



A multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices

Mark B. Shackleton^a, Stephen J. Taylor^{a,*}, Peng Yu^b

^a Department of Accounting and Finance, Lancaster University, UK

^b HSBC Bank, London, UK

ARTICLE INFO

Article history:

Received 8 October 2008

Accepted 11 May 2010

Available online 17 May 2010

JEL classification:

C14

C22

C53

G13

Keywords:

ARCH models

Density forecasts

Index options

Risk-neutral densities

Risk-transformations

ABSTRACT

We compare density forecasts of the S&P 500 index from 1991 to 2004, obtained from option prices and daily and 5-min index returns. Risk-neutral densities are given by using option prices to estimate diffusion and jump-diffusion processes which incorporate stochastic volatility. Three transformations are then used to obtain real-world densities. These densities are compared with historical densities defined by ARCH models. For horizons of two and four weeks the best forecasts are obtained from risk-transformations of the risk-neutral densities, while the historical forecasts are superior for the one-day horizon; our ranking criterion is the out-of-sample likelihood of observed index levels. Mixtures of the real-world and historical densities have higher likelihoods than both components for short forecast horizons.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Density predictions provide decision takers with more information than forecasts of expected returns and volatilities. This additional information is essential for many risk management activities. Central banks are prominent users of density predictions for interest rates, exchange rates, stock indices and commodity prices. They tend to prefer forward-looking densities obtained from option prices to conditional densities calculated from historical time series. Our paper provides the first comparison of the predictive accuracy of historical and option-based density forecasts across several forecast horizons.

As option forecasts of index volatility are often more accurate than historical forecasts, even when these are based upon intraday returns (Blair et al., 2001; Martens and Zein, 2004; Jiang and Tian, 2005), we might anticipate that a similar conclusion applies to density forecasts. We find, however, that our conclusions from comparing real-world density forecasts obtained from option prices with historical forecasts obtained from daily and intraday returns depend upon the forecast horizon. Historical forecasts rank the highest for

the one-day horizon and their performance is similar to option-based forecasts for the one-week horizon. Option-based forecasts are superior for horizons of two and four weeks. Furthermore, weighted combinations of historical and option densities outperform densities obtained from only one of the two sources of price information for the shortest horizons of one-day and one-week.

Option prices reflect competitive opinions about the risk-neutral density of the underlying asset when the option contracts expire. Several empirical methods convert option prices into an estimated risk-neutral density for *one* expiry date.¹ The more difficult problem of estimating the implied risk-neutral dynamics, from recent option prices for several expiry dates, has received much less attention. The first comprehensive empirical study is the pricing and hedging paper by Bakshi et al. (1997). They summarize daily estimates of jump-diffusion parameters for the S&P 500 index from 1988 to 1991. Likewise, we estimate risk-neutral parameters from S&P 500 futures prices on each day from 1990 to 2004 inclusive. It is then easy to derive the risk-neutral density for *any* time horizon.²

¹ See, for example, Jackwerth and Rubinstein (1996), Melick and Thomas (1997), Ait-Sahalia and Lo (1998), Bliss and Panigirtzoglou (2002) and Taylor (2005).

² The local volatility models of Dupire (1994) and Derman and Kani (1994) also provide risk-neutral densities for multiple horizons. However, their assumption that volatility is a deterministic function of time and the underlying asset price is counterfactual and it has been criticized by Dumas et al. (1998).

* Corresponding author. Tel.: +44 1524 593624; fax: +44 1524 847321.

E-mail addresses: m.shackleton@lancaster.ac.uk (M.B. Shackleton), s.taylor@lancaster.ac.uk (S.J. Taylor), peng.yu@hsbcib.com (P. Yu).

Transformations from risk-neutral to real-world³ densities have been estimated in several recent papers, commencing with Bakshi et al. (2003) and Bliss and Panigirtzoglou (2004), but for option expiry dates alone. Our first contribution is to obtain real-world densities from option prices for general forecast horizons; we evaluate forecasts for seven horizons that range from 1-day to 12-weeks.

All prior studies use full-sample datasets to estimate risk-transformations. The real-world densities are then *ex post*, because each density then depends on later asset prices. Our second contribution is to present results for *ex ante*, real-world densities constructed from the present and past prices for an asset and its options alone.⁴

There is a vast literature which compares volatility forecasts obtained from historical asset prices and current option prices.⁵ In contrast, we are only aware of two prior studies that make similar comparisons for density forecasts, namely Anagnou-Basioudis et al. (2005) and Liu et al. (2007) for small samples of forecasts for option expiry dates. Our third contribution is to compare ARCH and option-based forecasts for multiple horizons. These comparisons are the first to include results for historical density forecasts obtained from intraday returns.

Our methodology requires us to specify a risk-neutral process for the underlying asset price, whose parameters can be estimated rapidly from daily panels of option prices. An appropriate process for a stock index must incorporate a stochastic volatility component, whose increments are correlated with price increments. The price dynamics of Heston (1993) satisfy all of our requirements and they provide tractable formulae for densities and option prices, based upon the inversion of characteristic functions.

Like all parsimonious price models, the Heston model has been shown to be imperfect for pricing options.⁶ We also evaluate an affine process that includes price jumps generated by a Poisson process.⁷ We do not consider further jump processes for three reasons: firstly, it is difficult to estimate the additional parameters from daily panels of option prices; secondly, our transformations from risk-neutral to real-world densities are able to systematically improve mis-specified risk-neutral densities; and thirdly we find that incorporating risk-neutral price jumps does not lead to improved real-world densities.

It is possible that the sophistication of the Heston price dynamics and its jump-diffusion extensions may be counterproductive if the final goal is to produce real-world densities. Consequently, we also investigate transformations of risk-neutral, lognormal densities.

The positive risk premium for the aggregate equity market shows that some transformation must be applied to risk-neutral densities before appropriate, real-world, density forecasts can be made. Bliss and Panigirtzoglou (2004) evaluate single-parameter, utility transformations that can be motivated by a representa-

tive-agent model. They find that *ex post*, real-world densities for the S&P 500 and FTSE 100 indices are a significant improvement upon their risk-neutral densities; Anagnou-Basioudis et al. (2005), Kang and Kim (2006) and Liu et al. (2007) provide further utility-based results for these markets. However, as empirical estimates of implied risk aversion are incompatible with a standard consumption-based framework (Jackwerth, 2000; Rosenberg and Engle, 2002; Ziegler, 2007), standard utility transformations are unlikely to provide completely satisfactory real-world densities.

Liu et al. (2007) estimate the two parameters of a more flexible transformation, by maximizing the *ex post* likelihood of the index levels on monthly, option expiry dates. We apply the same transformation, but instead use *ex ante* parameters obtained separately for each forecast horizon. We also provide the first analysis of two further transformations: one assumes that affine (jump-) diffusion price dynamics apply in the real world by incorporating appropriate risk-premia, and the other applies a nonparametric calibration function.

All our density forecasting methods are described in Section 2. We consider risk-neutral (Q) densities that are either lognormal or provided by affine (jump-) diffusion price dynamics, real-world (P) densities given by the three transformations of the Q -densities, historical densities obtained from ARCH models that are estimated from daily and intraday returns, and mixture densities that use all of the information derived from historical and option prices. The econometric methodology used to obtain *ex ante* parameters and forecasts is presented in Section 3 and Appendix A. We also present our criteria for making out-of-sample comparisons between the various sets of density forecasts. The S&P 500 futures and options price data are described in Section 4, followed by all the empirical results in Section 5. Finally, Section 6 summarizes our conclusions.

2. Density forecasts

2.1. Risk-neutral densities

To obtain risk-neutral densities (RNDs) for *all* future times it is necessary to specify the risk-neutral dynamics of the underlying asset price. We consider three specifications for the *futures* price. The first simply assumes that prices follow geometric Brownian motion (GBM). All the RNDs are then lognormal. The second specifies a risk-neutral volatility process, while the third additionally incorporates price jumps.

The stochastic volatility process of Heston (1993) is a natural candidate because it has closed-form densities and theoretical option prices, whose implied volatilities display plausible “term structure” and “smile” effects. The risk-neutral dynamics for the futures price, p_t after t years, incorporate the stochastic variance V_t , which follows a square-root process:

$$dp/p = \sqrt{V}dW_1, \quad (1)$$

$$dV = \kappa(\theta - V)dt + \xi\sqrt{V}dW_2 \quad (2)$$

with correlation ρ between the increments of the two Wiener processes. The special case of GBM, with constant volatility θ , occurs when $V_0 = \theta$ and $\xi = 0$.

It is well-known that adding a jump component to (1) enhances the agreement between theoretical and observed option prices. Following Bates (1996) and Bakshi et al. (1997), we also evaluate the affine jump-diffusion defined by (2) and

$$dp/p = \sqrt{V}dW_1 + (e^J - 1)dN - \lambda\bar{\mu}_J dt \quad (3)$$

with N_t a Poisson process that has intensity λ ; the Poisson process is independent of the bivariate Wiener process $(W_{1,t}, W_{2,t})$. The jump events counted by N_t are matched with jumps of size J_t in $\log(p_t)$, that are normally distributed with mean μ_J and variance σ_J^2 ; the

³ Like Liu et al. (2007), we prefer ‘real-world’ to alternative adjectives, such as ‘subjective’, ‘objective’, ‘statistical’, ‘empirical’, ‘physical’, ‘true’, ‘risk-adjusted’ and ‘historical’. We use ‘historical’ to refer to densities that are obtained from time series of prices for the underlying asset.

⁴ Kostakis et al. (2009) provide *ex ante* comparisons for historical and real-world densities in a recent paper about asset allocation.

⁵ Recent evaluations of forecasts obtained from implied volatilities include Konstantinidi et al. (2008), Chalamandaris and Tsekrekos (2010), Taylor et al. (2010) and Yu et al. (2010).

⁶ For example by Bakshi et al. (1997), Andersen et al. (2002), Pan (2002), Jones (2003), Christoffersen et al. (2006) and Ait-Sahalia and Kimmel (2007). Also, Christoffersen et al. (2008, in press) show that two volatility components fit option prices more accurately than one component.

⁷ Special cases of the general affine jump-diffusions of Duffie et al. (2000) are investigated in Bates (1996, 2000, 2006), Bakshi et al. (1997), Andersen et al. (2002), Bollerslev and Zhou (2002), Pan (2002), Liu and Pan (2003), Eraker et al. (2003), Eraker (2004), Broadie et al. (2007) and Medvedev and Scaillet (2007). Processes incorporating jumps that arrive at an infinite rate are covered by Carr and Wu (2004) and Huang and Wu (2004).

average size of the proportional jumps in p_t equals $\bar{\mu}_j = \exp(\mu_j + 0.5\sigma_j^2) - 1$.

Several futures contracts, with different expiry dates, are traded at the same time. We suppose that their prices satisfy standard, no-arbitrage, equations; these imply the same continuous-time process and the same parameters are applicable to all contracts.

Heston (1993), Bakshi et al. (1997) and Duffie et al. (2000) provide analytic formulae for the characteristic function of $\log(p_T)$, conditional upon initial values p_0 and V_0 . Our notation for this conditional characteristic function is $g(\psi) = E^Q[\exp(i\psi \log(p_T))]$, with ψ a real number and Q the risk-neutral measure. The following inversion formula then gives the risk-neutral density of p_T , denoted by $f_{Q,T}(x)$, for positive values of x :

$$f_{Q,T}(x) = \frac{1}{\pi x} \int_0^\infty \text{Re}[\exp(-i\psi \log(x))g(\psi)]d\psi. \tag{4}$$

A straightforward numerical integration is required for each value of x . The fair price of a European call option, whose strike is K , can be written as:

$$c(p_0, K) = e^{-rT}(p_0P_1(p_0, K) - KP_2(p_0, K)), \tag{5}$$

where r is the risk-free rate, $P_2(p_0, K)$ is the risk-neutral probability that the option expires in-the-money and $P_1(p_0, K)$ is a probability for the same event when a different measure is applied. Both $P_1(p_0, K)$ and $P_2(p_0, K)$ are obtained from standard inversion formulae.

2.2. Real-world densities

A risk-neutral density should not be used to make statements about real-world probabilities determined by a real-world measure P , because investors require a premium that compensates them for price risk. It is possible that there are also volatility and jump risk premia. Transformations from risk-neutral to real-world densities rely on assumptions. These can be provided by specifying risk-premia functions, or by a representative-agent model, or by statistical calibration theory; Liu et al. (2007) show that there is a simple mathematical relationship between the marginal utility function of the representative agent and the statistical calibration function. We prefer the additional flexibility provided by either two or three risk-premium terms, or by a two-parameter calibration transformation, to the one-parameter utility transformations of Bliss and Panigirtzoglou (2004). We also investigate a nonparametric calibration transformation.

2.2.1. Risk-premia transformations

An affine real-world diffusion process is defined by including linear drift terms in both the price and the variance equations, thus:

$$\begin{aligned} dp/p &= \eta_1 V dt + \sqrt{V} d\tilde{W}_1, \\ dV &= [\eta_2 V + \kappa(\theta - V)]dt + \xi \sqrt{V} d\tilde{W}_2. \end{aligned} \tag{6}$$

The assumption of linear functions for the risk premia ensures analytic formulae for the real-world, characteristic functions of future prices. The inversion formula (4) then provides real-world densities $f_{P,T}(x)$ that depend on the parameters η_1 and η_2 . To define real-world, jump-diffusion dynamics we additionally increase the mean jump size by η_3 . Then

$$dp/p = (\eta_1 V - \lambda \bar{\mu}_j)dt + \sqrt{V} d\tilde{W}_1 + (e^J - 1)dN, \quad J_t \sim N(\mu_j + \eta_3, \sigma_j^2). \tag{7}$$

2.2.2. Calibration transformations

At time 0, suppose $f_{Q,T}(x)$ and $F_{Q,T}(x)$, respectively, denote the risk-neutral density and the cumulative distribution function

(c.d.f.) of the random variable p_T , and then define $U_T = F_{Q,T}(p_T)$.⁸ Following Bunn (1984), Dawid (1984) and Diebold et al. (1999), let the calibration function $C_T(u)$ be the real-world c.d.f. of the random variable U_T ; our notation emphasizes that the calibration function depends on the forecast horizon T . It is then well-known (see, for example, Liu et al. (2007)) that the real-world c.d.f. of p_T is

$$F_{P,T}(x) = C_T(F_{Q,T}(x)). \tag{8}$$

Also, the real-world density of p_T is given by

$$f_{P,T}(x) = f_{Q,T}(x)c_T(u) \tag{9}$$

with $u = F_{Q,T}(x)$ and with $c_T(u)$ representing the real-world density of U_T .

Our preferred parametric specification of the calibration function is the c.d.f. of the Beta distribution, applied by Liu et al. (2007) and Wang (2009) and recommended by Fackler and King (1990) in their innovative study of densities obtained from commodity option prices. The calibration density is then

$$c_T(u) = u^{a-1}(1-u)^{b-1}/B(a,b), \quad 0 \leq u \leq 1 \tag{10}$$

with $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. There are two calibration parameters, a and b , which depend on the horizon T . The special case $a = b = 1$ defines a uniform distribution and then the risk-neutral and real-world densities are identical. From (9), the real-world density is

$$f_{P,T}(x) = \frac{F_{Q,T}(x)^{a-1}(1-F_{Q,T}(x))^{b-1}}{B(a,b)}f_{Q,T}(x). \tag{11}$$

Alternatively, a nonparametric calibration function can be estimated from a historical set of observations of the quantity $u = F_{Q,T}(x)$, with x representing a typical observed value of the futures price p_T . We calculate kernel estimates of $C_T(u)$ and $c_T(u)$, using methods presented in Appendix A. These estimates, given by (34) and (37), are substituted into (8) and (9) to provide further real-world c.d.f.s and densities.⁹

2.3. Historical densities

By estimating ARCH models, the prices of the underlying asset up to and including time t can be used to produce historical density forecasts for the asset price at time $t + 1$. One period of time is defined by a constant forecast horizon in this section, that may be one-day, one-week or several weeks. The one-period returns are $r_t = \log(p_t/p_{t-1})$; here p_{t-1} and p_t are end-of-period futures prices for the same contract.

A specific ARCH model uses price information I_t , known at the end-of-period t , to produce a parametric density, $f_{ret}(r|I_t)$, for the next return, r_{t+1} . The historical density for the next end-of-period price, p_{t+1} , is then:

$$f(x|I_t) = f_{ret}(r|I_t)/x \tag{12}$$

with $r = \log(x/p_t)$. We describe four specifications for the historical density $f_{ret}(r|I_t)$.

The simplest credible ARCH model for a stock market index is the GJR(1, 1) model of Glosten et al. (1993). The conditional

⁸ For any random variable X , with c.d.f. $F(X)$ for a stated measure, the random variable $U = F(X)$ has a uniform distribution for the same measure. This result enables the correct specification of densities to be assessed empirically (Rosenblatt, 1952). The observed value u of $U_T = F_{Q,T}(p_T)$ is only a draw from a uniform distribution when $F_{Q,T}(x)$ is correctly specified and the risk-neutral measure is identical to the real-world measure.

⁹ Kling and Bessler (1989) and Diebold et al. (1999) have estimated nonparametric calibration functions from real-world densities, respectively for Treasury bill and foreign exchange rates.

variance h_t is then an asymmetric function of returns. We define the model as follows, with a constant conditional mean μ :

$$r_t - \mu = h_t^{1/2} z_t, \quad z_t \sim \text{i.i.d. } (0, 1),$$

$$\frac{h_t}{D_t} = \omega + \frac{(\alpha_1 + \alpha_2 d_{t-1}) \sigma_{t-1}^2 + \beta h_{t-1}}{D_{t-1}}, \quad (13)$$

$$d_{t-1} = 1 \quad \text{if } r_{t-1} < \mu, \quad \text{otherwise } d_{t-1} = 0.$$

The term D_t represents the number of trading days during period t , so that the conditional variance is proportional to the amount of trading time. Normal distributions for the i.i.d., standardized residuals z_t define our GJR model. As it is well-known that fat-tailed, conditional distributions are preferable for daily horizons, we also evaluate the GJR- t model defined by supposing the z_t have a standardized t -distribution, with degrees-of-freedom ν .¹⁰

Sums of squared intraday returns are superior to squared daily returns as measures of realized volatility (Andersen and Bollerslev, 1998; Andersen et al., 2001) and these sums can be used to improve volatility forecasts (Blair et al., 2001; Martens and Zein, 2004). Let $Intra_t$ represent the total of some set of squared intra-period returns for period t . Then the Intra and Intra- t models are here defined by the conditional variance equation:

$$\frac{h_t}{D_t} = \omega_{Intra} + \frac{(\gamma_1 + \gamma_2 d_{t-1}) Intra_{t-1} + \beta_{Intra} h_{t-1}}{D_{t-1}} \quad (14)$$

with, respectively, conditional normal distributions and conditional t -distributions. As in (13), the multiplier of the most recent volatility measurement ($Intra_{t-1}$) is an asymmetric function of the most recent excess return ($r_{t-1} - \mu$).

2.4. Mixture densities

Both ARCH and option-based densities may contain incremental information at a general time t about the price at time $t + T$. Consequently, we also evaluate the mixture (or combination) density:

$$f_{mix,T}(x) = \alpha f_{p,T}(x) + (1 - \alpha) f_{ARCH,T}, \quad 0 \leq \alpha \leq 1. \quad (15)$$

One component of the mixture may outperform the other during some periods and vice versa at other times. The mixture may then be preferred to its components. We let past data determine appropriate weights, α and $1 - \alpha$, in (15).

As option traders know the historical price information, it is possible that $\alpha = 1$ is optimal if the transformations are able to translate an “efficient” risk-neutral density into the best possible real-world density. At the other extreme, $\alpha = 0$ might occur if option prices contain no real-world information that is incremental to the historical record of asset prices.

3. Empirical methods

3.1. Estimation of parameters

The densities are all parametric and evaluated out-of-sample. We always use *ex ante* parameter values; all values required at time t are estimated from prices available at time t .

The parameters of the risk-neutral processes for asset prices are estimated at the end of each trading day. The estimated volatility of the GBM process is provided by the simplest credible estimate, namely the end-of-day, nearest-the-money implied volatility for the nearest-to-expiry options. On day t , for the Heston process defined by (1) and (2) we estimate the initial variance V_t , the three

volatility parameters, κ_t , θ_t and ξ_t , and the correlation ρ_t between the price and volatility differentials. Suppose N_t European, call¹¹ option contracts are traded on day t , labeled by $i = 1, \dots, N_t$, with strikes $K_{t,i}$, expiry times $T_{t,i}$ and market prices $c_{t,i}$; also, suppose $p_{t,i}$ is the futures price for the asset after $T_{t,i}$ years. Then the five Heston Q-parameters are estimated by minimizing

$$\sum_{i=1}^{N_t} (c_{t,i} - c(p_{t,i}, K_{t,i}, T_{t,i}, V_t, \kappa_t, \theta_t, \xi_t, \rho_t))^2, \quad (16)$$

with $c(\cdot)$ the Heston pricing formula, given by (5).¹² Likewise, the eight Q-parameters for the jump-diffusion process defined by (2) and (3) are also estimated. These are maximum likelihood estimates when the option pricing errors are Gaussian, independent and identically distributed.

The remaining parameters that appear in the real-world, historical and mixture densities are the price-of-risk parameters η_1 , η_2 and η_3 , the calibration function parameters a and b , the ARCH parameters ω , α_1 , α_2 , β , ω_{Intra} , γ_1 , γ_2 and β_{Intra} , and the mixture parameter α . As explained in Appendix A, all these parameters are estimated *ex ante* by maximizing the log-likelihood function of observed asset prices that are available when the forecasts are made. Separate estimates are obtained for each forecast horizon considered.

3.2. Evaluation of the density forecasts

Density forecasts can be assessed using a variety of criteria, including several surveyed by Tay and Wallis (2000). Christoffersen and Jacobs (2004) show that it is important to use the same criterion for calibration and evaluation in their study of option valuation. Our common criterion is the log-likelihood function, previously used by Bao et al. (2007) and Liu et al. (2007) to compare density forecasting methods applied to equity indices.

3.2.1. Likelihood criteria

For a fixed forecast horizon, suppose method m provides densities $f_{m,t}(x)$ at integer times i, \dots, j for the asset price at times $i + 1, \dots, j + 1$. Our preferred method maximizes the out-of-sample, log-likelihood of observed asset prices, defined for method m by:

$$L_m = \sum_{t=i}^j \log(f_{m,t}(p_{t+1})). \quad (17)$$

When some method M correctly specifies the real-world densities it will have the highest expected log-likelihood, as the information criterion of Kullback and Leibler (1951), namely

$$E^p[\log(f_{M,t}(p_{t+1})/f_{m,t}(p_{t+1}))] = \int_0^\infty f_{M,t}(x) \log(f_{M,t}(x)/f_{m,t}(x)) dx, \quad (18)$$

is positive whenever the densities $f_{M,t}$ and $f_{m,t}$ are continuous and distinct. Consequently, $E^p[L_M] > E^p[L_m]$ for $m \neq M$ and we may expect the sample value of L_M to exceed that of L_m when the number of forecasts made is sufficiently large. When none of the methods correctly specifies the densities, maximizing L_m will select the method whose densities are nearest to the true densities according to the information criterion (Bao et al., 2007).

The evidence for one method relative to a set of alternatives can be stated as a Bayesian probability. Assuming that one of K meth-

¹⁰ We assume the density for z_t is time-invariant. Variation in the density's shape might enhance density forecasts; see Jondeau and Rockinger (2003, 2009) who provide evidence for time-varying conditional skewness and kurtosis in daily returns for the S&P 500 index.

¹¹ As explained in Section 4.3, the put prices in our database are converted to equivalent European call prices, using put-call parity, and are then included in (16).

¹² Christoffersen and Jacobs (2004, p. 316) conclude that (16) is a “good general-purpose loss function in option valuation applications”. It is preferred in the study of S&P 500 dynamics by Christoffersen et al. (2006).

ods is correct, and selecting an uninformative prior distribution, the posterior probability that method m is correct equals

$$\pi_m = \exp(L_m) / \sum_{k=1}^K \exp(L_k). \quad (19)$$

The out-of-sample, log-likelihood is a special case of the weighted log-likelihood criterion used by Amisano and Giacomini (2007) to test for differences between the accuracy of competing forecasts. The null hypothesis that two methods m and n have equal expected log-likelihood can be tested by using the time series average of the log-likelihood differences:

$$d_t = \log(f_{m,t}(p_{t+1})) - \log(f_{n,t}(p_{t+1})), \quad i \leq t \leq j. \quad (20)$$

We find only minor dependence within a typical time series $\{d_t\}$ and consequently compare the AG test statistic,

$$AG = (j - i + 1)^{0.5} \bar{d} / s_d = (j - i + 1)^{-0.5} (L_m - L_n) / s_d \quad (21)$$

with a standard normal distribution; here \bar{d} and s_d are the average and the standard deviation of the terms d_t .

3.2.2. Diagnostic criteria

We also evaluate diagnostic tests that use a time series of observed cumulative probabilities to assess the adequacy of a set of forecasts, as recommended by Diebold et al. (1998).¹³ For a general method m these probabilities are defined by

$$u_{t+1} = \int_0^{p_{t+1}} f_{m,t}(x) dx, \quad i \leq t \leq j. \quad (22)$$

We check whether or not the values of u are consistent with i.i.d. observations from the uniform distribution between zero and one.¹⁴ Firstly, the Kolmogorov–Smirnov (KS) test is used, which relies on the maximum difference between the sample and theoretical cumulative functions. The sample c.d.f. of $\{u_{i+1}, \dots, u_{j+1}\}$, evaluated at u , is the proportion of outcomes less than or equal to u , i.e.:

$$\tilde{C}(u) = \frac{1}{j - i + 1} \sum_{t=i+1}^{j+1} S(u - u_t) \quad (23)$$

with $S(x) = 1$ if $x \geq 0$, and $S(x) = 0$ if $x < 0$. The test statistic is then

$$KS = \sup_{0 \leq u \leq 1} |\tilde{C}(u) - u|. \quad (24)$$

Secondly, we apply the test of Berkowitz (2001) to the numbers y_t defined by $\Phi(y_t) = u_t$, with $\Phi(\cdot)$ the c.d.f. of the standard normal distribution. This test assesses whether or not the values of y are consistent with the null hypothesis of i.i.d. observations from a standard normal distribution. The alternative hypothesis for the test is a stationary, Gaussian, AR(1) process with no restrictions on the mean, variance and autoregressive parameters. The test is decided by comparing a likelihood-ratio statistic (LR3) with χ_3^2 ; LR3 equals $2(L_1 - L_0)$, with L_0 and L_1 the maximum log-likelihoods of $\{y_{i+1}, \dots, y_{j+1}\}$, respectively, for the null and alternative hypotheses.

4. Data

Density forecasts are evaluated for S&P 500 index futures contracts. We investigate forecasts for the futures price, rather than

¹³ Interesting recent examples are Bliss and Panigirtzoglou (2004), Hong et al. (2004), Anagnou-Basioudis et al. (2005), Christoffersen and Mazzotta (2005), Hong and Li (2005), Liu et al. (2007) and Wang (2009).

¹⁴ The null hypothesis is that the set of density forecasts is correctly specified. The more general hypothesis that the forecasts are from a correct parametric specification can only be assessed if the tests are revised to take account of parameter estimation error. This is difficult when the forecasts are defined *ex ante* using option prices.

the spot index, because contemporaneous, settlement prices are available for futures and options contracts.¹⁵ A second advantage of working with futures is that the dividend payments on the index stocks are irrelevant.

4.1. Futures prices

Settlement prices and intraday prices for S&P 500 futures contracts are studied from 28 April 1982 until 31 December 2004, respectively, provided by the CME and Price-Data.com. Each return is calculated from the nearest-to-maturity contract, except on the final trading days and on the Thursdays that precede them when the next contract is used.

The realized variances are calculated from 5-min returns. This frequency provides a satisfactory trade-off between maximizing the accuracy of volatility estimates and minimizing the bias attributable to microstructure effects (Bandi and Russell, 2006). As the futures are traded from 08:30 to 15:15 at the CME, we use 81 intraday returns for each day. The realized variance for day t is the sum of the squares of the 5-min returns $r_{t,i}$:

$$\text{Intra}_t = \sum_{i=1}^{81} r_{t,i}^2. \quad (25)$$

4.2. Interest rates

Three-month, six-month and one-year Treasury bill rates are converted to continuously compounded rates. The risk-free rate r used in an option pricing formula is the three-month rate for option lives up to three months, otherwise the rate is given by linear interpolation.

4.3. Option prices

We study the settlement prices of options on S&P 500 futures for 15 years, from 2 January 1990 to 31 December 2004. We consider prices for all option-on-futures contracts that expire on the same Friday as their underlying futures contracts, after excluding contracts having seven or less calendar days until expiry.

Call and put settlement prices for the same strike and expiry date theoretically contain the same information. Either the call or the put will be out-of-the-money (OTM), except for the rare occasions when both are at-the-money (ATM). We choose to use the information provided by the prices of OTM and ATM options alone, because the in-the-money contracts are less actively traded and have higher early exercise premia.

The option contracts are American. We obtain equivalent European option prices from the American prices which have the same implied volatility when the pricing formulae are those of Black (1976) and Barone-Adesi and Whaley (1987). The early exercise premia are small for OTM options and hence only very small errors can be created by applying these formulae. Finally, the put-call parity equation is used to obtain equivalent European call prices from the European OTM put prices.

We study 435,100 option prices for the 3777 trading days from 1990 to 2004. The average number of prices used per day is 115, made up of 45 OTM calls and 70 OTM puts. The number of different expiry dates available on any day is 2, 3 or 4 and their average is 3.1. Table 1 summarizes the quantity, the moneyness and the time-to-expiry of the contracts that provide the observed prices.

¹⁵ Option prices for S&P 500 futures have been studied in related research by Bates (2000, 2006), Bliss and Panigirtzoglou (2004), Jones (2006) and Broadie et al. (2007).

Table 1

Summary statistics for the S&P 500 futures option dataset. Information about the numbers of daily settlement prices for out-of-the-money (OTM) options on S&P 500 futures, from 1990 to 2004. Moneyyness is defined by the futures price p divided by the strike price K .

		Total number	Average options per day	Max number per day	Min number per day
Calls		171,383	45	157	8
Puts		263,717	70	173	8
Overall		435,100	115	255	29
Number of cross-sections			3.1	4	2
Moneyyness/maturity	p/K	<1 Month	Between 1 and 6 months	>6 Months	Subtotal
Deep OTM put	>1.10	10,800 (2.48%)	89,779 (20.63%)	39,879 (9.17%)	140,458 (32.28%)
OTM put	1.03–1.10	8743 (2.01%)	52,427 (12.05%)	23,964 (5.51%)	85,134 (19.57%)
Near the money	0.97–1.03	7720 (1.77%)	47,325 (10.88%)	20,206 (4.64%)	75,251 (17.30%)
OTM call	0.90–0.97	6881 (1.58%)	45,519 (10.46%)	19,178 (4.41%)	71,578 (16.45%)
Deep OTM call	<0.90	2483 (0.57%)	42,253 (9.71%)	17,943 (4.12%)	62,679 (14.41%)
Subtotal		36,627 (8.42%)	277,303 (63.73%)	121,170 (27.85%)	435,100 (100%)

4.4. Volatility comparisons

The intraday volatility measure excludes the price variation from the market's close until the market reopens. From 1990 until 2004, the average of the daily realized variances (0.93×10^{-4}) equals 80% of the variance of the daily returns (1.17×10^{-4}). The options imply higher levels of variance: the average of the squares of the implied volatilities obtained from the nearest-to-the-money options that are nearest-to-expiry, stated in daily units (1.32×10^{-4}), is 113% of the variance of the daily returns. Converted to annualized, standard deviations, these average measures of variance are 15.3% for Intra, 17.2% for daily returns and 18.3% for near-the-money options. The higher level for risk-neutral volatility is to be expected, because empirical evidence for a negative risk premium for the volatility of US equity indices has been documented in several papers.¹⁶

5. Empirical results for density forecasts

Density forecasts are evaluated from January 1991 until December 2004 inclusive. The option prices during 1990 are only used to contribute to the *ex ante* information that is required to estimate the transformations from risk-neutral to real-world densities.

The density forecasts are made for seven horizons: one trading day and one, two, four, six, eight and 12 weeks. The first forecast for each horizon is made on Wednesday, 3 January 1990. The forecasts for the multi-day horizons do not overlap and they are all made on Wednesdays. The forecast density at time t for time $t + T$ always refers to the first futures contract that matures at least one calendar day after time $t + T$.

5.1. Illustrative density plots

Fig. 1 illustrates the one-day-ahead densities, calculated from the information available on 30 December 2004. Fig. 1a shows the four historical densities; the conditional t -densities have higher peaks and fatter tails than the conditional normal densities. Figs. 1b and 1c, respectively, show how the risk-transformations change the shapes of the lognormal and the Heston risk-neutral densities. The labels $P1$, $P2$ and $P3$ for the real-world densities, respectively, refer to the parametric calibration, the nonparametric calibration and the risk-premium transformations. Each transformation increases

the peak of the density and decreases the probability of a large price change, consistent with the real-world density having a lower standard deviation than the risk-neutral density. Illustrative densities for the longer horizon of four weeks, calculated on 17 November 2004, are shown on Fig. 2a–c. The real-world standard deviations are again less than the risk-neutral levels.

5.2. Historical densities

The parameters of the one-day-ahead, ARCH densities have been estimated from daily and intraday returns that commence on 4 January 1988. Referring to (13), the averages of the *ex ante*, GJR parameter estimates, used in the densities from 1991 onwards, include $\alpha_1 = 0.032$, $\alpha_2 = 0.043$ and $\beta = 0.913$. For the more credible GJR- t specification, with degrees-of-freedom ν , the averages include $\nu = 4.70$, $\alpha_1 = 0.009$, $\alpha_2 = 0.046$ and $\beta = 0.960$. The corresponding averages for the Intra- t specification, given by (14), are $\nu = 5.03$, $\gamma_1 = 0.039$, $\gamma_2 = 0.157$ and $\beta_{intra} = 0.893$.

The ARCH densities for the one-week and longer periods are estimated from prices that commence on 28 April 1982. All the averages for the degrees-of-freedom parameter imply a high level of excess kurtosis in the conditional distributions. The averages of ν are 6.20 and 7.52 for the one-week returns, respectively, for GJR- t and Intra- t , and they are 4.95 and 4.55 for the longest return period of 12 weeks.

5.3. Risk-neutral parameters

Table 2 presents our summary statistics for risk-neutral parameters estimated each day from 1990 to 2004. One set of statistics is given for Heston's diffusion process, with the risk-neutral, price dynamics given by:

$$dp/p = \sqrt{v}dW_1 \quad \text{and} \quad dV = \kappa(\theta - V)dt + \xi\sqrt{V}dW_2 \quad (26)$$

and with correlation ρ between the two Wiener processes. The second set of statistics are for the affine jump-diffusion dynamics defined by (3), which adds Gaussian price jumps in $\log(p_t)$ that have intensity λ , mean μ_j and standard deviation σ_j . The median estimates in Table 2 are generally similar to risk-neutral estimates in previous research.

Our risk-neutral parameters minimize the mean of the squared errors (MSE) for each day's option prices. The median MSE for the eight-parameter jump-diffusion process equals 69% of the median MSE for the five-parameter pure-diffusion process. This reduction in the MSE occurs when the median estimate of λ is 0.47 jumps

¹⁶ For example, in Jackwerth and Rubinstein (1996), Chernov and Ghysels (2000), Pan (2002), Bakshi and Kapadia (2003), Jones (2003) and Bollerslev et al. (in press).

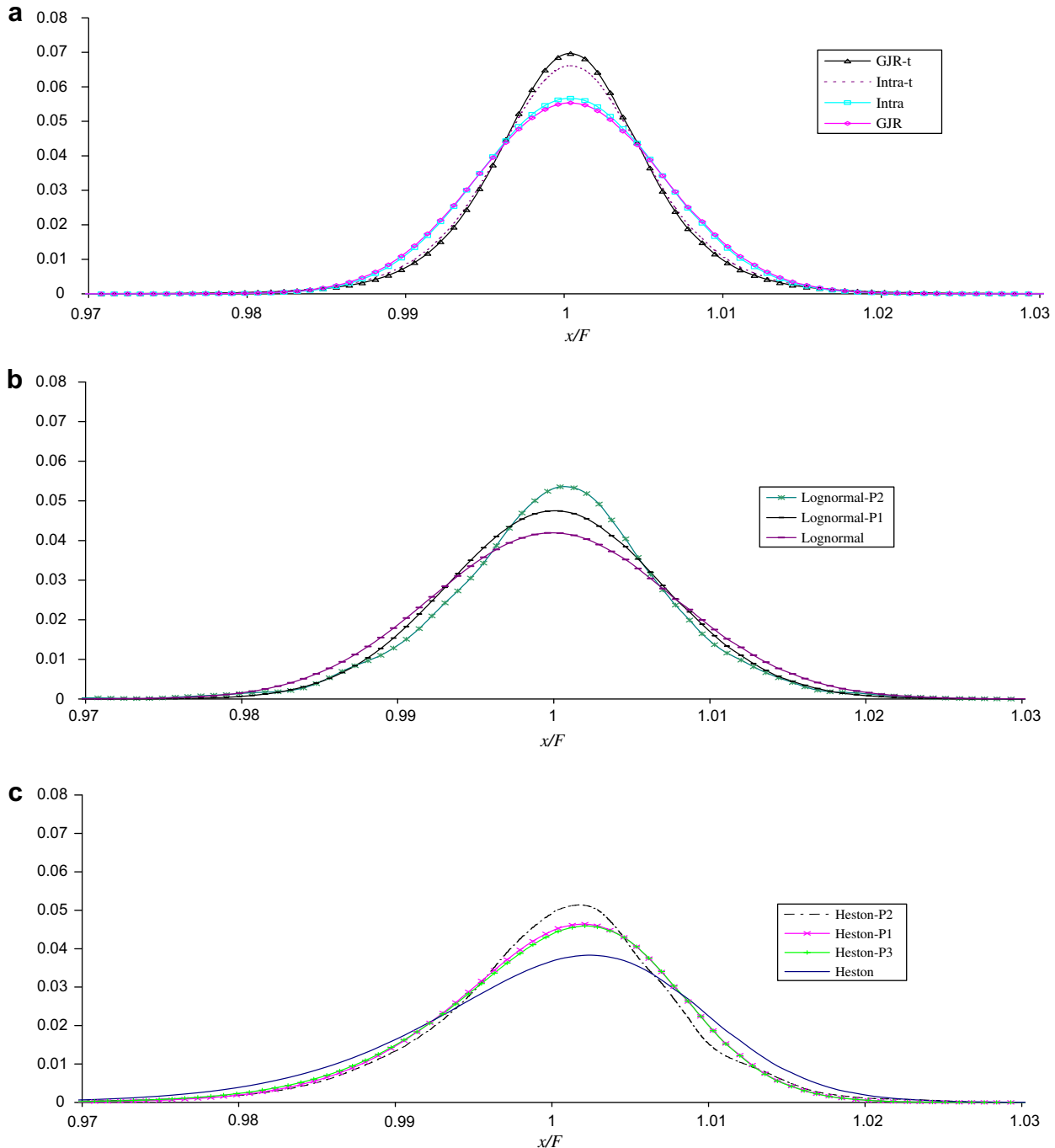


Fig. 1. (a) One-day ahead density forecasts obtained from ARCH models on December 30th, 2004. (b) One-day ahead density forecasts obtained from lognormal densities on December 30th, 2004. (c) One-day ahead density forecasts obtained from Heston's model on December 30th, 2004.

per annum and the median estimate of μ_j reduces the asset price by 6%. Our average risk-neutral estimate of λ is 0.69 which is similar to the estimates between 0.5 and 0.8 in Eraker (2004), obtained from index and option prices for an earlier period, while the higher value of 1.5 in Eraker et al. (2003) has been estimated from index returns between 1980 and 1999.¹⁷

Each time series of MSE values has a median which is much lower than the standard deviation, which indicates that the distri-

bution of MSE contains some extreme values. A positively skewed distribution is also noted for κ , λ and σ_j .

The stochastic variance V_t reverts towards θ . Our median estimate for the pure-diffusion process is 0.0452, equivalent to a volatility of 21.3%. The median estimate for the jump-diffusion process is lower, at 0.0347, because some of the total variation in prices is then attributed to the jump component. The rate of reversion towards θ is determined by κ . Our median estimates of κ are 4.15 (without jumps) and 3.09 (with jumps); the “half-life” parameter of the variance process is then between two and three months.

The kurtosis of returns is primarily controlled by the “volatility of volatility” parameter ξ . Estimates obtained solely from option prices, such as our median values of 0.79 (without jumps) and

¹⁷ The time-varying, real-world estimates of λ in Santa-Clara and Yan (2010) are similar. Their mean and standard deviation, from 1996 to 2002, are 0.795 and 0.714, while our statistics are 0.693 and 0.715.

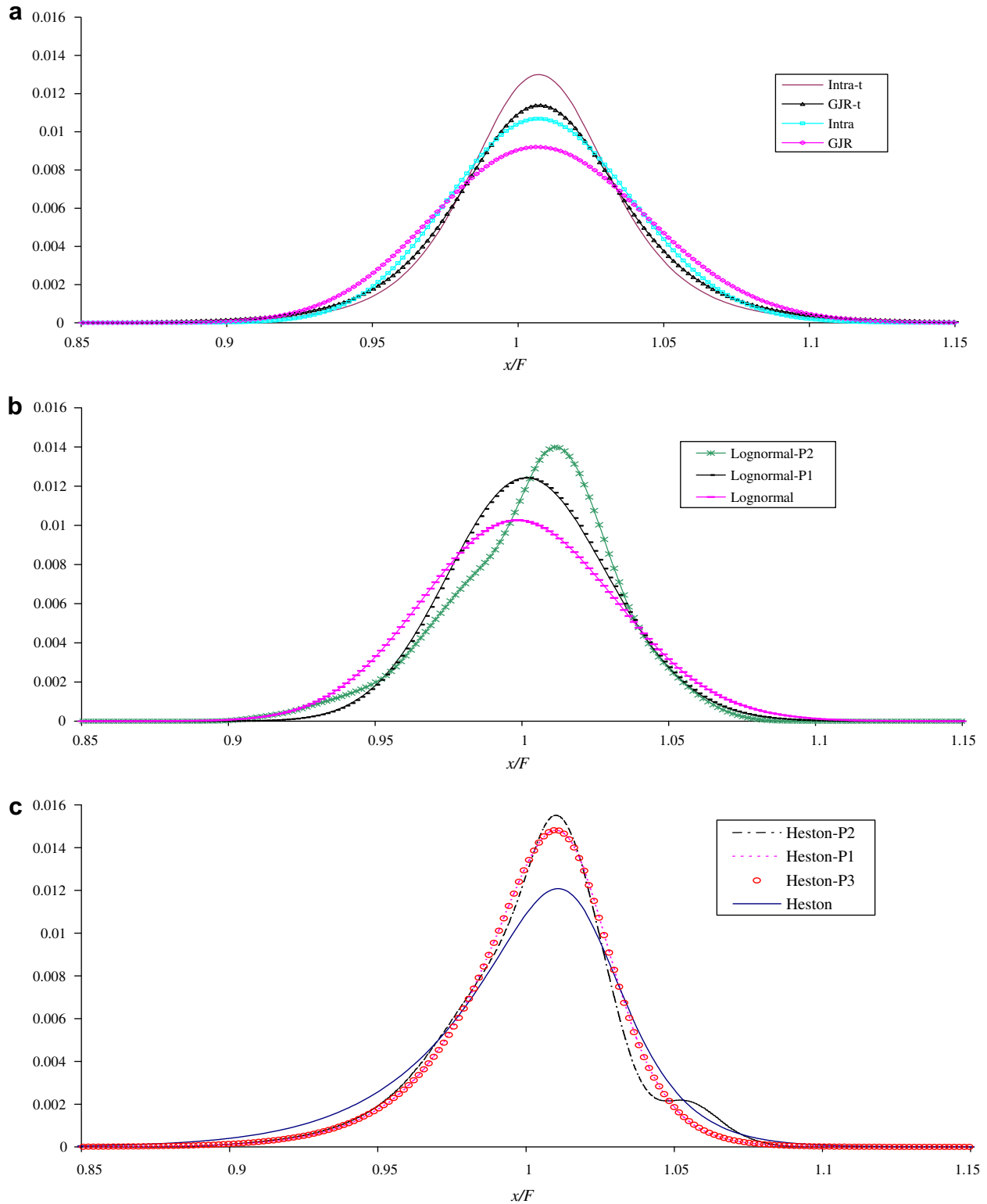


Fig. 2. (a) Four-week ahead density forecasts obtained from ARCH models on November 17th, 2004. (b) Four-week ahead density forecasts obtained from lognormal densities on November 17th, 2004. (c) Four-week ahead density forecasts obtained from Heston's model on November 17th, 2004.

0.64 (with jumps) and the 0.74 of Bates (2000), are much higher than those that are obtained from asset prices alone. A typical real-world estimate is 0.22, reported in Eraker (2004), while Ait-Sahalia and Kimmel (2007) obtain 0.48 from a bivariate time series model for price and volatility indices.

Our median estimates of the correlation ρ are -0.66 (without jumps) and -0.68 (with jumps). They are similar to the average, risk-neutral estimate of -0.64 in Bakshi et al. (1997), which is far more negative than their estimate of -0.28 obtained from time series of asset returns and changes in implied volatilities.

Table 2
Summary statistics for the daily estimates of the risk-neutral parameters. Estimates are summarized for the risk-neutral dynamics $dp/p = \sqrt{V}dW_1 + (e^{\lambda} - 1)dN - \lambda\bar{\mu}_j dt$ and $dV = \kappa(\theta - V)dt + \xi\sqrt{V}dW_2$, with $dW_1 dW_2 = \rho dt$. The Poisson process N_t has intensity λ , and is independent of the bivariate Wiener process $(W_{1,t}, W_{2,t})$. The jumps J_i in $\log(p_i)$ are normally distributed with mean μ_j and variance σ_j^2 .

Parameter	Median no jumps	Median with jumps	Mean no jumps	Mean with jumps	Standard deviation No jumps	Standard deviation With jumps
$\sqrt{V_0}$	0.1787	0.1651	0.1898	0.1877	0.0741	0.0694
κ	4.1528	3.0920	4.9292	3.8748	3.6598	3.2930
θ	0.0452	0.0347	0.0505	0.0421	0.0273	0.0381
ξ	0.7925	0.6400	0.9296	0.6977	0.5160	0.3731
ρ	-0.6624	-0.6795	-0.6590	-0.6788	0.0875	0.1181
λ		0.4746		0.6930		0.7147
μ_j		-0.0630		-0.1019		0.1609
σ_j		0.0385		0.0875		0.1059
MSE	0.0472	0.0327	0.1621	0.1072	0.3126	0.2052

The parameters are estimated each day from 1990 to 2004, from the out-of-the-money options on S&P 500 futures, by minimizing the mean squared error (MSE) of the fitted option prices. V_0 is the contemporaneous variance when the option prices are recorded. The constraint $\kappa \leq 36$ is applied, but only required when jumps are excluded by setting $\lambda = 0$.

We find that the jump-diffusion specifications do not yield more successful real-world densities than the simpler pure-diffusion specifications. Consequently, we initially focus on the pure-diffusion methods and defer our comparisons with the jump-diffusion methods until Section 5.9.

5.4. Cumulative probabilities from risk-neutral densities

A set of one-day-ahead RNDs provides cumulative distribution functions $F_{Q,t}(x)$ that can be evaluated at the next futures prices, p_{t+1} , to define observed probabilities $u_{t+1} = F_{Q,t}(p_{t+1})$. As expected, these probabilities are incompatible with a uniform distribution.

The sample c.d.f. $\tilde{C}(u)$ calculated from a time series $\{u_{t+1}\}$ is defined by (23). We show the differences between sample and uniform probabilities, $\tilde{C}(u) - u$, as the dark curve on Fig. 3 for the Heston RNDs. It is seen that there are too few outcomes for u near to either zero or one; only 5.7% of the observed u -values are below 0.1 and only 6.6% of them are above 0.9. The maximum value of $|\tilde{C}(u) - u|$ equals 6.6%; this value of the KS statistic rejects the null hypothesis of a uniform distribution at the 0.01% level. The differences for the lognormal RNDs are similar.

The shape of the deviation curve is primarily explained by the risk-neutral, standard deviations being higher, on average, than the historical standard deviations. Consequently, the risk-neutral probabilities of large price changes exceed the real-world probabilities.

5.5. Calibration transformations

A nonparametric estimate of the real-world density of the probabilities u_{t+1} is provided by the empirical calibration density stated in (37). This estimated density, $\hat{c}(u)$, is shown by the light curve on Fig. 4 for the one-day-ahead Heston case using the data from 1991 to 2004.

The time series averages of the *ex ante* estimates for the parametric calibration transformation applied to the one-day-ahead, risk-neutral Heston densities are $a = 1.434$ and $b = 1.412$. The corresponding calibration density, (10), is plotted as the dark curve on Fig. 4. The parametric and nonparametric calibration densities are similar, except near the end points of the distribution. The *ex ante* estimates of a and b vary between 1.3 and 1.6 and nearly always have $a > b$. The corresponding *ex ante* estimates for the risk-neutral lognormal densities are between 1.15 and 1.40 and their time series averages are 1.280 and 1.247.

The calibration methodology is intended to produce real-world densities whose observed probabilities u_{t+1} are uniformly distributed. After applying the parametric and nonparametric transformations, the one-day-ahead deviations $\tilde{C}(u) - u$ estimated *ex*

ante from all the data are shown as light curves on Fig. 3. These deviations are much nearer to zero than those for the risk-neutral densities, particularly for the nonparametric transformation.

Similar results and conclusions are obtained for the one-week-ahead densities. The time series averages of the Heston-estimates of a and b are 1.424 and 1.409, respectively. For horizons of two or more weeks, the average Heston-estimate of a is between 1.45 and 1.58, and it is always more than the average estimate of b which ranges from 1.30 to 1.43.¹⁸

5.6. Risk-premia transformations

The third transformation of the Heston RNDs into real-world densities changes the drift rates of the price and the volatility. The risk premia coefficients η_1 and η_2 in the bivariate diffusion (6) have been estimated for each of the seven horizons. These estimates should be similar across horizons if the assumed risk-neutral and real-world dynamics are correct. The seven estimates of η_1 (the return risk premium per unit variance) are indeed similar, including 2.41, 2.25 and 2.86 for the one-day, one-week and two-weeks horizons estimated from the entire sample from 1991 to 2004. All the full-period estimates of η_2 (the variance risk premium per unit of variance) provide further evidence for a negative volatility premium. The estimates, however, are approximately proportional to the reciprocal of the forecast horizon, varying from -197 for the one-day horizon to -4.2 for the 12-week horizon. This empirical effect is consistent with the real-world variance at time t being systematically lower than the estimated initial level V_t of the stochastic process for the risk-neutral variance.¹⁹

With $\eta_1 > 0$ and $\eta_2 < 0$, the risk-premia transformation ensures that the means and the standard deviations of the real-world densities are, respectively, above and below their risk-neutral counterparts. For the one-day horizon, Fig. 3 confirms that the risk-premia transformation reduces the magnitudes of the deviations $\tilde{C}(u) - u$. It can be seen that the deviations for the risk-premium and the parametric calibration transformations are similar.

A risk-premium transformation of the lognormal RNDs has also been investigated. Only the single risk parameter η_1 is then available, which improves the means but not the standard deviations of the densities. Consequently, the transformation only changes

¹⁸ The implied risk transformations cannot be reconciled with a representative agent model, because the estimates of b comprehensively fail to satisfy the constraint $b < 1$ for all horizons; see Liu et al. (2007). Ziegler (2007) provides a detailed theoretical analysis of several potential explanations of this empirical conclusion.

¹⁹ Consequently, all our transformations from risk-neutral to real-world densities can be reinterpreted as methods that jointly estimate risk premia and remove the systematic overpricing of option contracts.

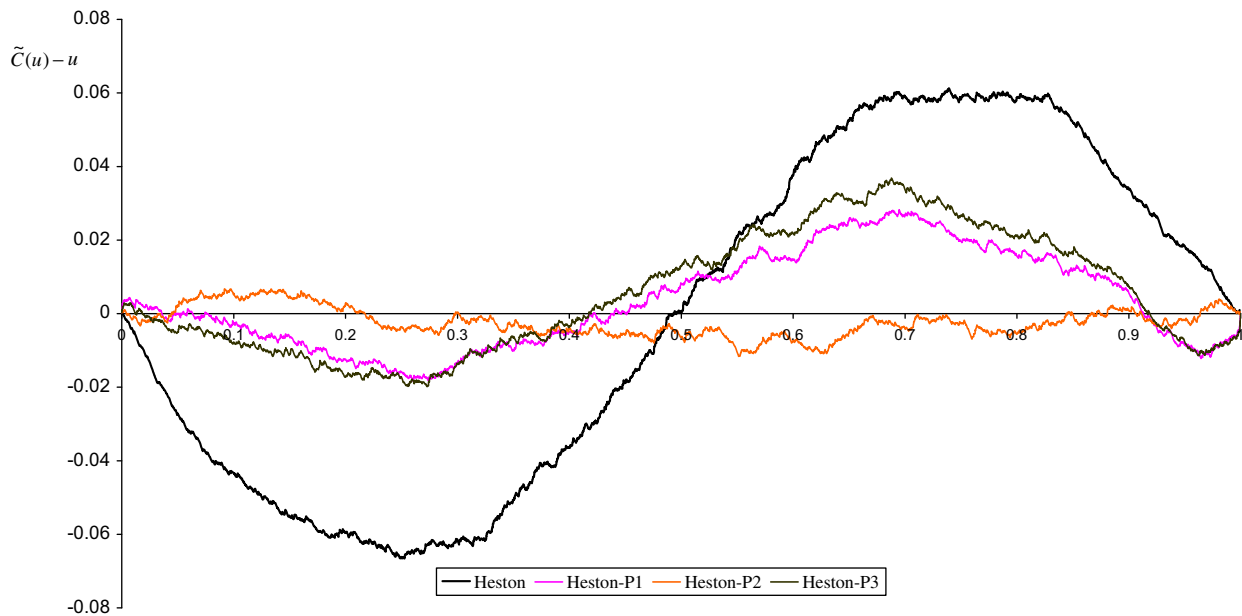


Fig. 3. The function $\tilde{C}(u) - u$ for one-day forecasts obtained from Heston's model and risk-transformations.

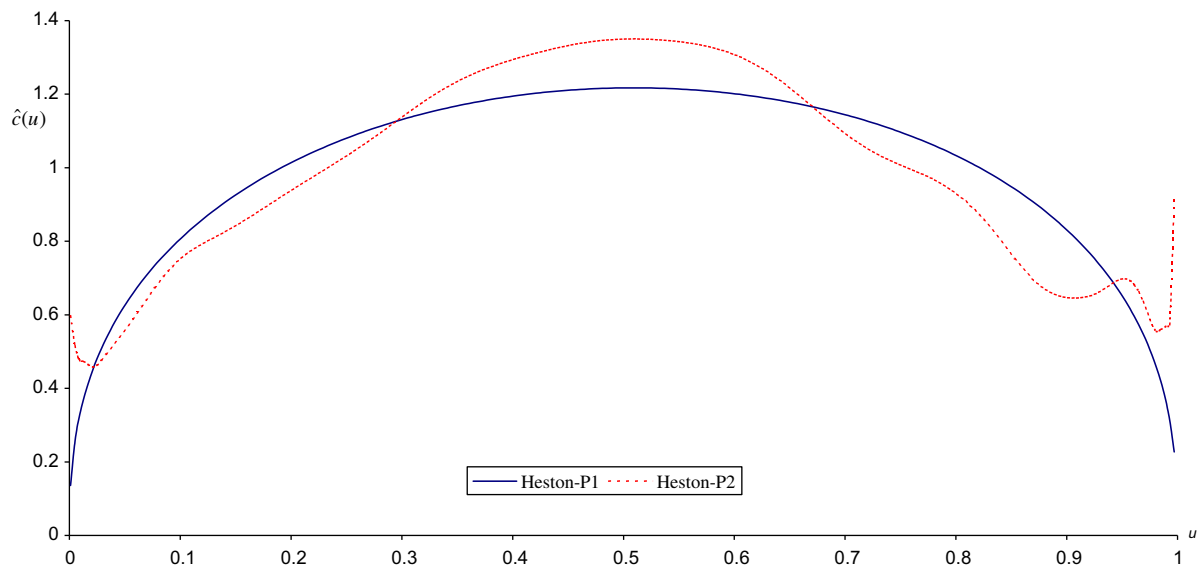


Fig. 4. The estimated calibration densities $\hat{c}(u)$ for cumulative probabilities u obtained from the one-day Heston forecasts.

the log-likelihoods by minor amounts and so the results are not reported.

5.7. Likelihood comparisons

Table 3 summarizes the log-likelihoods of the futures prices from January 1991 until December 2004, for 13 *ex ante* density forecasting methods. These log-likelihoods are given for non-overlapping forecasts, made for seven horizons that range from one-day to 12 weeks. We define the benchmark log-likelihoods as the values for the simplest historical method, namely the GJR densities. Table 3 shows the log-likelihood values in excess of the benchmark levels, for all other methods.

5.7.1. Historical methods

Initially consider the log-likelihoods of the four historical methods described in Section 2.3. These values are always higher for

conditional t -densities than for the matched conditional normal densities. They are also always higher for densities obtained from high-frequency returns than for the matched densities obtained from one-period returns. Consequently, the best of the four methods is the Intra- t method for all seven horizons. From (19), this best historical method has a posterior probability above 0.9999 for each of the two shortest horizons and between 0.88 and 0.95 for each of the five longer horizons; the total probability for the Intra and Intra- t methods exceeds 0.999 for all horizons of six or less weeks.

At the shortest horizon of one-day, incorporating non-normality adds more to the log-likelihood than incorporating intraday price information. The relative contributions of non-normality and intraday prices are similar when the horizon is either one or two weeks, while intraday prices contribute more when the horizon is either four or six weeks.

Applying a calibration transformation to the Intra- t densities improves some log-likelihood values. Table 3 shows that one

Table 3
Log-likelihoods for sets of density forecasts. The numbers tabulated are the log-likelihoods of the GJR density forecasts and the log-likelihoods of the other sets of forecasts in excess of the GJR benchmark values. The letter *Q* refers to risk-neutral densities. The risk transformation *P1* refers to the parametric calibration transformation, *P2* to the nonparametric calibration transformation, and *P3* to the risk-premia transformation, respectively, defined by (11), (36) and (6). The number in each row shown in bold font indicates the highest ranked method for the corresponding horizon.

Forecast horizon	Number of obs.	GJR	GJR- <i>t</i>	Intra	Intra- <i>t</i>	Risk-transformed Intra- <i>t</i>		Log normal <i>Q</i>	Risk-transformed lognormal		Heston <i>Q</i>	Risk-transformed Heston		
						<i>P1</i>	<i>P2</i>		<i>P1</i>	<i>P2</i>		<i>P1</i>	<i>P2</i>	<i>P3</i>
Data source		Daily returns	Daily returns	Intraday returns	Intraday returns	Intraday returns	Intraday returns	Options	Options	Options	Options	Options	Options	Options
1 Day	3520	-11951.2	91.4	56.4	135.9	134.0	140.8	27.0	73.5	100.9	-2.4	103.9	127.4	93.7
1 Week	711	-2961.9	13.5	16.1	34.8	36.7	38.4	17.0	32.8	36.5	18.5	41.5	35.2	40.6
2 Weeks	351	-1574.0	10.4	16.2	18.5	16.6	18.0	13.6	27.8	26.4	14.8	26.9	22.4	25.8
4 Weeks	176	-853.6	4.1	10.9	13.2	11.9	9.6	12.5	13.4	15.8	16.0	20.3	20.2	23.2
6 Weeks	115	-596.9	5.7	14.0	16.4	13.1	17.6	16.0	17.1	16.9	19.6	16.3	18.7	20.3
8 Weeks	86	-446.1	1.5	2.6	5.8	4.3	2.8	4.9	9.4	9.3	6.7	7.8	7.5	7.5
12 Weeks	58	-310.2	5.2	3.7	7.4	5.1	5.7	5.6	6.9	6.8	9.8	8.5	9.1	7.6

transformation provides an improvement for the one-day horizon and that both transformations improve the one-week ahead densities. However, the transformations provide inferior log-likelihoods for longer horizons.

5.7.2. Ten univariate methods

Comparisons are now made between 10 methods, which define three sets of historical densities and seven sets of option-based densities. The historical densities are referred to as Intra-*t*, Intra-*t-P1* and Intra-*t-P2*. We refer to the RNDs as the lognormal-*Q* and the Heston-*Q* densities. The real-world densities are labeled lognormal-*P1*, lognormal-*P2*, Heston-*P1*, Heston-*P2* and Heston-*P3*; *P1*, *P2* and *P3* are, respectively, the parametric calibration, the non-parametric calibration, and the risk-premia transformations, defined by (11), (36) and (6).

5.7.3. One-day horizon

The ranked log-likelihoods, in excess of the benchmark value, are: Intra-*t-P2* 141, Intra-*t* 136, Intra-*t-P1* 134, Heston-*P2* 127, Heston-*P1* 104, lognormal-*P2* 101, Heston-*P3* 94, lognormal-*P1* 73, lognormal-*Q* 27 and Heston-*Q* -2. These numbers are summarized in five remarks.

First, the Intra-*t* densities obtained from high-frequency returns have high log-likelihoods compared with the option-based densities. Second, the nonparametric risk transformation *P2* is superior to the parametric transformation *P1*, with the differences, respectively, equal to 6.8, 23.5 and 27.4 for the Intra, Heston and lognormal cases. Third, the Heston *P*-densities have higher log-likelihoods than the lognormal *P*-densities, the differences being 30.4 for *P1* and 27.5 for *P2*. Fourth, the risk-premia transformation *P3* ranks behind the statistical transformations *P1* and *P2* for the Heston densities. Finally, as expected the *Q*-densities are far inferior to their related *P*-densities.

The posterior probability of the Intra-*t-P2* method equals 0.992, although four of the AG test statistics are insignificant at the 5% level when this method is compared with the nine alternatives: AG equals 0.74, 0.99, 1.94 and 0.94, respectively, for tests against the Intra-*t*, Intra-*t-P1*, Heston-*P1* and Heston-*P2* methods.²⁰

5.7.4. Horizons from 1 to 4 weeks

The best method depends on the horizon, being Intra-*P2* for one day, Heston-*P1* for one week, lognormal-*P1* for two weeks and Hes-

ton-*P3* for four weeks. The absence of a uniformly best method reflects the similarity of the log-likelihoods for the five option-based, *P*-densities, for all but the one-day horizon.

Each set of option *P*-densities always outperforms the corresponding set of *Q*-densities. We note that the option *P*-densities are superior to the Intra-*t* densities for 8 of the 15 possible comparisons when the horizon is one week. In contrast, all the option *P*-densities are superior to all the Intra-*t* densities for the two-week and the four-week horizons. For the one-week horizon, the average excess log-likelihood for the three sets of Intra-*t* densities is 36.6 and the average for the five sets of option *P*-densities is 37.3. The corresponding comparisons are 17.7 versus 25.9 for two weeks and 11.6 versus 18.6 for four weeks.

The total posterior probability for the Heston methods equals 0.957, 0.305 and 0.999, respectively, for the one, two and four-week horizons. The corresponding totals for the lognormal methods are 0.005, 0.695 and 0.001, and for the Intra-*t* methods they equal 0.038, 0.000 and 0.000. The AG test concludes that the best method for the one-week horizon is only significantly better than two of the other nine methods (namely the *Q*-methods) at the 5% level; the best method is better than four methods, at the 5% level, for the two and four-week horizons.²¹

5.7.5. Horizons from 6 to 12 weeks

The differences between the log-likelihoods of the various methods decrease as the horizon increases, primarily because the numbers of non-overlapping forecasts decrease. The best methods for the longer horizons are Heston-*P3* for six weeks, lognormal-*P1* for eight weeks and Heston-*Q* for 12 weeks, while the worst is either Intra-*t-P1* or Intra-*t-P2*. The differences between the best and the worst methods are 7.2, 6.6 and 4.1 for these horizons.

The total posterior probability for the Heston methods now equals 0.907, 0.226 and 0.888, respectively, for the six, eight and 12-week horizons. The totals for the lognormal methods are 0.046, 0.761 and 0.057, and for the Intra-*t* methods they are 0.047, 0.014 and 0.055.

5.7.6. Mixtures

We now consider the log-likelihoods for mixtures defined by a fraction α of an option-based density added to a fraction $1 - \alpha$ of the Intra-*t* density. Table 4 shows the mixture log-likelihoods and the time series averages of the *ex ante* estimates of α .²²

²⁰ The AG test produces the same conclusions and similar test values when a Newey-West adjustment is made to the estimated variance of \bar{d} . The four insignificant values change to 0.73, 0.98, 1.95 and 0.93 when ten autocorrelations are used. The first-lag autocorrelations range from -0.054 to 0.010 for the 45 time series of differences calculated from the log-likelihoods of the 10 univariate methods.

²¹ The Newey-West adjustment to the test changes only 1 of these 27 test results.

²² The log-likelihoods attained by setting $\alpha = 0.5$ are often near those obtained by *ex ante* selection of α . Averages are notably inferior, however, when *Q*-densities contribute to the mixture and the horizon is short.

Table 4

Log-likelihoods for mixtures of historical densities and option-based densities. Each log-likelihood is the value in excess of the GJR benchmark given in Table 3. The mixture densities are a fraction α of the option-based density plus a fraction $1 - \alpha$ of the Intra- t density. α is estimated *ex ante*. The risk transformation $P1$ refers to the parametric calibration transformation, $P2$ to the nonparametric calibration transformation, and $P3$ to the risk-premia transformation, respectively, defined by (11), (36) and (6) and applied to risk-neutral densities denoted by Q .

Forecast horizon	Intra- t only	Intra- t combined with													
		Lognormal						Heston							
		Q	Average α (%)	P1	Average α (%)	P2	Average α (%)	Q	Average α (%)	P1	Average α (%)	P2	Average α (%)	P3	Average α (%)
1 Day	135.9	147.7	17	150.8	35	142.0	31	151.2	17	159.8	38	149.9	48	159.5	34
1 Week	34.8	33.5	21	36.6	60	38.3	60	34.0	32	42.5	62	38.2	51	42.2	64
2 Weeks	18.5	18.9	23	27.1	83	24.9	71	20.4	33	28.0	68	24.8	53	27.8	71
4 Weeks	13.2	14.1	11	13.6	49	15.1	49	15.7	70	20.1	83	19.8	83	23.0	89
6 Weeks	16.4	19.1	48	16.7	76	17.7	56	19.8	53	18.6	47	18.7	51	20.2	77
8 Weeks	5.8	6.1	33	8.8	82	8.0	60	6.7	49	8.1	58	6.7	44	8.2	58
12 Weeks	7.4	7.4	9	9.5	72	7.2	40	6.6	31	9.5	72	8.8	65	8.9	68

Table 5

Results from the Kolmogorov–Smirnov test. The tabulated numbers are the p -values for the Kolmogorov–Smirnov test of the null hypothesis that the variables u_t have a uniform distribution. The risk transformation $P1$ refers to the parametric calibration transformation, $P2$ to the nonparametric calibration transformation, and $P3$ to the risk-premia transformation, respectively, defined by (11), (36) and (6) and applied to risk-neutral densities denoted by Q .

Forecast horizon	Number of obs.	GJR (%)	GJR- t (%)	Intra (%)	Intra- t (%)	Risk-transformed Intra- t (%)		Log normal (%)	Risk-transformed lognormal (%)		Heston (%)	Risk-transformed Heston (%)				
						P1	P2		Q	P1		P2	Q	P1	P2	P3
1 Week	711	0.36	25.09	0.11	12.73	2.15	39.95	0.04	1.14	74.35	0.60	96.08	78.80	67.58		
2 Weeks	351	13.36	76.69	0.73	89.15	13.11	63.64	0.03	21.60	67.63	3.52	92.71	66.42	78.16		
4 Weeks	176	13.09	82.55	0.14	64.28	16.24	63.78	0.04	0.67	24.86	7.75	82.81	49.13	69.38		
6 Weeks	115	94.00	83.30	4.97	42.91	3.81	29.25	1.57	14.28	18.18	54.23	84.58	80.00	41.38		
8 Weeks	86	59.42	72.82	2.42	77.81	38.37	88.08	1.40	48.79	98.95	9.90	84.75	99.45	51.56		
12 Weeks	58	85.19	89.06	0.31	69.28	35.18	63.81	0.18	21.62	61.28	31.29	90.03	91.03	85.18		

Table 6

Berkowitz test values and parameter estimates. The null hypothesis that the variables $y_t = \Phi^{-1}(u_t)$ are i.i.d with a standard normal distribution is tested against the alternative of an AR(1), Gaussian process. The tabulated numbers are the test statistic LR3 and the estimates of the AR and variance parameters.

Forecast horizon		GJR	GJR- t	Intra	Intra- t	Risk-transformed Intra- t		Log normal	Risk-transformed lognormal		Heston	Risk-transformed Heston				
						P1	P2		Q	P1		P2	Q	P1	P2	P3
	Variance	0.93	1.07	0.90	1.02	1.05	0.99	0.79	1.04	1.00	0.68	1.04	1.02	1.02		
	LR3	9.78*	8.44*	19.84*	1.89	4.52	0.84	96.82*	2.58	0.09	241.46*	3.72	1.02	0.64		
1 Week	AR	-0.11	-0.12	-0.10	-0.11	-0.11	-0.11	-0.10	-0.10	-0.09	-0.09	-0.09	-0.10	-0.08		
	Variance	0.78	0.88	0.71	0.83	1.04	0.97	0.71	0.97	0.93	0.65	0.98	0.94	0.94		
	LR3	26.57*	13.93*	43.35*	19.06*	10.00*	9.15*	46.13*	8.51*	8.27*	61.63*	6.14	8.50*	5.52		
2 weeks	AR	-0.03	-0.05	-0.03	-0.05	-0.04	-0.04	-0.02	-0.02	-0.01	-0.01	-0.01	0.00	0.00		
	Variance	0.86	1.01	0.78	0.90	1.12	1.02	0.63	0.99	0.91	0.63	1.02	0.94	1.02		
	LR3	3.90	0.89	13.07*	2.85	4.35	1.72	35.95*	0.94	2.11	32.32*	0.34	1.39	0.33		
4 Weeks	AR	0.00	-0.01	0.05	0.02	0.02	0.02	0.04	0.04	0.04	0.06	0.05	0.05	0.08		
	Variance	0.77	0.92	0.81	0.90	1.10	0.97	0.63	0.99	0.90	0.74	1.00	0.92	1.00		
	LR3	6.04	0.98	6.54	1.74	1.16	0.36	19.61*	0.93	1.43	7.39	0.71	1.41	1.29		
6 Weeks	AR	-0.22	-0.17	-0.01	-0.15	-0.12	-0.12	-0.14	-0.11	-0.10	-0.10	-0.09	-0.07	-0.03		
	Variance	0.93	1.09	1.01	1.13	1.34	1.08	0.71	1.17	1.05	0.77	1.11	1.07	1.22		
	LR3	6.03	4.19	3.32	3.83	7.93	2.42	11.63*	3.81	1.89	6.48	1.89	1.74	2.50		
8 Weeks	AR	0.02	0.06	0.09	0.07	0.08	0.09	0.14	0.16	0.18	0.11	0.18	0.18	0.22		
	Variance	0.85	1.24	0.76	0.90	1.13	0.96	0.54	0.94	0.79	0.74	0.93	0.80	0.91		
	LR3	1.65	2.90	6.16	1.01	1.50	0.90	18.36*	2.39	4.35	5.63	2.71	4.40	4.36		
12 Weeks	AR	-0.08	-0.04	0.04	0.02	0.03	0.04	0.06	0.08	0.11	0.09	0.10	0.15	0.19		
	Variance	1.40	1.07	0.90	0.93	1.29	1.03	0.59	1.12	0.93	0.75	1.04	0.98	1.16		
	LR3	4.41	0.46	4.45	0.40	2.54	0.40	10.95*	1.06	0.82	4.92	0.90	1.62	3.18		

* Indicate that the null is rejected at the 5% level, when LR3 > 7.81.

Each mixture of an option-based P -density and the corresponding Intra- t density has a higher log-likelihood than both the P -den-

sity and the Intra- t density components for the one-day and one-week horizons. However, for horizons of two or more weeks each

set of P -densities has a log-likelihood near to that of its mixture while the mixture log-likelihoods are almost always higher than the respective values for the Intra- t densities.

For the one-day horizon, the excess log-likelihoods of the mixtures range from 142 to 160 and all exceed the best univariate result, which is 141 for the Intra- t - $P2$ densities. Combining Intra- t with either Heston- $P1$ or Heston- $P3$ gives the best results. The total Bayesian probability of these mixtures equals 0.9998, when the methods considered are Intra- t , 7 option-based methods and the seven mixture methods. The AG test rejects the null hypothesis of no difference between one component of a mixture and the mixture of Intra- t and an option-based method for 13 of the 14 possible comparisons (at the 5% significance level).

As the forecast horizon increases, first the average weight given to the five sets of P -densities increases, being 37%, 59%, 69% and 71% for the one-day, one-week, two-week and four-week horizons. Second, the total Bayesian probability for the mixtures decreases, from 100% to 77%, 61% and 45%. Third, the AG test values decrease; beyond the one-day horizon, none of these values reject the null hypothesis (at the 5% level) that a real-world, option-based method has the same expected log-likelihood as its mixture with the Intra- t method.

5.8. Diagnostic tests

The KS statistic tests the null hypothesis that a set of densities are correctly specified, but it assumes the cumulative probabilities u_t are independent random variables. Table 5 lists the p -values for the KS test for the 13 *ex ante* density forecasting methods, for each of the seven horizons. As the null is rejected at the α -level whenever $p < \alpha$, it is found that 27 of the 91 test values reject the null hypothesis at the 5% level. Nineteen of the 27 rejections occur for densities which might be expected to be mis-specified, namely the RNDs and the ARCH densities that are conditionally normal.

It is noteworthy that the Intra- t densities and the $P2$ -densities obtained by applying the nonparametric transformation to the Intra- t densities and the RNDs have the most satisfactory p -values: 19 of the 28 p -values exceed 50% and their minimum is 18%. The Heston- $P1$ and Heston- $P3$ densities also have satisfactory p -values beyond the one-day horizon.

A good specification of the tails may be more important than a good overall fit for some purposes. Tail comparisons are made by calculating the maximum of $|\tilde{C}(u) - u|$ for the regions $0 \leq u \leq \beta$ and $1 - \beta \leq u \leq 1$. With β equal to 0.025, 0.05 or 0.1, the Intra- t - $P2$, lognormal- $P2$ and Heston- $P2$ methods generally outrank the others. For all six tail regions, each of these $P2$ -methods is always in the top five for the one-day and one-week horizons.

The LR3 statistic of Berkowitz (2001) tests the null hypothesis that the observations $y_t = \Phi^{-1}(u_t)$ are i.i.d. from a standard normal distribution against the alternative that they are from a Gaussian, AR(1) process with no restrictions on the mean, variance and autoregressive parameters. Table 6 contains the values of LR3 and the MLEs of the variance and autoregressive parameters for the 13 density forecasting methods, by horizon.

The MLEs of the AR parameter are all between -0.012 and 0.005 for the sets of one-day forecasts, so they provide no evidence to doubt that the time series $\{u_t\}$ contain independent observations. The MLEs for the one-week forecasts are, however, all between -0.12 and -0.08 and they all reject the null hypothesis that the AR parameter is zero at the 5% level. There is no significant evidence of time-series dependence for longer horizons.

The MLEs of the variance parameter are usually near one, as required for correctly specified densities. The low estimates for the RNDs, such as the 0.68 and the 0.79 for the one-day ahead Heston- Q and lognormal- Q forecasts, are a direct consequence of his-

Table 7

The impact of a jump component upon the log-likelihood values. The tabulated numbers are the differences between *ex ante* log-likelihoods for sets of jump-diffusion densities and the corresponding sets of pure-diffusion densities. Positive numbers indicate that the jump-diffusion specification has a higher log-likelihood. The letter Q refers to risk-neutral densities. The risk transformation $P1$ refers to the parametric calibration transformation, $P2$ to the nonparametric calibration transformation, and $P3$ to the risk-premia transformation.

Forecast horizon	Q	$P1$	$P2$	$P3$
1 Day	60.9	-4.8	-8.5	-7.6
1 Week	12.6	2.8	2.6	-1.4
2 Weeks	4.4	0.0	-0.2	-2.9
4 Weeks	1.1	0.2	-0.8	-3.5
6 Weeks	-0.5	-0.5	-3.7	-2.1
8 Weeks	0.2	0.2	0.0	-0.1
12 Weeks	1.2	-1.0	-1.2	-1.3

torical volatility being lower (on average) than risk-neutral volatility.

The null distribution of LR3 is χ^2_3 and thus a test value is significant at the 5% level if it exceeds 7.81. Table 6 shows that the null is always rejected for the lognormal- Q densities at the 5% level. For the other methods, the null is rejected at the 5% level for various sets of one-day, one-week and two-week forecasts; there are no rejections for the longer horizons, which may reflect low power when few forecasts are evaluated. The only methods whose densities always pass the LR3-test at the 5% level are the Heston- $P1$ and Heston- $P3$ methods. With only one exception, all the significant values of LR3 can be explained by assuming risk neutrality, or by assuming the conditional shape of a historical density is normal or by the negative estimates of the AR parameter for the one-week horizon.

5.9. Comparisons between diffusion and jump-diffusion methods

The robustness of the empirical evidence presented so far for the option-based densities can be assessed by comparing the Heston (H) densities derived from (1) and (2) with the jump-diffusion (JD) densities derived from (2) and (3). Four sets of JD-densities are compared with the corresponding four Heston sets, for each horizon. The same labels Q , $P1$, $P2$ and $P3$ are used for the JD sets. An extra risk premium parameter is estimated for the JD- $P3$ densities; for the one-day horizon, the average estimated size of the jumps in $\log(p_t)$ increases from -10% for the JD- Q densities to -5% for the JD- $P3$ densities.

Table 7 shows the differences D between the *ex ante* log-likelihoods of sets of JD-densities and their corresponding sets of H-densities. The risk-neutral densities are significantly improved by including the jump component, for the three shortest horizons. The values of D for the one-day, one-week and two-weeks horizons equal 60.9, 12.6 and 4.4, and the respective AG test statistics are 5.64, 4.01 and 1.99.

In contrast, there is no evidence that estimating a jump component improves the option-based, real-world densities. All three differences D are negative for the one-day comparisons.²³ There are only four positive real-world values of D in Table 7, 15 values are negative and 2 are reported as 0.0. None of the real-world, AG test statistics is significant at the 5% level and the diagnostic test statistics are generally similar for the Heston and JD-densities.

²³ Risk-neutral jumps improve the specification of risk-neutral densities and subsequent transformations to real-world densities will both incorporate risk preferences and help to remove any remaining mis-specification of the RNDs. In contrast, each of the three transformations applied directly to Heston RNDs achieves a similar result in a single step. These transformations appear to use past prices to efficiently learn about systematic differences between risk-neutral and real-world densities.

Further differences D have been calculated for mixtures of the Intra- t and option-based densities. The one-day difference is only 3.4 ($AG = 1.50$) for the mixtures of Intra- t and risk-neutral densities, compared with 61 in Table 7, because more than 80% of these mixture weights are allocated to the Intra- t densities. The real-world, one-day differences are all between -2 and -1 . The differences for the one-week and longer horizons are similar to the numbers in Table 7 and they are all insignificant at the 5% level.

6. Conclusions

Hitherto, most option-based density estimation methods have only provided results within a risk-neutral context and most methods have required the forecast horizon to coincide with an option expiry date. In contrast, we are the first to show it is possible to construct informative, *ex ante*, real-world densities for many forecast horizons by using currently available price information.

Our most important conclusions for the S&P 500 index depend upon the forecast horizon. For the one-day horizon, ARCH densities obtained from 5-min returns are more informative than densities obtained from option prices but the most informative densities are provided by mixtures of historical and option-based methods. We say the mixture densities are the most informative because they rank highest according to the out-of-sample likelihood criterion. At the one-week horizon, the mixture densities continue to outrank historical and option-based methods but now historical methods have likelihoods similar to those of option-based methods.

For the two-week and four-week horizons we find strong evidence that three transformations of risk-neutral densities, estimated from index levels, option prices and Heston's pricing formula, all provide real-world densities that are more informative than the historical densities estimated from ARCH models. At even longer horizons, up to 12 weeks, the empirical evidence continues to favor option-based methods.²⁴ Furthermore, mixture densities are not preferable to option-based densities when the horizon is at least two weeks. Neither are real-world densities based upon jump-diffusion processes superior to those that are estimated from pure-diffusion dynamics.

Jiang and Tian (2005) have shown that the information content of option prices is higher than that of daily and intraday index values when forecasting the volatility of the S&P 500 index over horizons from one to six months. Our study shows that the same conclusion applies to *ex ante* density forecasts of the S&P 500 index when the forecast horizon is two or more weeks, but it does not hold for shorter horizons. We attribute the superior performance of option-based forecasts over medium-term horizons to the forward-looking property of option prices. As we only use prices for option contracts that have eight or more days until expiry, the one-day and one-week-ahead risk-neutral densities are always extrapolations which are not supported by competitive trading. This partially explains why the best historical densities are relatively more successful than the real-world densities for the two shortest horizons. Furthermore, we conjecture that the historical density is superior for the shortest horizon of one-day because an accurate forecast of tomorrow's volatility can be obtained from high-frequency returns summarized by daily measures of realized volatility.

We have described three transformations of the risk-neutral densities that produce real-world densities. For our data, these real-world densities always outrank the risk-neutral densities for horizons between one-day and four-weeks inclusive. The nonpara-

metric calibration transformation produces the best diagnostic test results and it also enhances the historical densities for the two shortest horizons.

Risk managers, central bankers and other users of density forecasts for equity indices should not rely on risk-neutral densities extracted from option prices. They can instead obtain more accurate densities by applying a risk transformation to risk-neutral densities. Our empirical evidence shows that it is reasonable to seek further improvements for short horizons by mixing the transformed option-based densities with historical densities which utilize the information provided by high-frequency returns.

Acknowledgements

For their helpful comments on earlier versions of this paper, we thank anonymous referees, Yacine Aït-Sahalia, Torben Andersen, Tim Bollerslev, Raffaella Giacomini, Martin Martens, Chyawat Orn-thanalai, Marc Paolella, Andrew Patton, Esther Ruiz, Mark Salmon, Christoph Schleicher, Allan Timmermann and Kenneth Wallis. We also thank seminar participants at the Bank of England, Athens University of Economics and Business, University of Basel, University of Cyprus, Imperial College, Lancaster University, London School of Economics, Carlos III University of Madrid, University of Manchester, National Central University, National Chengchi University, National Taiwan University, University of Nottingham, University of Reading, University of Warwick, University of Zurich, and participants at meetings held by the Bachelier Finance Society, the European Finance Association, the European Financial Management Association, the International Institute of Forecasters, Peking University and the Journal of Banking and Finance, the Centre for Analytical Finance, and Frankfurt MathFinance.

Appendix A. Estimation methods

We, respectively, denote the five diffusion and the eight jump-diffusion risk-neutral parameters estimates obtained from option prices at the end-of-day t by Θ_t and Θ_t^j . Here we show how risk parameters, other relevant parameters and the nonparametric calibration density are estimated within a diffusion framework. The same methods are also used to estimate parameters in a jump-diffusion framework by replacing Θ_t by Θ_t^j and, when appropriate, estimating the additional jump risk parameter.

A.1. Parametric estimation methods

The risk-premium parameters in (6), namely η_1 and η_2 , are estimated separately for each forecast horizon. Each horizon defines a set of non-overlapping time periods. For a specific set, at the end-of-period s corresponding to day n_s , we can evaluate the real-world density $f_{p,s,T}(x|\eta_1, \eta_2, \Theta_{n_s})$ for the asset price T years later, at the end-of-period $s+1$. The *ex ante* maximum likelihood estimates (MLEs) of η_1 and η_2 at time n_t are given by maximizing the log-likelihood of the observed asset prices p_s at the ends of periods $s = 1, 2, \dots, t$. Thus we maximize:

$$\log L(p_1, \dots, p_t | \eta_1, \eta_2) = \sum_{s=0}^{t-1} \log(f_{p,s,T}(p_{s+1} | \eta_1, \eta_2, \Theta_{n_s})). \quad (27)$$

In the same way, the parameters of the parametric calibration function (10), namely a and b , are also estimated separately for each forecast horizon. The risk-neutral density $f_{Q,s,T}(x|\Theta_{n_s})$ and its c.d.f are used to evaluate the real-world density $f_{p,s,T}(x|a, b, \Theta_{n_s})$ given by (11). The *ex ante* MLEs of a and b at time n_t are given by maximizing:

²⁴ The ARCH specifications may be mis-specified and it is thus possible that other ARCH specifications might provide more informative densities.

$$\log L(p_1, \dots, p_t | a, b) = \sum_{s=0}^{t-1} \log(f_{p,s,T}(p_{s+1} | a, b, \Theta_{n_s})). \tag{28}$$

The ARCH densities for one-period returns, specified by (13) and (14), have the general form $f_{ret}(r_{s+1} | I_s, \vartheta)$, that depends on a parameter vector ϑ and a set I_s of historical returns. The MLE at time t is the vector $\hat{\vartheta}_t$ that maximizes the log-likelihood of all the returns since some earlier time τ (assumed to precede the first available option prices):

$$\log L(r_\tau, \dots, r_1, \dots, r_t | \vartheta) = \sum_{s=\tau-1}^{t-1} \log(f_{ret}(r_{s+1} | I_s, \vartheta)). \tag{29}$$

From (12), the *ex ante* density of the next end-of-period price, p_{t+1} , is then given by

$$f_{ARCH,t}(x | I_t, \hat{\vartheta}_t) = f_{ret}(r | I_t, \hat{\vartheta}_t) / x \quad \text{with } r = \log(x/p_t). \tag{30}$$

Similarly, the MLE of the mixture parameter α , that determines the weights given to option-based and historical densities in (15), can be obtained by maximizing an appropriate log-likelihood function. We use a two-step method. The first step provides estimates of all the parameters except α . Then, at time n_t we will know the observed values of the components of the mixture, for example we know $\tilde{f}_{P,s} = f_{P,s,T}(p_{s+1} | \hat{a}_s, \hat{b}_s, \Theta_{n_s})$ and $\tilde{f}_{A,s} = f_{ARCH,s}(p_{s+1} | I_s, \hat{\vartheta}_s)$ for times $0 \leq s < t$. The MLE of α at time n_t is given by the number $\hat{\alpha}_t$ that maximizes

$$\log L(p_1, p_2, \dots, p_t | \alpha) = \sum_{s=0}^{t-1} \log(\hat{\alpha}_t \tilde{f}_{P,s} + (1 - \hat{\alpha}_t) \tilde{f}_{A,s}). \tag{31}$$

The *ex ante* mixture density for p_{t+1} is then

$$\hat{\alpha}_t f_{P,t,T}(x | \hat{a}_t, \hat{b}_t, \Theta_{n_t}) + (1 - \hat{\alpha}_t) f_{ARCH,t}(x | I_t, \hat{\vartheta}_t). \tag{32}$$

A.2. Nonparametric estimation methods

The nonparametric calibration function is re-estimated at the end of each period t . The observed futures prices define a set of t cumulative, risk-neutral probabilities, $u_{s+1} = F_{Q,s,T}(p_{s+1} | \Theta_{n_s}), 0 \leq s \leq t - 1$. We assume these observations are i.i.d. with c.d.f. given by the calibration function $C_T(u)$.

Let $\varphi(\cdot)$ and $\Phi(\cdot)$, respectively, denote the density and the c.d.f. of the standard normal distribution. We transform the observations u_i to new variables $y_i = \Phi^{-1}(u_i)$ and then fit a nonparametric, kernel c.d.f. to the set $\{y_1, y_2, \dots, y_t\}$. We use a normal kernel, with bandwidth B , and so obtain the kernel density and c.d.f., respectively, as

$$\hat{h}_T(y) = \frac{1}{tB} \sum_{i=1}^t \phi\left(\frac{y - y_i}{B}\right) \quad \text{and} \quad \hat{H}_T(y) = \frac{1}{t} \sum_{i=1}^t \Phi\left(\frac{y - y_i}{B}\right). \tag{33}$$

The bandwidth B in (16) should decrease as t increases. We have used the standard formula of Silverman (1986), $B = 0.9\sigma_y t^{0.2}$, with σ_y the standard deviation of the terms y_i .

The empirical calibration function is then

$$\hat{C}_T(u) = \hat{H}_T(\Phi^{-1}(u)). \tag{34}$$

From (8), the real-world c.d.f. for the next observed futures price becomes

$$F_{P,T}(x) = \hat{C}_T(F_{Q,T}(x)). \tag{35}$$

Also, with $u = F_{Q,T}(x)$ and $y = \Phi^{-1}(u)$, the real-world density is

$$f_{P,T}(x) = \frac{d}{dx} \hat{H}_T(y) = \frac{dy}{dx} \frac{d\hat{H}_T(y)}{dy} = \frac{du}{dx} \frac{dy}{du} \hat{h}_T(y) = \frac{f_{Q,T}(x) \hat{h}_T(y)}{\phi(y)}. \tag{36}$$

Finally, the nonparametric calibration density is

$$\hat{c}_T(u) = \hat{h}_T(y) / \phi(y). \tag{37}$$

References

Ait-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics* 83, 413–452.

Ait-Sahalia, Y., Lo, A.W., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance* 53, 499–547.

Amisano, G.G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business Economics and Statistics* 25, 177–190.

Anagnou-Basioudis, I., Bedendo, M., Hodges, S.D., Tompkins, R., 2005. Forecasting accuracy of implied and GARCH-based probability density functions. *Review of Futures Markets* 11, 41–66.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.

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

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.

Bakshi, G., Cao, C.G., Chen, Z., 1997. Empirical performance of alternative option pricing models. *Journal of Finance* 52, 2003–2049.

Bakshi, G., Kapadia, N., 2003. Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies* 16, 527–566.

Bakshi, G., Kapadia, N., Madan, D.B., 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies* 16, 101–143.

Bandi, F.M., Russell, J.R., 2006. Separating microstructure noise from volatility. *Journal of Financial Economics* 79, 655–692.

Bao, Y., Lee, T.-H., Saltoglu, B., 2007. Comparing density forecast models. *Journal of Forecasting* 26, 203–225.

Barone-Adesi, G., Whaley, R.E., 1987. Efficient analytic approximation of American option values. *Journal of Finance* 42, 301–320.

Bates, D.S., 1996. Jumps and stochastic volatility: exchange rate processes implicit in Deutsche mark options. *Review of Financial Studies* 9, 69–107.

Bates, D.S., 2000. Post-'87 crash fears in S&P 500 futures options market. *Journal of Econometrics* 94, 181–238.

Bates, D.S., 2006. Maximum likelihood estimation of latent affine processes. *Review of Financial Studies* 19, 909–965.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics* 19, 465–474.

Black, F., 1976. The pricing of commodity contracts. *Journal of Financial Economics* 3, 167–179.

Blair, B.J., Poon, S.-H., Taylor, S.J., 2001. Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high frequency index returns. *Journal of Econometrics* 105, 5–26.

Bliss, R., Panigirtzoglou, N., 2002. Testing the stability of implied probability density functions. *Journal of Banking and Finance* 26, 381–422.

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

Bollerslev, T., Gibson, M.S., Zhou, H., in press. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics*, doi:10.1016/j.jeconom.2010.03.033.

Bollerslev, T., Zhou, H., 2002. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics* 109, 33–65.

Broadie, M., Chernov, M., Johannes, M., 2007. Model specification and risk premia: evidence from futures options. *Journal of Finance* 62, 1453–1490.

Bunn, D.W., 1984. *Applied Decision Analysis*. McGraw-Hill.

Carr, P., Wu, L., 2004. Time-changed Lévy processes and option pricing. *Journal of Financial Economics* 71, 113–141.

Chalamandaris, G., Tsekrekos, A.E., 2010. Predictable dynamics in implied volatility surfaces from OTC currency options. *Journal of Banking and Finance* 34, 1175–1188.

Chernov, M., Ghysels, E., 2000. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics* 56, 407–458.

Christoffersen, P., Jacobs, K., 2004. The importance of the loss function in option valuation. *Journal of Financial Economics* 72, 291–318.

Christoffersen, P., Jacobs, K., Mimouni, K., 2006. *Models for S&P 500 Dynamics: Evidence from Realized Volatility, Daily Returns, and Option Prices*. McGill University.

Christoffersen, P., Jacobs, K., Ornathanalai, C., Wang, Y., 2008. Option valuation with long-run and short-run volatility components. *Journal of Financial Economics* 90, 272–297.

Christoffersen, P., Jacobs, K., Dorion, C., Wang, Y., in press. Volatility components: affine restrictions and non-normal innovations. *Journal of Business and Economic Statistics*, doi:10.1198/jbes.2009.06122.

Christoffersen, P., Mazzotta, S., 2005. The accuracy of density forecasts from foreign exchange options. *Journal of Financial Econometrics* 3, 578–605.

Dawid, A.P., 1984. Statistical theory, the prequential approach. *Journal of the Royal Statistical Society, Series A* 147, 278–290.

- Derman, E., Kani, I., 1994. Riding on the smile. *Risk* 7 (January), 32–39.
- Diebold, F.X., Hahn, J., Tay, A.S., 1999. Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange. *Review of Economics and Statistics* 81, 661–673.
- Diebold, F.X., Todd, A.G., Tay, A.S., 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863–883.
- Duffie, D., Pan, J., Singleton, K.J., 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.
- Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: empirical tests. *Journal of Finance* 53, 2059–2106.
- Dupire, B., 1994. Pricing with a smile. *Risk* 7 (January), 18–20.
- Eraker, B., 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance* 59, 1367–1403.
- Eraker, B., Johannes, M., Polson, N.G., 2003. The impact of jumps in volatility and returns. *Journal of Finance* 58, 1269–1300.
- Fackler, P.L., King, R.P., 1990. The calibration of option-based probability assessments in agricultural commodity markets. *American Journal of Agricultural Economics* 72, 73–83.
- Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.
- Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6, 327–343.
- Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies* 18, 37–84.
- Hong, Y., Li, H., Zhao, F., 2004. Out-of-sample performance of spot interest rate models. *Journal of Business Economics and Statistics* 22, 457–473.
- Huang, J.-Z., Wu, L., 2004. Specification analysis of option pricing models based on time-changed Lévy processes. *Journal of Finance* 59, 1405–1439.
- Jackwerth, J.C., 2000. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* 13, 433–451.
- Jackwerth, J.C., Rubinstein, M., 1996. Recovering probability distributions from option prices. *Journal of Finance* 51, 1611–1631.
- Jiang, G.J., Tian, Y.S., 2005. The model-free implied volatility and its information content. *Review of Financial Studies* 18, 1305–1342.
- Jondeau, E., Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: existence, persistence and comovements. *Journal of Economic Dynamics and Control* 27, 1699–1737.
- Jondeau, E., Rockinger, M., 2009. The impact of shocks on higher moments. *Journal of Financial Econometrics* 7, 77–105.
- Jones, C.S., 2003. The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics* 116, 181–224.
- Jones, C.S., 2006. A nonlinear factor analysis of S&P 500 index option returns. *Journal of Finance* 61, 2325–2363.
- Kang, B.J., Kim, T.S., 2006. Option-implied risk preferences: an extension to wider classes of utility functions. *Journal of Financial Markets* 9, 180–198.
- Kling, J.L., Bessler, D.A., 1989. Calibration-based predictive distributions: an application of prequential analysis to interest rates, money, prices and output. *Journal of Business* 62, 477–499.
- Konstantinidi, E., Skiadopoulos, G., Tzagkarakaki, E., 2008. Can the evolution of implied volatility be forecasted? *Journal of Banking and Finance* 32, 2401–2411.
- Kostakis, A., Panigirtzoglou, N., Skiadopoulos, G., 2009. Asset Allocation with Option-implied Distributions: A Forward-looking Approach. University of Piraeus.
- Kullback, S., Leibler, R.A., 1951. On information and sufficiency. *Annals of Mathematical Statistics* 22, 79–86.
- Liu, J., Pan, J., 2003. Dynamic derivative strategies. *Journal of Financial Economics* 69, 401–430.
- Liu, X., Shackleton, M.B., Taylor, S.J., Xu, X., 2007. Closed-form transformations from risk-neutral to real-world distributions. *Journal of Banking and Finance* 31, 1501–1520.
- Martens, M., Zein, J., 2004. Predicting financial volatility: high-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets* 24, 1005–1028.
- Medvedev, A., Scaillet, O., 2007. Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *Review of Financial Studies* 20, 427–459.
- Melick, W.R., Thomas, C.P., 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.
- Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics* 63, 3–50.
- Rosenberg, J.V., Engle, R.F., 2002. Empirical pricing kernels. *Journal of Financial Economics* 64, 341–372.
- Rosenblatt, M., 1952. Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23, 470–472.
- Santa-Clara, P., Yan, S., 2010. Crashes, volatility, and the equity premium: lessons from S&P 500 options. *Review of Economics and Statistics* 92, 435–451.
- Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis*. Chapman Hall.
- Tay, A.S., Wallis, K.F., 2000. Density forecasting: a survey. *Journal of Forecasting* 19, 235–254.
- Taylor, S.J., 2005. *Asset Price Dynamics Volatility and Prediction*. Princeton University Press.
- Taylor, S.J., Yadav, P.K., Zhang, Y., 2010. The information content of implied volatilities and model-free volatility expectations: evidence from options written on individual stocks. *Journal of Banking and Finance* 34, 871–881.
- Wang, Y.-H., 2009. The impact of jump dynamics on the predictive power of option-implied densities. *Journal of Derivatives* 16 (3), 9–22.
- Yu, W.W., Lui, E.C.K., Wang, J.W., 2010. The predictive power of the implied volatility of options traded OTC and on exchanges. *Journal of Banking and Finance* 34, 1–11.
- Ziegler, A., 2007. Why does implied risk aversion smile? *Review of Financial Studies* 20, 859–904.