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

Mark B. Shackleton, Stephen J. Taylor and Peng Yu
Department of Accounting and Finance, Lancaster University, England

E-Mail: m.shackleton@lancaster.ac.uk, s.taylor@lancaster.ac.uk, p.yu1@lancaster.ac.uk
Telephone: +44 1524 593624

First draft: January 2006
This version: March 2008

For their helpful comments on earlier versions of this paper we thank Yacine Aït-Sahalia, Torben Andersen, Tim Bollerslev, Martin Martens, 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, Lancaster University, London School of Economics, Juan Carlos III University of Madrid, University of Manchester, University of Warwick, University of Zurich, and participants at meetings held by Bachelier Finance Society (Tokyo), Journal of Banking and Finance (Beijing), Centre for Analytical Finance (Sandbjerg), MathFinance workshop (Frankfurt) and European Financial Management Association (Madrid).
A multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices

Abstract

We compare density forecasts of the S&P 500 index from 1991 to 2004, obtained from option prices and daily and five-minute index returns. Risk-neutral densities are given by using option prices to estimate diffusion and jump-diffusion processes, that incorporate stochastic volatility, and then three transformations are used to obtain real-world densities. These densities are compared with historical densities defined by ARCH models.

The best forecasts are obtained from risk-transformations of the risk-neutral densities, for horizons of two and four weeks, 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.

JEL classifications: C14; C22; C53; G13

Keywords: ARCH models; Density forecasts; Index options; Risk-neutral densities; Risk transformations
1. Introduction

Density predictions provide decision takers with more information than forecasts of expected returns and volatilities. The additional information is essential for many risk management activities including the calculation of ‘value at risk’. Central banks are prominent users of density predictions for interest rates, exchange rates, stock market indices and commodity prices. They tend to prefer forward-looking densities obtained from option prices to conditional densities calculated from historical time series, although there is almost no published evidence to support their preference. Our paper provides the first comparison of the predictive accuracy of historical and option-based density forecasts across several forecast horizons.

Option prices reflect competitive opinions about the risk-neutral density of the underlying asset when a set of option contracts expire. Several empirical methods are able to convert option prices into an estimated risk-neutral density for one expiry date, as has been illustrated by Jackwerth and Rubinstein (1996), Melick and Thomas (1997), Ait-Sahalia and Lo (1998), Bliss and Panigirtzoglou (2002) and Taylor (2005). The more difficult problem of estimating the risk-neutral dynamics of the underlying asset, from the most recent option prices for several expiry dates, has received much less attention. The first comprehensive empirical study that incorporates stochastic volatility dynamics is the pricing and hedging paper by Bakshi, Cao and Chen (1997). They summarize daily estimates of jump-diffusion parameters for the S&P 500 index for the four-year period from 1988 to 1991. We estimate risk-neutral parameters from S&P 500 futures prices on each day in the fifteen years from 1990 to 2004 inclusive, initially for the diffusion of Heston (1993) and then with jumps
included in the price process. It is then easy to derive the risk-neutral density for any time horizon\(^1\).

Transformations from risk-neutral (\(Q\)) to real-world (\(P\))\(^2\) densities have been proposed and estimated in several recent papers, commencing with Bakshi, Kapadia and Madan (2003) and Bliss and Panigirtzoglou (2004). These real-world densities have, however, been obtained for option expiry dates alone. Our first contribution is to obtain real-world densities for general forecast horizons; we evaluate forecasts for seven horizons that range from one day to twelve weeks.

All previous studies use full-sample datasets to estimate risk-transformation parameters. The real-world densities are then \textit{ex post}, because a typical density defined at some time \(s\) for the asset price at a later time \(t\) depends on asset prices recorded after time \(s\). Our second contribution is to present results for \textit{ex ante}, real-world densities, which are constructed from the present and past prices for an asset and its options alone.

There is a vast literature that compares volatility forecasts obtained from historical asset prices and current option prices, surveyed by Poon and Granger (2003) and Taylor (2005). 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.

As option forecasts of index volatility are often more accurate than historical forecasts, even when these are based upon intraday returns [Blair, Poon and Taylor (2001), Martens and

---

1 The local volatility models of Dupire (1994) and Derman and Kani (1994) can also provide risk-neutral densities for any horizon until the latest option expiry time. 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, Fleming and Whaley (1998).

2 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.
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 have the highest ranks
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.

Our methodology requires us to specify a risk-neutral, continuous-time, stochastic
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 have a general level of correlation with
price increments. The price dynamics of Heston (1993) are the simplest that satisfy all of our
requirements. These dynamics state that the variance of asset prices follows a square-root
process and, based upon the numerical inversion of characteristic functions, they have closed-
form formulae for densities and option prices.

Like all parsimonious price models, the Heston model has been shown to be imperfect
for pricing options. Analytic characteristic functions are also available for more complicated
price dynamics, that include jumps in prices and/or volatility. We also evaluate an affine
process that includes price jumps generated by a Poisson process. We do not consider further

---

3 For example by Bakshi et al (1997), Andersen, Benzoni and Lund (2002), Pan (2002), Jones (2003),
Christoffersen, Jacobs and Moinou (2006) and Ait-Sahalia and Kimmel (2007). Also, Christoffersen et al
(2008, forthcoming) show that two volatility components fit option prices more accurately than one component.
4 Special cases of the general affine jump-diffusions of Duffie, Pan and Singleton (2000) have been investigated
(2002), Liu and Pan (2003), Eraker, Johannes and Polson (2003), Eraker (2004), Broadie, Chernov and Johannes
(2007) and Medvedev and Scaillet (2007). Processes incorporating jumps that arrive at an infinite rate are
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 mathematical sophistication of the Heston price dynamics and its jump-diffusion extensions may be counterproductive when 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 representative-agent model. They estimate the risk parameter by minimizing the diagnostic test statistic of Berkowitz (2001); this test assesses the uniformity and independence of the cumulative probabilities of observed prices. Bliss and Panigirtzoglou 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. Anagnostou-Basioudis et al (2005) for the S&P 500, Kang and Kim (2006) and Liu et al (2007) for the FTSE 100, also provide empirical results for utility transformations. It must be noted, however, that empirical estimates of implied risk aversion are incompatible with a standard consumption-based framework (Jackwerth (2000), Rosenberg and Engle (2002), Ziegler (2007)). Consequently, 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 observed index levels on the monthly, option expiry
dates. We apply the same calibration transformation, but instead employ \textit{ex ante} transformation parameters. These are obtained separately for seven forecast horizons, that are not restricted by the timing of option expiry dates. We also provide the first analysis of two further transformations from risk-neutral to real-world densities: one assumes that affine (jump-) diffusion price dynamics apply in the real world by incorporating appropriate risk-premium functions, and the other applies a non-parametric 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 \textit{ex ante} parameters and forecasts is presented in Section 3 and the Appendix. 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. The empirical results are all contained in Section 5. We find strong evidence that the best forecasts are given by the risk-transformations of the risk-neutral densities, for horizons of two and four weeks, when the ranking criterion is the out-of-sample likelihood of observed index levels. For horizons between six and twelve weeks, the empirical evidence also favors densities extracted from option prices. In contrast, the one-day-ahead historical densities are superior to all option-based densities but they obtain comparable likelihoods for the one-week horizon. A mixture of the best historical densities and each set of $P$-densities outperforms both components of the mixture, for these two shortest horizons. Standard diagnostic tests show that the best historical densities and the risk-transformed, option-based densities nearly
always pass these tests, while the other density forecasting methods have more test failures. Finally, Section 6 summarizes our conclusions.

2. Density forecasts

2.1 Risk-neutral densities

Almost all of the methods that estimate risk-neutral densities (RNDs) from option prices can only provide densities for the underlying asset at its option expiry times. To obtain densities for all future times it is necessary to specify the risk-neutral dynamics of the underlying asset price. We consider three specifications. 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. Ideally the assumed price dynamics provide panels of theoretical option prices that are similar to panels of observed option prices, when the price dynamics parameters are optimized.

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 continuous-time, risk-neutral dynamics for a futures price, \( F_t \), are given by supposing that the stochastic variance, \( V_t \), follows the square-root process of Cox, Ingersoll and Ross (1985):

\[
dF/F = \sqrt{V} dW_1, \tag{1}
\]

and

\[
dV = \kappa(\theta - V)dt + \xi \sqrt{V} dW_2, \tag{2}
\]
with correlation $\rho$ between the increments of the two Wiener processes, $W_{1,t}$ and $W_{2,t}$. The time $t$ is measured in years. The special case of GBM, with constant volatility $\theta$, 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

$$dF_s/F = \sqrt{V} dW_t + (e^{J} - 1) dN - \lambda \bar{\mu}_J dt,$$

with $N_t$ a Poisson process that has intensity $\lambda$; 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 $\log(F_t)$, that are normally distributed with mean $\mu_J$ and variance $\sigma_J^2$; the average size of the proportional jumps in $F_t$ 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 that imply the same continuous-time process and the same parameters ($\kappa, \theta, \xi, \rho, \lambda, \mu_J, \sigma_J$) are applicable to all contracts.

Heston (1993), Bakshi et al (1997) and Duffie et al (2000) provide analytic formulae for the characteristic function of $\log(F_T)$, conditional upon initial values $F_0$ and $V_0$. Our notation for this conditional characteristic function is $\tilde{g}(\psi) = E^Q[\exp(i\psi \log(F_T))]$, with $\psi$ a real number and $Q$ the risk-neutral measure. The following inversion formula then gives the risk-neutral density of $F_T$, denoted by $g_{Q,T}(x)$, for positive values of $x$:

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

(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:

$$c(F_0, K) = e^{-rT} (F_0P_1(F_0, K) - KP_2(F_0, K))$$

(5)

where $r$ is the risk-free rate, $P_2(F_0, K)$ is the risk-neutral probability that the option expires in-the-money and $P_1(F_0, K)$ is a probability for the same event when a different measure is applied. Both $P_1(F_0, K)$ and $P_2(F_0, K)$ are obtained from standard inversion formulae.

2.2 Real-world densities

The risk-neutral density will always be incorrectly specified if it is used to make statements about real-world probabilities determined by a real-world measure $P$. Although a premium that compensates for price risk ensures this conclusion, it is possible that there are also volatility and jump risk premia. Furthermore, the assumption of a square-root process for volatility is at best a convenient approximation.

Transformations from risk-neutral to real-world densities rely on assumptions. These can be provided by a representative agent model, by specifying risk-premia functions or by statistical calibration theory. We prefer the additional flexibility provided by either two or three risk-premium terms, and by the two-parameter calibration transformation of Fackler and King (1990) and Liu et al (2007), to the one-parameter utility transformations of Bliss and Panigirtzoglou (2004) and Anagnou-Basioudis et al (2005). We also investigate a non-parametric calibration transformation.

Risk-premium transformations

An affine real-world diffusion process is defined by including linear drift terms in both the price and the variance equations, thus:
\[ \frac{dF}{F} = \eta_1 V dt + \sqrt{V} d\tilde{W}_1, \]

and

\[ dV = [\eta_2 V + \kappa(\theta - V)]dt + \xi \sqrt{V} d\tilde{W}_2. \]  \hspace{1cm} (6)

The assumption of linear functions for the risk premia ensures the availability of analytic formulae for the real-world, characteristic functions of future prices. The inversion formula (4) then provides real-world densities \( g_{P,T}(x) \) that depend on the premium coefficients, \( \eta_1 \) and \( \eta_2 \). To define real-world, jump-diffusion dynamics we additionally increase the mean jump size by \( \eta_3 \). Then

\[ \frac{dF}{F} = (\eta_1 V - \lambda \mu_J) dt + \sqrt{V} dW_1 + (e^J - 1) dN, \quad J_t \sim N(\mu_J + \eta_3, \sigma_J^2), \]  \hspace{1cm} (7)

and the risk premium becomes \[ [\eta_1 V + \lambda (\exp(\eta_3) - 1) \exp(\mu_J + 0.5\sigma_J^2)] F dt. \]

**Calibration transformations**

At time 0, suppose \( g_{Q,T}(x) \) and \( G_{Q,T}(x) \) respectively denote the risk-neutral density and cumulative distribution function (c.d.f.) of the random variable \( F_T \), and then define \( U_T = G_{Q,T}(F_T) \). Following Bunn (1984), Dawid (1984) and Diebold, Hahn and Tay (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 \( F_T \) is

\[ G_{P,T}(x) = C_T(G_{Q,T}(x)). \]  \hspace{1cm} (8)

Also, the real-world density of \( F_T \) is given by

\[^5\] For any random variable \( X \), with c.d.f. \( G(x) \) for a stated measure, the random variable \( U = G(X) \) 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 \( U_T = G_{Q,T}(F_T) \) is only a draw from a uniform distribution when \( G_{Q,T}(x) \) is correctly specified and the risk-neutral measure is identical to the real-world measure.
with \( u = G_{Q,T}(x) \) 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, recommended by Fackler and King (1990) in their innovative study of densities obtained from commodity option prices. The calibration density is then the Beta density,

\[
c_T(u) = u^{j-1}(1-u)^{k-1} / B(j,k), \quad 0 \leq j \leq 1,
\]

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

\[
g_{P,T}(x) = \frac{G_{Q,T}(x)^{j-1}(1-G_{Q,T}(x))^{k-1}}{B(j,k)} g_{Q,T}(x).
\]

Alternatively, a non-parametric calibration function can be estimated from a historical set of observations of the quantity \( u = G_{Q,T}(x) \), with \( x \) representing a typical observed value of the futures price \( F_T \). We calculate kernel estimates of \( C_T(u) \) and \( c_T(u) \), using methods presented in the Appendix. These estimates, given by (34) and (37), are substituted into (8) and (9) to provide further real-world c.d.f.s and densities.\(^6\)

### 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,

---

\(^6\)Kling and Bessler (1989) and Diebold et al (1999) have estimated non-parametric calibration functions from real-world densities, respectively for Treasury Bill and foreign exchange rates.
one week or several weeks. The one-period returns are \( r_t = \log(F_t/F_{t-1}) \); here \( F_{t-1} \) and \( F_t \) 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, \( f(r|I_t) \), for the next return, \( r_{t+1} \). The historical density for the next end-of-period price, \( F_{t+1} \), is then:

\[
g(x|I_t) = \frac{f(r|I_t)}{x},
\]

with \( r = \log(x/F_t) \). We describe four specifications for the historical density \( f(r|I_t) \).

The simplest credible ARCH model for a stock market index is the GJR(1, 1) model of Glosten, Jagannathan and Runkle (1993). The conditional variance \( h_t \) is then an asymmetric function of returns. We define the model as follows, with a constant conditional mean \( \mu \):

\[
r_t = \mu + \varepsilon_t
\]

\[
\varepsilon_t = h_t^{1/2} z_t, \quad z_t \sim \text{i.i.d.}(0,1),
\]

\[
\frac{h_t}{D_t} = \omega + \frac{(\alpha_1 + \alpha_2 d_{t-1})\varepsilon_{t-1}^2 + \beta_{GJR} h_{t-1}}{D_{t-1}},
\]

\[
d_{t-1} = 1 \quad \text{if} \quad \varepsilon_{t-1} < 0,
\]

\[
= 0 \quad \text{otherwise}.
\]

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 a specification that we refer to as the 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 \( \nu \), as first evaluated in Bollerslev (1987).
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:

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

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 ($\varepsilon_{t-1}$).

### 2.4 Mixture densities

At some general time $t$, both ARCH densities and option-based densities may contain incremental information about the asset price at a later time $t + T$. Consequently, we also evaluate the mixture density:

\[
g_{mix,T}(x) = \alpha g_{P,T}(x) + (1 - \alpha)g_{ARCH,T}(x), \quad 0 \leq \alpha \leq 1.
\]

As option traders know the historical price information, it is possible that $\alpha = 1$ 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 risk-neutral, real-world and historical densities are all parametric. We always use \textit{ex ante} estimates of parameters, to ensure that all density forecasts are evaluated out-of-sample. Consequently, all parameters required at time \( t \) are estimated by only using information available at time \( t \). While the risk-neutral parameters are estimated daily from option prices by minimizing a least-squares function, all the other parameters are estimated by maximizing the log-likelihood function of selected, observed asset prices.

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. For the Heston process defined by (1) and (2), at the end of day \( n \) we estimate the initial variance \( V_n \), the three volatility parameters, \( \kappa_n, \theta_n \) and \( \xi_n \), and the correlation \( \rho_n \) between the price and volatility differentials. Suppose \( N_n \) European, call option contracts are traded on day \( n \), labeled by \( i = 1, \ldots, N_n \), with strikes \( K_{n,i} \), expiry times \( T_{n,i} \) and market prices \( c_{n,i} \); also, suppose \( F_{n,i} \) is the futures price for the asset after \( T_{n,i} \) years. Then the five Heston \( Q \)-parameters are estimated by minimizing

\[
\sum_{i=1}^{N_n} (c_{n,i} - c(F_{n,i}, K_{n,i}, T_{n,i}, V_n, \kappa_n, \theta_n, \xi_n, \rho_n))^2, \tag{16}
\]

\footnote{We only include call prices in the estimation function, (16). We explain in Section 4.3 that the put prices in our database are converted to equivalent European call prices, using the put-call parity relationship, and are then included in (16).}
with \( c(.) \) the Heston pricing formula, given by (5)\(^8\). Likewise, the eight \( Q \)-parameters for the jump-diffusion process defined by (2) and (3) are also estimated at the end of each trading day.

The remaining parameters that appear in the real-world, historical and mixture densities are the price-of-risk parameters \( \eta_1, \eta_2 \) and \( \eta_3 \), the calibration function parameters \( j \) and \( k \), the ARCH parameters \( \omega, \alpha_1, \alpha_2, \beta, \omega_{Intra}, \gamma_1, \gamma_2 \) and \( \beta_{Intra} \), and the mixture parameter \( \alpha \). As explained in the Appendix, all these parameters are estimated \( \text{ex ante} \) 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 methods, including several surveyed by Tay and Wallis (2000). Our forecasts are assessed using numerical criteria, that include the out-of-sample likelihood and the values of diagnostic test statistics.

**Likelihood criteria**

For a fixed forecast horizon, suppose method \( m \) provides a series of density forecasts \( g_{m,t}(x) \), made at integer times \( v, v+1, \ldots, w \) for the asset price at times \( v+1, v+2, \ldots, w+1 \). The out-of-sample, log-likelihood of observed asset prices for method \( m \) equals

\[
L_m = \sum_{t=v}^{w} \log(g_{m,t}(F_{t+1})).
\] (17)

\(^8\) Christoffersen and Jacobs (2004, page 316) conclude that (16) is a “good general-purpose loss function in option valuation applications”. It is the preferred loss function in the study of S&P 500 dynamics by Christoffersen et al (2006).
Our preferred method has the maximum value of $L_m$. The same criterion has been used by Bao, Lee and Saltoglu (2007) and Liu et al (2007) to compare density forecasting methods applied to equity indices.

Note that if one of the methods, say method $M$, correctly specifies the densities then it will have the highest expected log-likelihood. This follows from a property of the information criterion of Kullback and Liebler (1951), defined by

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

This criterion is positive whenever the two densities are continuous and distinct. Consequently, $E[L_M] > E[L_m]$ 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 the likelihood criterion $L_m$ across methods 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, which is determined by the likelihoods of the methods. Assuming that one of $N$ methods is correct, and selecting an uninformative prior distribution, the posterior probability that method $m$ is correct equals

$$p_m = \frac{\exp(L_m)}{\sum_{n=1}^N \exp(L_n)}. \quad (19)$$

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\(^9\). The null hypothesis that two methods $m$ and $n$ have equal expected

---

\(^9\) Corradi and Swanson (2006) describe an alternative criterion that can be used to perform such tests.
log-likelihood can be tested by assessing the statistical significance of the sample mean of the
time series of log-likelihood differences:

\[ d_t = \log(g_{m,t}(F_{t+1})) - \log(g_{n,t}(F_{t+1})), \quad v \leq t \leq w. \]  

(20)

We calculate the AG test statistic as

\[ AG = (w - v + 1)^{0.5} \frac{\bar{d}}{s_d} = (w - v + 1)^{-0.5} \frac{(L_m - L_n)}{s_d}, \]  

(21)

with \( \bar{d} \) and \( s_d \) the average and the standard deviation of the terms \( d_t \). We suppose AG has a
standard normal distribution when the null is true, since we find there is no discernible
autocorrelation in the time series \( \{d_t\} \).

**Diagnostic criteria**

We 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, Gunther and Tay
(1998) and applied in several financial studies.\(^\text{10}\) For a general method \( m \) these probabilities
are defined by

\[ u_{t+1} = \frac{F_{t+1}}{g_{m,t}(x)}dx, \quad v \leq t \leq w. \]  

(22)

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

\[ \tilde{C}(u) = \frac{1}{w-v+1} \sum_{t=v+1}^{w+1} S(u-u_t), \]  

(23)

\(^\text{10}\) Interesting recent examples are Bliss and Panigirtzoglou (2004), Hong, Li and Zhao (2004), Anagnou-
with \( S(x) = 1 \) if \( x \geq 0 \), and \( S(x) = 0 \) if \( x < 0 \). The test statistic is then

\[
KS = \sup_{0 \leq u \leq 1} \left| \hat{C}(u) - u \right|.
\] (24)

Secondly, we apply the test of Berkowitz (2001) to the numbers \( y_t \) defined by \( \Phi(y_t) = u_t \), with \( \Phi(.) \) 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 \( \chi^2_3 \); LR3 equals \( 2(L_1 - L_0) \), with \( L_0 \) and \( L_1 \) the maximum log-likelihoods of \( \{y_{v+1}, \ldots, y_{w+1}\} \), respectively for the null and alternative hypotheses.

4. Data

The underlying assets for the density forecasts are futures contracts written on the S&P 500 index, traded at the CME. We investigate density forecasts for the futures price, rather than the spot level of the index, because of the availability of contemporaneous, settlement prices for futures and options contracts. A second advantage of working with futures data is that it is not necessary to consider dividend payments on the stocks that define the index. Option prices for S&P 500 futures have also been studied in related research by Bates (2000, 2006), Bliss and Panigirtzoglou (2004), Anagnou-Basioudis et al (2005), Jones (2006) and Broadie et al (2007).
4.1 Futures prices

End-of-day settlement prices and intraday prices for S&P 500 futures contracts are studied from 28 April 1982 until 31 December 2004. The settlement and intraday prices are respectively provided by the CME and Price-Data.com. S&P 500 futures are contracts written on the index level on the third Fridays of March, June, September and December. All returns are calculated from the nearest contract, except on the expiry days and on the Thursdays that precede them when the next contract is used.

The high-frequency, realized variances are calculated from five-minute 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 S&P 500 futures contracts are traded from 08:30 to 15:15 local time at the CME, we use 81 intraday returns for each day. The realized variance for day $t$ is defined as the sum of the squares of the five-minute returns $r_{t,i}$:

$$Intra_t = \sum_{i=1}^{81} r_{t,i}^2.$$  \hspace{1cm} (25)

4.2 Interest rates

Three-month, six-month and one-year U.S. Treasury bill rates are converted to continuously compounded rates. The risk-free rate $r$ employed in an option pricing formula is then 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 end-of-day settlement prices of options on S&P 500 futures contracts for fifteen years, from 2 January 1990 to 31 December 2004. These prices post-date the crash of
October 1987, after which the skewness of risk-neutral densities became significantly negative (Jackwerth and Rubinstein (1996)).

We consider prices for each option-on-futures contract that expires on the same Friday as its underlying futures contract. We do not use the prices of the remaining option contracts that expire one or two months earlier than the futures, because they are less actively traded. Option contracts with seven or less calendar days until expiry are excluded.

Call and put settlement prices for the same strike and expiry date should 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 only use the information provided by the prices of OTM and ATM options, since the in-the-money option contracts are usually less actively traded. No volume filter is applied to the option price data.

The option contracts are American. We obtain equivalent European option prices from the American prices, that 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 approximating the premia by using the formulae of Barone-Adesi and Whaley. Finally, the put-call parity equation for options on futures is used to obtain equivalent European call prices from the European OTM put prices.

We use 435,100 option prices for the 3,777 trading days from 1990 to 2004. The average number of option prices studied 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 (defined as $F/K$), and the time-to-expiry of the contracts that provide the observed prices. There are far more prices for deep-OTM put options than for deep-OTM call options, reflecting the greater demand for put options.
4.4 Volatility comparisons

Figure 1 shows, as a dark curve, the time series of the implied volatilities obtained from the nearest-to-the-money options that are nearest to expiry. Figure 1 also shows, as light dots, the time series of the annualized, intraday, realized standard deviation, defined by $\sqrt{252Intra_t}$. 

The intraday and option measures of volatility move together, as expected, but the implied level is higher than the intraday, realized level. The first reason for this is that the intraday 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 ($0.93 \times 10^{-4}$) equals 80% of the variance of the daily returns ($1.17 \times 10^{-4}$). The second reason is a systematic difference between historical variance and risk-neutral variance: the average of the squared implied volatilities on Figure 1, stated in daily units ($1.32 \times 10^{-4}$), 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 U.S. equity indices has been documented in several papers\footnote{For example, in Jackwerth and Rubinstein (1996), Chernov and Ghysels (2000), Pan (2002), Bakshi and Kapadia (2003), Jones (2003) and Bollerslev, Gibson and Zhou (2005).}.
5. Empirical results for density forecasts

Density forecasts are evaluated for the fourteen years from January 1991 until December 2004 inclusive. The option prices during 1990 are only used to contribute to the \textit{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 twelve 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

The final one-day ahead densities, calculated from the information available on 30 December 2004, are shown on Figures 2a, 2b and 2c. Figure 2a shows the four historical densities; the conditional $t$-densities have higher peaks and fatter tails than the conditional normal densities. Figures 2b and 2c 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 non-parametric calibration and the risk-premium transformations. It can be seen that 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 Figures 3a, 3b and 3c. The real-world standard deviations are again less
than the risk-neutral levels. From Figures 2c and 3c it is seen that the illustrative P1-densities are very similar to the corresponding P3-densities.

5.2 Historical densities

The parameters of the one-day ahead, ARCH densities have been estimated from daily and intraday return data that commences on 4 January 1988 and thus post-dates the crash of October 1987. 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 $\nu$, the averages include $\nu = 4.70$, $\alpha_1 = 0.009$, $\alpha_2 = 0.046$ and $\beta = 0.960$, with the persistence parameter equal to $\alpha_1 + 0.5 \alpha_2 + \beta = 0.992$. 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 data that commences on 28 April 1982. All the averages for the degrees-of-freedom parameter indicate a high level of excess kurtosis in the conditional distributions. The averages of $\nu$ are 6.20 and 7.52 for the one-week returns, respectively for the GJR-t and Intra-t models, and they are 4.95 and 4.55 for the longest return period of twelve weeks.

Comparisons of the log-likelihoods of the four ARCH specifications favor the Intra-t model for all return periods, as will be shown in Section 5.7. It is possible that the Intra-t densities can be enhanced by using either the parametric or the non-parametric calibration transformation, implemented using (11), (36) and $ex \ ante$ information. We give results for these transformations later in Sections 5.7 and 5.8.

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:

\[
\frac{dF}{F} = \sqrt{V}dW_1 \quad \text{and} \quad dV = \kappa(\theta - V)dt + \xi \sqrt{V}dW_2, \quad (26)
\]

and with correlation \( \rho \) 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 \( \log(F_t) \) that have intensity \( \lambda \), mean \( \mu_J \) and standard deviation \( \sigma_J \).

Several previous studies have estimated these parameters from S&P 500 index levels and/or option prices. As researchers use a variety of markets (underlying and/or options), derivatives prices (none, spot options or futures options), different sample periods and different estimation methodologies, it is not surprising that their parameter estimates are rather diverse. We find that our median estimates are generally similar to previous risk-neutral estimates.

Our risk-neutral parameters are estimated by minimizing the criterion in (16), which is equivalent to minimizing 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 the jump intensity \( \lambda \) is 0.47 per annum and the median estimate of the average jump size \( \mu_J \) reduces the asset price by 6%. Our average risk-neutral estimate of \( \lambda \) equals 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.\(^{12}\)

\(^{12}\) The time-varying, real-world estimates of \( \lambda \) in Santa-Clara and Yan (2006) provide similar summary statistics to our risk-neutral estimates. Their mean and standard deviation, for the period from 1996 to 2002, are respectively 0.795 and 0.714, while our statistics are 0.693 and 0.715.
Each time series of MSE values has a median that is much lower than the standard deviation, which indicates that the distribution of MSE contains some extreme values. A positively skewed distribution is also noted for $\kappa$, $\lambda$, and $\sigma_j$.

The stochastic variance $V_t$ reverts towards the level $\theta$. Our median estimate for the pure-diffusion process is 0.0452, which is 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 $\theta$ is determined by $\kappa$. Our median estimates of $\kappa$ 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 $\xi$. Estimates obtained solely from option prices, such as our median values of 0.79 (without jumps) and 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 $\rho$ 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.

The divergence between risk-neutral and real-world estimates of $\xi$ and $\rho$ has recently been emphasized by Broadie et al (2007). It follows that the affine price dynamics considered are unable to explain all the empirical properties of asset and option prices. To do so, we might consider more complicated dynamics and additional sources of risk, provided for
example by uncertainty about the frequency and magnitude of jumps (Liu, Pan and Wang (2005)). Alternatively, the divergence might be explained by the buying pressure effect investigated by Bollen and Whaley (2004).

There is no guarantee that a more complicated option pricing model will produce more accurate predictive densities. We find that the eight-parameter, jump-diffusion specification does not yield more successful real-world densities than the simpler five-parameter, pure-diffusion specification. Consequently, we focus on the pure-diffusion methods in Sections 5.4 to 5.8 and defer our comparisons with the jump-diffusion methods to Section 5.9.

5.4 Cumulative probabilities from risk-neutral densities

A set of one-day-ahead, risk-neutral densities, now denoted by \( g_{Q,t}(x) \), provides cumulative distribution functions \( G_{Q,t}(x) \) that can be evaluated at the next, observed futures prices, \( F_{t+1} \), to define observed probabilities defined by \( u_{t+1} = G_{Q,t}(F_{t+1}) \). As expected, the observed probabilities are incompatible with a uniform distribution.

The sample c.d.f. calculated from a time series \( \{u_{t+1}\} \) is denoted by \( \tilde{C}(u) \) and defined by (23). We show the differences between sample and uniform probabilities, \( \tilde{C}(u) - u \), as the dark curve on Figure 4b for the Heston densities. It is obvious 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 \( \left| \tilde{C}(u) - u \right| \) equals 6.6%; this value of the Kolmogorov-Smirnov test statistic rejects the null hypothesis of a uniform distribution at the 0.01% level. The deviations \( \tilde{C}(u) - u \) for the lognormal densities are similar, as can be seen from Figure 4a.

The shape of the deviation curves can be explained primarily by the fact that the risk-neutral, standard deviations are, on average, significantly higher than the historical standard
deviations, as we observed in Section 4.4. Consequently, the risk-neutral probabilities of extreme price changes exceed the real-world probabilities.

5.5 Calibration transformations

A non-parametric estimate of the real-world density of the probabilities \( u_{t+1} \) is provided by the empirical, non-parametric, calibration density function stated in (37). This estimated density, \( \tilde{c}(u) \), is shown by the light curves on Figures 5a and 5b, respectively for the one-day-ahead lognormal and Heston cases using the data from 1991 to 2004.

The time series averages of the \textit{ex ante} estimates for the parametric calibration transformation applied to the one-day-ahead, risk-neutral Heston densities are \( j = 1.434 \) and \( k = 1.412 \). The corresponding calibration density can be found from (10) and it is plotted as the dark curve on Figure 5b. It can be seen that the parametric and non-parametric calibration densities have similar shapes, except near the end points of the distribution. The \textit{ex ante} estimates of \( j \) and \( k \) vary between 1.3 and 1.6 and nearly always have \( j > k \). The corresponding \textit{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, which define the calibration density shown by the dark curve on Figure 5a.

The calibration methodology is intended to produce real-world densities whose observed probabilities \( u_{t+1} \) are uniformly distributed. After applying the parametric and the non-parametric calibration transformations, the one-day-ahead deviations \( \tilde{C}(u) - u \) estimated \textit{ex ante} from all the data are shown as light curves on Figures 4a and 4b. It can be seen that these deviations are much nearer to zero than those for the risk-neutral densities, particularly for the non-parametric transformation.

Similar results and conclusions are obtained for the one-week-ahead densities. The time series averages of the Heston-estimates of \( j \) and \( k \) are 1.424 and 1.409 respectively. For
horizons of two or more weeks, the average Heston-estimate of $j$ is between 1.45 and 1.58, and it is always more than the average estimate of $k$ which ranges from 1.30 to 1.43.\textsuperscript{13}

Figure 6a shows the four-week-ahead deviations $\tilde{C}(u) - u$ for the risk-neutral lognormal densities and their derived real-world densities, while Figure 6b shows the corresponding deviations that are based upon the risk-neutral Heston densities.

\textbf{5.6 Risk-premium transformations}

The third transformation of the risk-neutral Heston densities into real-world densities adjusts the drift rates of the price and the volatility and thereby incorporates both price and volatility risk premia. The premia coefficients $\eta_1$ and $\eta_2$ in the bivariate diffusion defined by (6) have been estimated separately for each of the seven horizons. These estimates should be similar across horizons if the assumed risk-neutral and real-world dynamics are correct. We find that the seven estimates of $\eta_1$ (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 $\eta_2$ (the variance risk premium per unit of variance) are negative, which is compatible with the evidence for a negative volatility premium cited in Section 4.4. 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.\textsuperscript{14}

\textsuperscript{13} The estimates of the calibration parameter $k$ comprehensively fail to satisfy the constraint $k < 1$ for all horizons and therefore the empirical risk transformations implied by the parametric calibration functions can not be reconciled with a representative agent model (Liu et al (2007)). Ziegler (2007) provides a detailed theoretical analysis of several potential explanations of this empirical conclusion.

\textsuperscript{14} 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.
With $\eta_1 > 0$ and $\eta_2 < 0$, the risk-premium 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, Figure 4b shows that the risk-premium transformation reduces the magnitudes of the deviations $\tilde{C}(u) - u$ as expected. It can also be seen that the deviation curves for the risk-premium and the parametric calibration transformations are very similar.

A risk-premium transformation of the risk-neutral lognormal densities has also been investigated. Only the single risk parameter $\eta_1$ is then available, which improves the means but not the standard deviations of the densities. Consequently, the transformation only changes 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 observed futures prices from January 1991 until December 2004, for thirteen ex ante density forecasting methods. These log-likelihoods are given for non-overlapping forecasts, made for seven horizons that range from one day to twelve 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.

Historical methods

Initially we 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. The AG test defined by (21) rejects the null that the Intra-t method is not better than an alternative method at the 1% level, for each of the three possible alternatives and each of the two shortest horizons.

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 one of the two calibration transformations to the Intra-t densities improves some of the log-likelihood values. Table 3 shows that one 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.

Ten univariate methods

Likelihood comparisons are now made between ten 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 risk-neutral densities 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 refers to the parametric calibration transformation, P2 to the non-parametric calibration transformation and P3 to the risk-premium transformation, respectively defined by (11), (36) and (6).
One-day horizon

The log-likelihoods, in excess of the GJR benchmark value, are as follows in descending order: 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 non-parametric 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-premium transformation P3 ranks behind the statistical transformations P1 and P2 for the Heston densities. Finally, as expected the Q-densities are seen to be far inferior to their related P-densities, reflecting the fact that the transformations are able to diminish the impact of systematic mis-specification in the risk-neutral 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.

Horizons from one to four 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 Heston-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 shortest horizon; the five one-week numbers in Table 3 range from 32.8 to 41.5, for two weeks from 22.4 to 27.8 and for four weeks from 13.4 to 23.2.
Next 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.

Each set of option $P$-densities always outperforms the corresponding set of $Q$-densities. The differences vary between 15.8 and 33.0 for one week, 7.6 and 14.2 for two weeks and from 0.9 to 7.2 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.

**Horizons from six to twelve 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 twelve 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 twelve-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.

**Mixtures**

Mixture densities are defined by (15). We now consider the log-likelihoods for mixtures defined by a fraction $\alpha$ of an option-based density added to a fraction $1-\alpha$ of the Intra-t density. We assess these mixtures because we do not know *ex ante* that the Intra-t-P2 densities are superior to the Intra-t densities. Table 4 shows the log-likelihoods of *ex ante* mixture densities and the time series averages of the *ex ante* estimates of $\alpha$.\(^{15}\)

Each mixture of an option-based P-density and the corresponding Intra-t density has a higher log-likelihood than both the P-density 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 142.0 to 159.8 and all these statistics exceed the best univariate result, namely 140.8 for the Intra-t-P2 densities. Combining Intra-t with either the Heston-P1 or the Heston-P3 densities 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 7 mixture methods. The most successful mixture gives a 38% weight, on average, to the Heston-P1 densities and the remaining 62% weight to the Intra-t densities. 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).

\(^{15}\) The log-likelihoods attained by simple averages ($\alpha = 0.5$) are often near to those obtained by *ex ante* selection of $\alpha$. Simple averages are notably inferior, however, when Q-densities contribute to the mixture and the horizon is short.
Three effects are noted as the forecast horizon increases. First, the average weight given to the five sets of $P$-densities increases. These averages are 37% for the one-day horizon, 59% for one week, 69% for two weeks and 71% for four weeks. Second, the total Bayesian probability for the mixtures decreases, from 100% for the one-day horizon to 77% for one week, 61% for two weeks and 45% for four weeks. Third, the AG test values decrease; for the one, two and four-week horizons 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 Kolmogorov-Smirnov (KS) test statistic, defined by (24) as the maximum value of $\left|\bar{C}(u) - u\right|$, can be used to test the null hypothesis that a set of densities are correctly specified. This test makes the assumption that the observed cumulative probabilities are observations of independent random variables. Figures 4a and 4b show that there are high values of $\left|\bar{C}(u) - u\right|$ for the risk-neutral densities (RNDs) for short horizons, so that the KS test establishes that these densities are indeed mis-specified.

Table 5 lists the percentage p-values for the KS test for the thirteen ex ante density forecasting methods, for each of the seven horizons. As the null hypothesis is rejected at the $\alpha\%$ 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 that 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 non-parametric transformation to the Intra-t and the risk-neutral densities have the most satisfactory p-values: 19 of the 28 p-values exceed 50% and their minimum is 18%. Selected
KS p-values for the one-day horizon are 88% for lognormal-P2, 87% for Intra-t-P2, 47% for Heston-P2, 15% for Intra-t and 9% for Intra-t-P1, compared with a maximum of 0.8% for the other eight sets of one-day densities. We also note that the Heston-P1 and the Heston-P3 densities provide satisfactory p-values, except for the one-day horizon when the p-values are less than 0.5%.

A good specification of the tail values of the densities may be more important than a good overall fit for some purposes. Tail comparisons can be made by restricting the calculation of the maximum of \(|\hat{C}(u) - u|\) to the tail regions \(0 \leq u \leq a\) and \(1 - a \leq u \leq 1\). We have set \(a\) equal to 0.025, 0.05 or 0.1 and find that the Intra-t-P2, lognormal-P2 and Heston-P2 generally outrank the other methods. For all six tail areas evaluated, each of these three P2-methods is always ranked in the top five methods for the one-day and one-week horizons.

The likelihood-ratio test statistic of Berkowitz (2001), denoted by LR3, tests the null hypothesis that the numbers \(y_t = \Phi^{-1}(u_t)\) are i.i.d. observations from a standard normal distribution against the alternative that they are from a stationary, 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, once more for the thirteen \textit{ex ante} density forecasting methods and for each of the seven horizons.

The MLEs of the autoregressive 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 \(\{u_t\}\) are composed of independent observations. The corresponding MLEs for the one-week forecasts are, however, all between \(-0.12\) and \(-0.08\) inclusive; they reject the null hypothesis that the autoregressive parameter is zero at the 5% level for all the historical and all the option-based sets of densities. There is no significant evidence of time-series dependence for horizons of two weeks or longer.
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 historical volatility being lower (on average) than risk-neutral volatility.

The null distribution of LR3 is \( \chi^2_3 \) 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, rejections of the null at the 5% level occur 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 density forecasting methods whose densities always pass the LR3-test at the 5% level are the Heston-P1 and Heston-P3 methods. The highest significant values of LR3 are for the Q-densities, which is a consequence of the substantial difference between the MLEs of their variances and the null value of one. With only one exception, all of the other significant values of LR3 can be explained either by an incorrect normal assumption about the conditional shape of a historical density or by the negative estimates of the autoregressive 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 \( \log(F_t) \) 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\textsuperscript{16}. The three values of the differences $D$ are all negative for the one-day comparisons. There are two positive values for the one-week horizon, but they are smaller than the magnitudes of the respective negative values for the one-day horizon. There are only 4 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.

Further differences $D$ have been calculated for mixtures of the Intra-t and option-based densities. The one-day difference is only 3.4 ($\text{AG} = 1.50$) for the mixtures of Intra-t and risk-neutral densities, compared with 62 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.

The diagnostic test statistics are generally similar for the Heston and JD-densities. At the 5\% level, there are 5 out of 28 significant KS test values for the four Heston sets combined with the seven forecast horizons. The corresponding rejections count is 3 out of 28 for the JD-densities. These counts are 4 and 7 out of 28 for the LR3 test, respectively for the Heston and JD-densities, again at the 5\% level. Excluding the risk-neutral densities reduces the counts to 2 (H, KS), 1 (JD, KS), 1 (H, LR3) and 2 (JD, LR3).

\textsuperscript{16} The three risk transformations use past prices to learn about systematic differences between risk-neutral and real-world densities. The increases in the log-likelihood values from this learning are higher when jumps are excluded from the RNDs. This effect offsets the improvements in the RNDs obtained by incorporating jumps into the RNDs.
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 have provided the first evidence that 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 five-minute 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 twelve weeks, the empirical evidence continues to favor option-based methods. 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. 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 extrapolations. This may explain why the best historical densities are relatively more successful than the real-world densities for the two shortest horizons.

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 non-parametric 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 that utilize the information provided by high-frequency returns.

**Appendix: 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 *n* by $\Theta_n$ and $\Theta^J_n$. Here we show how risk parameters, other relevant parameters and the non-parametric calibration density are estimated within a diffusion framework. The same methods are also used to
estimate parameters in a jump-diffusion framework by replacing $\Theta_n$ by $\Theta_n^J$ and, when appropriate, estimating the additional jump risk parameter.

**Parametric estimation methods**

The parameters of the two risk-premium functions that are assumed in (6), namely $\eta_1$ and $\eta_2$, are estimated separately for each of seven forecast horizons. Each horizon defines a set of non-overlapping time periods. For one of these sets, at the end of period $s$ corresponding to day $n_s$, we can use numerical methods to evaluate the real-world density function $g_{P,s,T}(x|\eta_1,\eta_2,\Theta_{n_s})$ for the asset price $T$ years later, at the end of period $s+1$. The *ex ante* maximum likelihood estimates (MLEs) of $\eta_2$ and $\eta_2$ at time $t$ are given by maximizing the log-likelihood function of the observed asset prices $F_s$ at the ends of periods $s = 1,2,...,t$. Thus we maximize:

$$\log L(F_1,...,F_t|\eta_1,\eta_2) = \sum_{s=0}^{t-1} \log(g_{P,s,T}(F_{s+1}|\eta_1,\eta_2,\Theta_{n_s})). \quad (27)$$

In the same way, the parameters of the parametric calibration function (10), namely $j$ and $k$, are also estimated separately for each of several forecast horizons. The risk-neutral density $g_{Q,s,T}(x|\Theta_{n_s})$ and its c.d.f are used to evaluate the real-world density $g_{P,s,T}(x|j,k,\Theta_{n_s})$ given by (11). The *ex ante* MLEs of $j$ and $k$ at time $t$ are also given by maximizing the log-likelihood function of the available, observed asset prices, i.e.:

$$\log L(F_1,...,F_t|j,k) = \sum_{s=0}^{t-1} \log(g_{P,s,T}(F_{s+1}|j,k,\Theta_{n_s})). \quad (28)$$

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

$$
\log L(r_{\tau},...,r_t|\theta) = \sum_{s=\tau-1}^{t-1} \log(f(r_{s+1}|I_s, \theta)).
$$

(29)

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

$$
g_{ARCH,t}(x|I_t, \hat{\theta}_t) = f(r|I_t, \hat{\theta}_t)/x, \quad \text{with } r = \log(x/F_t).
$$

(30)

Similarly, the MLE of the mixture parameter $\alpha$, 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 $\alpha$. Then, at time $t$ we will know the observed values of the components of the mixture, for example we know $\tilde{g}_{p,s} = g_{p,s,T}(F_{s+1}|\hat{j}_s, \hat{k}_s, \Theta_n)$ and $\tilde{g}_{A,s} = g_{ARCH,s}(F_{s+1}|\hat{j}_s, \hat{\theta}_s)$ for times $0 \leq s < t$. The MLE of $\alpha$ at time $t$ is given by the number $\hat{\alpha}_t$ that maximizes

$$
\log L(F_1, F_2, ..., F_t|\alpha) = \sum_{s=0}^{t-1} \log(\alpha \tilde{g}_{p,s} + (1-\alpha)\tilde{g}_{A,s}).
$$

(31)

The ex ante mixture density for $F_{t+1}$ is then

$$
\hat{\alpha}_t g_{P,t,T}(x|\hat{j}_t, \hat{k}_t, \Theta_n) + (1-\hat{\alpha}_t) g_{ARCH,t}(x|I_t, \hat{\theta}_t).
$$

(32)

Non-parametric estimation methods

The non-parametric 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,

$$
u_{s+1} = G_{Q,s,T}(F_{s+1}|\Theta_n), \quad 0 \leq s \leq t-1.
$$

We assume these observations are i.i.d. with c.d.f. given by the calibration function $C_T(u)$. 

Let \( \phi(.) \) and \( \Phi(.) \) 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 \( \{y_1, y_2, \ldots, y_t\} \). We use a normal kernel, with bandwidth \( B \), and so obtain the kernel density and c.d.f. respectively as

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

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

The empirical calibration function is then

\[
\hat{C}_T(u) = \hat{H}_T(\Phi^{-1}(u)). \quad (34)
\]

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

\[
G_{P,T}(x) = \hat{C}_T(G_{Q,T}(x)). \quad (35)
\]

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

\[
g_{P,T}(x) = \frac{d}{dx} \hat{H}_T(y) = \frac{dy}{dx} \frac{d\hat{H}_T(y)}{dy} = \frac{du}{dy} \frac{dy}{dx} \frac{\hat{H}_T(y)}{\phi(y)} = \frac{g_{Q,T}(x)}{\phi(y)} \hat{H}_T(y). \quad (36)
\]

Finally, the non-parametric calibration density is

\[
\hat{c}_T(u) = \hat{h}_T(y)/\phi(y). \quad (37)
\]

References


Hong, Yongmiao, and Haitao Li, 2005, Nonparametric specification testing for continuous-time models with applications to term structure of interest rates, *Review of Financial Studies* 18, 37-84.


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. Moneyness is defined by the futures price $F$ divided by the strike price $K$.

<table>
<thead>
<tr>
<th>Moneyness /Maturity</th>
<th>$F/K$</th>
<th>Total number</th>
<th>Average options per day</th>
<th>Max number per day</th>
<th>Min number per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calls</td>
<td></td>
<td>171,383</td>
<td>45</td>
<td>157</td>
<td>8</td>
</tr>
<tr>
<td>Puts</td>
<td></td>
<td>263,717</td>
<td>70</td>
<td>173</td>
<td>8</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>435,100</td>
<td>115</td>
<td>255</td>
<td>29</td>
</tr>
<tr>
<td>Number of cross-sections</td>
<td></td>
<td>3.1</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Deep OTM put</td>
<td>&gt;1.10</td>
<td>10,800</td>
<td>89,779</td>
<td>39,879</td>
<td>140,458</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.48%)</td>
<td>(20.63%)</td>
<td>(9.17%)</td>
<td>(32.28%)</td>
</tr>
<tr>
<td>OTM put</td>
<td>1.03–1.10</td>
<td>8,743</td>
<td>52,427</td>
<td>23,964</td>
<td>85,134</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.01%)</td>
<td>(12.05%)</td>
<td>(5.51%)</td>
<td>(19.57%)</td>
</tr>
<tr>
<td>Near the money</td>
<td>0.97–1.03</td>
<td>7720</td>
<td>47,325</td>
<td>20,206</td>
<td>75,251</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.77%)</td>
<td>(10.88%)</td>
<td>(4.64%)</td>
<td>(17.30%)</td>
</tr>
<tr>
<td>OTM call</td>
<td>0.90–0.97</td>
<td>6,881</td>
<td>45,519</td>
<td>19,178</td>
<td>71,578</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.58%)</td>
<td>(10.46%)</td>
<td>(4.41%)</td>
<td>(16.45%)</td>
</tr>
<tr>
<td>Deep OTM call</td>
<td>&lt;0.90</td>
<td>2,483</td>
<td>42,253</td>
<td>17,943</td>
<td>62,679</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.57%)</td>
<td>(9.71%)</td>
<td>(4.12%)</td>
<td>(14.41%)</td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td>36,627</td>
<td>277,303</td>
<td>121,170</td>
<td>435,100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.42%)</td>
<td>(63.73%)</td>
<td>(27.85%)</td>
<td>(100%)</td>
</tr>
</tbody>
</table>
Estimates are summarized for the risk-neutral dynamics
\[ \frac{dF}{F} = \sqrt{V} dW_1 + (\mu^J - 1)dN - \lambda \mu_J dt \quad \text{and} \quad dV = \kappa (\theta - V) dt + \xi \sqrt{V} dW_2, \]
with \( dW_1 dW_2 = \rho dt \). The Poisson process \( N \) has intensity \( \lambda \), and is independent of the bivariate Wiener process \( (W_{1,t}, W_{2,t}) \). The jumps \( J_t \) in \( \log(F_t) \) are normally distributed with mean \( \mu_J \) and variance \( \sigma_J^2 \); the average size of the proportional jumps in \( F_t \) equals \( \bar{\mu}_J \).

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 \( \lambda = 0 \).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Median No jumps</th>
<th>Median With jumps</th>
<th>Mean No jumps</th>
<th>Mean With jumps</th>
<th>Standard deviation No jumps</th>
<th>Standard deviation With jumps</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sqrt{V_0} )</td>
<td>0.1787</td>
<td>0.1651</td>
<td>0.1898</td>
<td>0.1877</td>
<td>0.0741</td>
<td>0.0694</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>4.1528</td>
<td>3.0920</td>
<td>4.9292</td>
<td>3.8748</td>
<td>3.6598</td>
<td>3.2930</td>
</tr>
<tr>
<td>( \theta )</td>
<td>0.0452</td>
<td>0.0347</td>
<td>0.0505</td>
<td>0.0421</td>
<td>0.0273</td>
<td>0.0381</td>
</tr>
<tr>
<td>( \xi )</td>
<td>0.7925</td>
<td>0.6400</td>
<td>0.9296</td>
<td>0.6977</td>
<td>0.5160</td>
<td>0.3731</td>
</tr>
<tr>
<td>( \rho )</td>
<td>-0.6624</td>
<td>-0.6795</td>
<td>-0.6590</td>
<td>-0.6788</td>
<td>0.0875</td>
<td>0.1181</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>0.4746</td>
<td>0.6930</td>
<td>0.7147</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu_J )</td>
<td>-0.0630</td>
<td>-0.1019</td>
<td>0.1609</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma_J )</td>
<td>0.0385</td>
<td>0.0875</td>
<td>0.1059</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>0.0472</td>
<td>0.0327</td>
<td>0.1621</td>
<td>0.1072</td>
<td>0.3126</td>
<td>0.2052</td>
</tr>
</tbody>
</table>
### 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).

<table>
<thead>
<tr>
<th>Forecast horizon</th>
<th>Number of obs.</th>
<th>GJR</th>
<th>GJR-t</th>
<th>Intra</th>
<th>Intra-t</th>
<th>Risk-transformed Intra-t</th>
<th>Log Normal Q</th>
<th>Risk-transformed Lognormal Q</th>
<th>Heston Q</th>
<th>Risk-transformed Heston Q</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data source</td>
<td>Daily returns</td>
<td>Daily returns</td>
<td>Intraday returns</td>
<td>Intraday returns</td>
<td>Intraday returns</td>
<td>Intraday returns</td>
<td>Options</td>
<td>Options</td>
<td>Options</td>
</tr>
<tr>
<td>1 day</td>
<td>3520</td>
<td>-11951.2</td>
<td>91.4</td>
<td>56.4</td>
<td>135.9</td>
<td>134.0</td>
<td>140.8</td>
<td>27.0</td>
<td>73.5</td>
<td>100.9</td>
</tr>
<tr>
<td>1 week</td>
<td>711</td>
<td>-2961.9</td>
<td>13.5</td>
<td>16.1</td>
<td>34.8</td>
<td>36.7</td>
<td>38.4</td>
<td>17.0</td>
<td>32.8</td>
<td>36.5</td>
</tr>
<tr>
<td>2 weeks</td>
<td>351</td>
<td>-1574.0</td>
<td>10.4</td>
<td>16.2</td>
<td>18.5</td>
<td>16.6</td>
<td>18.0</td>
<td>13.6</td>
<td>27.8</td>
<td>26.4</td>
</tr>
<tr>
<td>4 weeks</td>
<td>176</td>
<td>-853.6</td>
<td>4.1</td>
<td>10.9</td>
<td>13.2</td>
<td>11.9</td>
<td>9.6</td>
<td>12.5</td>
<td>13.4</td>
<td>15.8</td>
</tr>
<tr>
<td>6 weeks</td>
<td>115</td>
<td>-596.9</td>
<td>5.7</td>
<td>14.0</td>
<td>16.4</td>
<td>13.1</td>
<td>17.6</td>
<td>16.0</td>
<td>17.1</td>
<td>16.9</td>
</tr>
<tr>
<td>8 weeks</td>
<td>86</td>
<td>-446.1</td>
<td>1.5</td>
<td>2.6</td>
<td>5.8</td>
<td>4.3</td>
<td>2.8</td>
<td>4.9</td>
<td>9.4</td>
<td>9.3</td>
</tr>
<tr>
<td>12 weeks</td>
<td>58</td>
<td>-310.2</td>
<td>5.2</td>
<td>3.7</td>
<td>7.4</td>
<td>5.1</td>
<td>5.7</td>
<td>5.6</td>
<td>6.9</td>
<td>6.8</td>
</tr>
</tbody>
</table>
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 $\alpha$ of the option-based density plus a fraction $1-\alpha$ of the Intra-$t$ density. $\alpha$ 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.

<table>
<thead>
<tr>
<th>Forecast Horizon</th>
<th>Intra-t only</th>
<th>Lognormal</th>
<th>Intra-t combined with</th>
<th>Heston</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Q Average $\alpha$</td>
<td>P1 Average $\alpha$</td>
<td>P2 Average $\alpha$</td>
</tr>
<tr>
<td>1 day</td>
<td>135.9</td>
<td>147.7 17%</td>
<td>150.8 35%</td>
<td>142.0 31%</td>
</tr>
<tr>
<td>1 week</td>
<td>34.8</td>
<td>33.5 21%</td>
<td>36.6 60%</td>
<td>38.3 60%</td>
</tr>
<tr>
<td>2 weeks</td>
<td>18.5</td>
<td>18.9 23%</td>
<td>27.1 83%</td>
<td>24.9 71%</td>
</tr>
<tr>
<td>4 weeks</td>
<td>13.2</td>
<td>14.1 11%</td>
<td>13.6 49%</td>
<td>15.1 49%</td>
</tr>
<tr>
<td>6 weeks</td>
<td>16.4</td>
<td>19.1 48%</td>
<td>16.7 76%</td>
<td>17.7 56%</td>
</tr>
<tr>
<td>8 weeks</td>
<td>5.8</td>
<td>6.1 33%</td>
<td>8.8 82%</td>
<td>8.0 60%</td>
</tr>
<tr>
<td>12 weeks</td>
<td>7.4</td>
<td>7.4 9%</td>
<td>9.5 72%</td>
<td>7.2 40%</td>
</tr>
</tbody>
</table>
Table 5

Results from the Kolmogorov-Smirnov test

The tabulated numbers are the $p$-values for the Kolmogorov-Smirnov test of the null hypotheses that the variables $u_t$ have a uniform distribution. The risk transformation $P_1$ refers to the parametric calibration transformation, $P_2$ to the nonparametric calibration transformation, and $P_3$ to the risk-premia transformation, respectively defined by (11), (36) and (6) and applied to risk-neutral densities denoted by $Q$. 

<table>
<thead>
<tr>
<th>Forecast horizon</th>
<th>Number of obs.</th>
<th>GJR</th>
<th>GJR-t</th>
<th>Intra</th>
<th>Intra-t</th>
<th>Risk-transformed Intra-t</th>
<th>Log Normal</th>
<th>Risk-transformed Lognormal</th>
<th>Heston</th>
<th>Risk-transformed Heston</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P1</td>
<td>P2</td>
<td>Q P1</td>
<td>P2</td>
<td>Q P1</td>
</tr>
<tr>
<td>1 day</td>
<td>3520</td>
<td>0.00%</td>
<td>0.75%</td>
<td>0.00%</td>
<td>15.19%</td>
<td>8.55%</td>
<td>87.42%</td>
<td>0.00%</td>
<td>0.06%</td>
<td>87.56% 0.32% 46.72% 0.01%</td>
</tr>
<tr>
<td>1 week</td>
<td>711</td>
<td>0.36%</td>
<td>25.09%</td>
<td>0.11%</td>
<td>12.73%</td>
<td>2.15%</td>
<td>39.95%</td>
<td>0.04%</td>
<td>1.14%</td>
<td>74.35% 0.60% 96.08% 78.80% 67.58%</td>
</tr>
<tr>
<td>2 weeks</td>
<td>351</td>
<td>13.36%</td>
<td>76.69%</td>
<td>0.73%</td>
<td>89.15%</td>
<td>13.11%</td>
<td>63.64%</td>
<td>0.03%</td>
<td>21.60%</td>
<td>67.63% 3.52% 92.71% 66.42% 78.16%</td>
</tr>
<tr>
<td>4 weeks</td>
<td>176</td>
<td>13.09%</td>
<td>82.55%</td>
<td>0.14%</td>
<td>64.28%</td>
<td>16.24%</td>
<td>63.78%</td>
<td>0.04%</td>
<td>0.67%</td>
<td>24.86% 7.75% 82.81% 49.13% 69.38%</td>
</tr>
<tr>
<td>6 weeks</td>
<td>115</td>
<td>94.00%</td>
<td>83.30%</td>
<td>4.97%</td>
<td>42.91%</td>
<td>3.81%</td>
<td>29.25%</td>
<td>1.57%</td>
<td>14.28%</td>
<td>18.18% 54.23% 84.58% 80.00% 41.38%</td>
</tr>
<tr>
<td>8 weeks</td>
<td>86</td>
<td>59.42%</td>
<td>72.82%</td>
<td>2.42%</td>
<td>77.81%</td>
<td>38.37%</td>
<td>88.08%</td>
<td>1.40%</td>
<td>48.79%</td>
<td>98.95% 9.90% 84.75% 99.45% 51.56%</td>
</tr>
<tr>
<td>12 weeks</td>
<td>58</td>
<td>85.19%</td>
<td>89.06%</td>
<td>0.31%</td>
<td>69.28%</td>
<td>35.18%</td>
<td>63.81%</td>
<td>0.18%</td>
<td>21.62%</td>
<td>61.28% 31.29% 90.03% 91.03% 85.18%</td>
</tr>
</tbody>
</table>
Table 6
Berkowitz test values and parameters

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. Stars indicate that the null is rejected at the 5% level, when LR3 > 7.81.

<table>
<thead>
<tr>
<th>Forecast horizon</th>
<th>GJR</th>
<th>GJR-t</th>
<th>Intra</th>
<th>Intra-t</th>
<th>Risk-transformed</th>
<th>Log</th>
<th>Risk-transformed</th>
<th>Heston</th>
<th>Risk-transformed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AR</td>
<td>P1</td>
<td>P2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.93</td>
<td>0.79</td>
<td>1.04</td>
<td>1.00</td>
<td>0.68</td>
<td>0.68</td>
<td>1.04</td>
<td>1.02</td>
<td>1.02</td>
</tr>
<tr>
<td>LR3</td>
<td>9.78*</td>
<td>19.84*</td>
<td>4.52</td>
<td>0.84</td>
<td>241.46*</td>
<td>3.72</td>
<td>1.02</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-0.11</td>
<td>-0.10</td>
<td>-0.11</td>
<td>-0.10</td>
<td>-0.10</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.10</td>
</tr>
<tr>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.78</td>
<td>0.83</td>
<td>1.04</td>
<td>0.97</td>
<td>0.63</td>
<td>0.63</td>
<td>1.02</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>LR3</td>
<td>26.57*</td>
<td>19.06*</td>
<td>46.13*</td>
<td>9.15*</td>
<td>61.63*</td>
<td>6.14</td>
<td>8.50*</td>
<td>5.52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>2 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.86</td>
<td>0.90</td>
<td>1.12</td>
<td>1.02</td>
<td>0.63</td>
<td>0.63</td>
<td>1.02</td>
<td>0.94</td>
<td>1.02</td>
</tr>
<tr>
<td>LR3</td>
<td>3.90</td>
<td>2.85</td>
<td>4.35</td>
<td>1.72</td>
<td>32.32*</td>
<td>3.41</td>
<td>1.39</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>4 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.77</td>
<td>0.92</td>
<td>1.10</td>
<td>0.97</td>
<td>0.63</td>
<td>0.63</td>
<td>0.90</td>
<td>0.74</td>
<td>0.92</td>
</tr>
<tr>
<td>LR3</td>
<td>6.04</td>
<td>1.74</td>
<td>1.16</td>
<td>0.36</td>
<td>19.61*</td>
<td>0.93</td>
<td>1.43</td>
<td>7.39</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-0.22</td>
<td>-0.17</td>
<td>-0.15</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.10</td>
<td>-0.09</td>
<td>-0.07</td>
</tr>
<tr>
<td>6 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.93</td>
<td>1.01</td>
<td>1.13</td>
<td>1.08</td>
<td>0.71</td>
<td>0.71</td>
<td>1.05</td>
<td>1.11</td>
<td>1.07</td>
</tr>
<tr>
<td>LR3</td>
<td>6.03</td>
<td>3.32</td>
<td>3.83</td>
<td>2.42</td>
<td>6.48</td>
<td>1.89</td>
<td>1.74</td>
<td>2.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>0.02</td>
<td>0.06</td>
<td>0.09</td>
<td>0.08</td>
<td>0.14</td>
<td>0.16</td>
<td>0.11</td>
<td>0.18</td>
</tr>
<tr>
<td>8 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.85</td>
<td>1.24</td>
<td>0.90</td>
<td>0.96</td>
<td>0.54</td>
<td>0.49</td>
<td>0.79</td>
<td>0.74</td>
<td>0.93</td>
</tr>
<tr>
<td>LR3</td>
<td>1.65</td>
<td>2.90</td>
<td>6.16</td>
<td>1.01</td>
<td>2.39</td>
<td>4.35</td>
<td>5.63</td>
<td>4.40</td>
<td>4.36</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-0.08</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.02</td>
<td>0.06</td>
<td>0.08</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>12 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>1.40</td>
<td>1.07</td>
<td>0.93</td>
<td>1.29</td>
<td>0.59</td>
<td>0.59</td>
<td>1.03</td>
<td>0.98</td>
<td>1.16</td>
</tr>
<tr>
<td>LR3</td>
<td>4.41</td>
<td>4.45</td>
<td>4.45</td>
<td>2.54</td>
<td>10.95*</td>
<td>1.06</td>
<td>0.82</td>
<td>4.92</td>
<td>0.90</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Forecast horizon</th>
<th>Q</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 day</td>
<td>60.9</td>
<td>-4.8</td>
<td>-8.5</td>
<td>-7.6</td>
</tr>
<tr>
<td>1 week</td>
<td>12.6</td>
<td>2.8</td>
<td>2.6</td>
<td>-1.4</td>
</tr>
<tr>
<td>2 weeks</td>
<td>4.4</td>
<td>0.0</td>
<td>-0.2</td>
<td>-2.9</td>
</tr>
<tr>
<td>4 weeks</td>
<td>1.1</td>
<td>0.2</td>
<td>-0.8</td>
<td>-3.5</td>
</tr>
<tr>
<td>6 weeks</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-3.7</td>
<td>-2.1</td>
</tr>
<tr>
<td>8 weeks</td>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
<td>-0.1</td>
</tr>
<tr>
<td>12 weeks</td>
<td>1.2</td>
<td>-1.0</td>
<td>-1.2</td>
<td>-1.3</td>
</tr>
</tbody>
</table>
Figure 1

Implied volatilities from the at-the-money, shortest-maturity options, and realized volatilities from intra-day returns

Implied volatilities are on the dark curve, realized volatilities are shown as light dots.
Figure 2a
One-day ahead density forecasts obtained from ARCH models on Dec 30th, 2004

Figure 2b
One-day ahead density forecasts obtained from lognormal densities on Dec 30th, 2004

Figure 2c
One-day ahead density forecasts obtained from Heston’s model on Dec 30th, 2004
Figure 3a
Four-week ahead density forecasts obtained from ARCH models on Nov 17th, 2004

Figure 3b
Four-week ahead density forecasts obtained from lognormal densities on Nov 17th, 2004

Figure 3c
Four-week ahead density forecasts obtained from Heston’s model on Nov 17th, 2004
Figure 4a
The function $\tilde{C}(u) - u$ for one-day forecasts obtained from lognormal densities and risk-transformations

Figure 4b
The function $\tilde{C}(u) - u$ for one-day forecasts obtained from Heston’s model and risk-transformations
Figure 5a
The estimated calibration densities $\hat{c}(u)$ for cumulative probabilities $u$
obtained from the one-day lognormal forecasts

![Figure 5a](image)

Figure 5b
The estimated calibration densities $\hat{c}(u)$ for cumulative probabilities $u$
obtained from the one-day Heston forecasts

![Figure 5b](image)
Figure 6a
The function $\tilde{C}(u) - u$ for four-week forecasts obtained from lognormal densities and risk-transformations

Figure 6b
The function $\tilde{C}(u) - u$ for four-week forecasts obtained from Heston's model and risk-transformations