previous values of the estimators. The choice of the particular functional above can be justified in terms of ease of implementation, as it can be embedded in a least-squares framework by adding a further dummy variable \(\lambda\), and an additional set of data, with a dummy strike. \(\lambda\) is then set equal to 1 for regular data, and to 0 for the additional one. We set:

\[
f(\sigma, \theta) = \lambda(P_K - \Pi_K) + (1 - \lambda)(\alpha|\sigma - \tilde{\sigma}|^2 + \beta|\theta - \tilde{\theta}|^2)
\]

so that:

\[
\tilde{\chi}^2(\sigma, \theta) = \sum_{K \in \tilde{K} \cup \{\tilde{K}\}} f(\sigma, \theta)^2
\]

where the dummy strike \(\tilde{K}\) can take any value.
The first two set of figures show the values of volatility (\(\sigma\)), skewness and kurtosis for the first two expirations, estimated with or without the smoothing term in the functional: \(\alpha\) and \(\beta\) were both set equal to 10. While the addition of the penalization term greatly reduces the variance of the estimators, virtually eliminating outliers, in principle it may create a bias. Calculating the differences between the two estimators (with or without smoothing), and discarding those values lying outside the centered 95% confidence interval, it turns out that the average biases on volatility, skewness and kurtosis are respectively \(-1.2\%\), \(-71.0\%\) and \(4.8\%\) of the parameter averages. In other words, the bias is not serious for volatility and excess kurtosis, while it is significant for skewness.

The third set of figures shows the estimates for a moving horizon of one-month, obtained using the procedure in section 3 from the estimates on the first two expirations. In the last set we have the graphs of the parameters estimates versus the index. In the sample period there have been two major events affecting the Italian market: the admission to the core group of countries participating to the EMU, and the global crisis of world markets following the default of Russia. The first event caused a strong rally in the Italian index, on anticipation of the admission to the EMU, followed by a sharp drop during April. In this case, the option market correctly anticipated the chance of large movements both implied volatility and kurtosis rising from mid March, and with skewness decreasing over the same period. On the contrary, the Russian crisis, which caused a much larger drop in the index, was not anticipated at all by market participants, as volatility began to rise only at the end of August, kurtosis continued to shrink until the end of October, and skewness even rose all along the crisis.

References


Table 3: First terms of $C_n$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$C_n(S, K, r, \tau, \sigma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$e^{-\delta \tau} S \Phi(d_1) - e^{-\tau r} K \Phi(d_2)$</td>
</tr>
<tr>
<td>1</td>
<td>$-e^{-\delta \tau} S \sigma \sqrt{\tau} \Phi(d_1)$</td>
</tr>
<tr>
<td>2</td>
<td>$d_2 e^{-\tau r} K \sigma \sqrt{\tau} \phi(d_2) + e^{-\delta \tau} S \sigma^2 \tau \Phi(d_1)$</td>
</tr>
<tr>
<td>3</td>
<td>$\sigma \sqrt{\tau} \left( e^{-\tau r} K (-1 + d_2^2 - d_2 \sigma \sqrt{\tau} \phi(d_2)) - e^{-\delta \tau} S \sigma^2 \tau \Phi(d_1) \right)$</td>
</tr>
<tr>
<td>4</td>
<td>$e^{-\tau r} K \sigma \sqrt{\tau} \left( d_2^3 + \sigma \sqrt{\tau} - d_2^2 \sigma \sqrt{\tau} + d_2 (-3 + \sigma^2 \tau) \right) \phi(d_2) + e^{-\delta \tau} S \sigma^4 \tau^2 \Phi(d_1)$</td>
</tr>
<tr>
<td>5</td>
<td>$\sigma \sqrt{\tau} \left( e^{-\tau r} K \left( 3 + d_2^4 - d_2^2 \sigma \sqrt{\tau} - \sigma^2 \tau + d_2^3 (-6 + \sigma^2 \tau) + d_2 \left( 3 \sigma \sqrt{\tau} - \sigma^3 \tau^2 \right) \right) \phi(d_2) - e^{-\delta \tau} S \sigma^4 \tau^2 \Phi(d_1) \right)$</td>
</tr>
</tbody>
</table>

\[
d_1 = \log \frac{S_t}{K} + \left( r - \delta + \frac{\sigma^2}{2} \right) \tau \left( \sigma \sqrt{\tau} \right) \\
d_2 = \log \frac{S_t}{K} + \left( r - \delta - \frac{\sigma^2}{2} \right) \tau \left( \sigma \sqrt{\tau} \right) \\
\phi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \\
\Phi(x) = \int_{-\infty}^{x} \phi(t) dt\]
Second expiration

Volatility

Skewness

Excess kurtosis
Constant horizon

Volatility

Skewness

Excess kurtosis