Density forecast comparisons for stock prices, obtained from high-frequency returns and daily option prices

Rui Fan [a, *] Stephen J. Taylor [b] Matteo Sandri [b]

Swansea University Lancaster University Lancaster University

This version**:** March 13, 2017

Abstract

This paper presents the first comparison of the accuracy of density forecasts for stock prices. Six sets of forecasts are evaluated for DJIA stocks, across four forecast horizons. Two forecasts are risk-neutral densities implied by the Black-Scholes and Heston models. The third set are historical lognormal densities with dispersion determined by forecasts of realized variances obtained from 5-minute returns. Three further sets are defined by transforming risk-neutral and historical densities into real-world densities. The most accurate method applies the risk transformation to the Black-Scholes densities. This method outperforms all others for 87% of the comparisons made using the likelihood criterion.

The authors thank the editor, the referee and participants at the EFMA 2015 conference for their helpful advice and comments.

[*] Contact author. Tel.: +44 (0)1792 602671. Email: Rui.Fan@swansea.ac.uk (R. Fan).

[a] School of Management, Swansea University, Bay Campus, Fabian Way, Swansea, Wales, SA1 8EN.

[b] Department of Accounting and Finance, Lancaster University Management School, LA1 4YX, UK.

1. Introduction

Density forecasts for future asset prices are of importance to central bankers, risk managers and other decision makers for activities such as policy-making, risk management and derivatives pricing. They can also be used to assess market beliefs about economic and political events when derived from option prices.

Our contribution in this paper is to provide the first comparison of density forecasts for the future prices of individual stocks. We compare forecasts obtained from option prices and the pricing models of Black and Scholes (1973) and Heston (1993) with forecasts obtained from high-frequency historical returns and the regression model of Corsi (2009). We recommend to decision makers that the best predictive densities are provided by firstly using option prices, secondly using simple models to define risk-neutral densities and thirdly transforming the risk-neutral densities into real-world densities. Our conclusions differ from those in prior research for other assets.

We investigate density forecasts for the prices of seventeen constituents of the Dow Jones Industrial Average (DJIA) for the period from 2003 to 2012, for four horizons ranging from one day to one month. Previously, equity market comparisons have been made by Liu et al. (2007) for FTSE 100 futures, Shackleton et al. (2010) for S & P 500 futures and Yun (2014) for the S&P-500 index. Comparisons for other asset classes include Høg and Tsiaras (2011) for crude oil futures and Trujillo-Barrera et al. (2012) for lean hog futures, while related studies of Euribor interest rate futures are Gutiérrez and Vincent-Humphreys (2012), Vergote and Gutiérrez (2012) and Ivanova and Gutiérrez (2014).

As option prices reflect both historical and forward-looking information, forecasters are motivated to prefer forecasts derived from option market data. Evidence for superior predictions using option prices is widespread. Regarding volatility forecasts, Blair et al.

(2001), Jiang and Tian (2005), Giot and Laurent (2007) and Busch et al. (2011) all state that option forecasts are more informative and accurate than historical forecasts of index volatility even when the historical information set includes intraday, high-frequency returns.[1] Chang et al. (2012) show option-implied betas have significant predictive power compared with historical betas, while Kempf et al. (2015) show that option prices can be used to enhance portfolio optimization. Liu et al. (2007) provide the first evidence that option prices provide density forecasts that are superior to forecasts which rely solely on historical prices.

We show that options prices are more informative than intraday, high-frequency returns when constructing density forecasts for our selection of DJIA stocks, for all forecast horizons varying from one day to one month. This contrasts with the index results of Shackleton et al. (2010), which favor option-based methods for two-week and four-week horizons but high-frequency methods for forecasts one-day ahead. We note that option-based density forecasts generally outperform historical forecasts obtained from daily returns for a one-month horizon, according to Liu et al. (2007), Høg and Tsiaras (2011) and Trujillo-Barrera et al. (2012). Furthermore, Yun (2014) shows that "options & returns" models outperform "only returns" models for all horizons from one day to four weeks.

Many methods have been proposed to obtain risk-neutral densities from option prices. Almost all methods only provide distributions for horizons which match option expiry dates[2]. We instead estimate a stochastic process for the underlying asset price, by matching theoretical and market option prices for all expiry dates, to obtain densities for all horizons. We estimate the mean-reverting, square-root, stochastic volatility process of Heston (1993) which provides a closed form solution for option prices and a tractable density formula based on inverting characteristic functions. More complicated specifications including jumps (e.g.

---

1 Further comparisons are in Poon and Granger (2003), Martens and Zein (2004) and Taylor et al. (2010).

2 For example, applying methodologies such as explicit parametric distributions (Ritchey, 1990; Madan and Milne, 1994; Anagnou-Basioudis et al., 2005), discrete probabilities (Jackwerth and Rubinstein, 1996), a nonparametric kernel regression (Aït-Sahalia and Lo, 1998) and implied volatility splines (Bliss and Panigirtzoglou, 2002).

Duffie et al., 2000; Eraker, 2004) are not evaluated, because Shackleton et al. (2010) find that adding jumps does not significantly improve their forecasting results because their risk transformations systematically improve mis-specified risk-neutral densities.

We compare density forecasts derived from option prices using the Black-Scholes (1973) and Heston (1993) models with forecasts obtained from historical time series using the Corsi (2009) heterogeneous autoregressive model of realized variance (HAR-RV). However, the risk-neutral density is a suboptimal forecast of the future distribution of the asset price as there is no risk premium in the risk neutral world, while in reality investors are risk-averse. Hence we need to use economic models and/or econometric methods to transform risk-neutral densities into real-world[3] densities. Economic theory motivates pricing kernel transformations using power and/or exponential utility functions (Bakshi et al., 2003; Bliss and Panigirtzoglou, 2004; Liu et al., 2007) and the hyperbolic absolute risk aversion (HARA) function (Kang and Kim, 2006). Liu et al. (2007) use both utility and statistical calibration transformations, and they find that a statistical, parametric calibration gives a higher log-likelihood for observed outcomes than a utility transformation. Shackleton et al. (2010) compare parametric and nonparametric econometric transformations, obtaining the best diagnostic test results for the latter. The nonparametric method avoids making assumptions about the correct transformation and relies on learning from past outcomes how best to change measure from risk-neutral to real-world. Following Shackleton et al. (2010), Høg and Tsiaras (2011) and Ivanova and Gutiérrez (2014), we also transform risk-neutral densities into real-world densities using a nonparametric transformation. Some further parametric transformation results are provided in Yun (2014).

This paper is structured as follows. Section 2 describes the density forecasting methods,

[3] Like Liu et al. (2007), we use "real-world" rather than other alternative adjectives, such as "risk-adjusted", "statistical", "empirical", "physical", "true", "subjective" and "objective", etc., which are all used in the literature to indicate that the price distributions incorporate risk preferences.

namely the Black-Scholes (1973) and Heston (1993) models for densities inferred from option prices, the Corsi (2009) HAR-RV model for density forecasts obtained from historical high-frequency stock prices and the nonparametric transformation of Shackleton et al. (2010). It also includes the econometric methods used to obtain ex-ante parameters and evaluate density forecasts. Section 3 describes the DJIA stock and option prices data employed in the study. Section 4 focuses on the empirical analysis. Section 5 summarizes the findings and concludes.

2. Methodology

2.1 Option pricing with stochastic volatility

We extract risk-neutral densities from option prices and a specific pricing model. A realistic stochastic process for an individual stock price must incorporate a stochastic volatility component, whose increments are correlated with the price increments. We need to calculate an enormous number of theoretical option prices, so fast calculations are essential. The stochastic volatility process of Heston (1993) meets all our requirements as it has closed-form densities and option prices. Furthermore, the Heston pricing formula provides a generally satisfactory match to observed implied volatilities and it is a significant improvement on the Black-Scholes formula in empirical research (e.g. Bakshi et al., 1997; Lin et al., 2001; Dupoyet, 2006). More complicated affine jump-diffusion processes are described by Duffie et al. (2000) and these can fit implied volatilities even better. We do not consider these extensions, noting that Shackleton et al. (2010) obtained no forecasting advantages from including price jumps in their study.

The risk-neutral price dynamics for the stock price S, which incorporate the stochastic variance V, are defined by

$$\frac{dS}{S} = (r - q)dt + \sqrt{V}dW_1 \tag{1}$$

and

$$dV = \kappa(\theta - V)dt + \sigma\sqrt{V}dW_2, \tag{2}$$

where r is the risk-free interest rate and q is the dividend yield. We let ρ denote the correlation between the two Wiener processes W_1 and W_2, while θ is the level towards which the stochastic variance V reverts and κ denotes the rate of reversion of V towards θ. The volatility of volatility parameter σ controls the kurtosis of the returns.

Similar to the Black-Scholes formula, the Heston call price formula depends on two prices, the initial volatility, two discount factors and two probabilities:

$$C(S_0, V_0) = S_0 e^{-qT} P_1 - K e^{-rT} P_2. \tag{3}$$

Here S_0 is the current spot price, K is the strike price of an European option and T is the time to expiry. Each P_j in equation (3) is a conditional probability that the call option expires in-the-money. The term P_2 is for the risk-neutral measure Q, while P_1 is for a related measure Q^* having different drift rates. All probabilities and density functions are provided by integrals of functions of the characteristic function of $\log(S_T)$, conditional on the initial state values S_0 and V_0, see Heston (1993) or the textbook explanation in Taylor (2005, Section 14.6).

2.2 High-frequency HAR methods

The HAR-RV model of Corsi (2009) is an AR-type model for the realized variance which combines volatility components calculated over different time horizons. The model uses the multiperiod realized variance, which is the total of one-period measures denoted as

$$RV_{t,t+h} = RV_{t+1} + RV_{t+2} + \cdots + RV_{t+h}, h \geq 1. \tag{4}$$

By definition $RV_{t,t+1} \equiv RV_{t+1}$ and we use $h = 5$ and $h = 22$ to represent the weekly and monthly realized variance. Here the time period for volatility measures is from t to $t + h$, both counting trading days. In contrast, our options notation is a time period from 0 to T, both measured in years.

As the logarithmic daily realized variances are approximately unconditionally normally distributed, Andersen et al. (2007) predict the logarithm of the realized variance for the next h-day period by applying the regression specification:

$$\log\left(RV_{t,t+h}\right) = \beta_{0,h} + \beta_{D,h} \log\left(RV_{t-1,t}\right) + \beta_{W,h} \log\left(RV_{t-5,t}\right) + \beta_{M,h} \log\left(RV_{t-22,t}\right) + \varepsilon_{t,t+h}. \tag{5}$$

We also use this logarithmic specification. As noted by Pong et al. (2004), following Granger and Newbold (1976) an unbiased prediction of $RV_{t,t+h}$ is provided by the exponential of the forecast of $\log(RV_{t,t+h})$ multiplied by $\exp\left(\frac{1}{2}S^2(h)\right)$ with $S^2(h)$ an estimate of the variance of $\varepsilon_{t,t+h}$, here assuming $\log(RV_{t,t+h})$ is a Gaussian process. This is a standard assumption for realized variance, first shown to be appropriate for equities by Andersen et al. (2001).

2.3 Lognormal densities, from the Black-Scholes model and HAR-RV forecasts

In the Black-Scholes model, we assume the risk-neutral dynamics is a geometric Brownian motion:

$$dS/S = (r - q)dt + \sigma dW. \tag{6}$$

The risk-neutral distribution of $\log(S_T)$ is then normal:

$$\log(S_T) \sim N\left(\log(S_0) + (r - q)T - \frac{1}{2}\sigma^2 T, \sigma^2 T\right),$$

and

$$E^Q[S_T] = S_0 e^{(r-q)T} = F, \tag{7}$$

where F is the no-arbitrage, futures price for time T.

The risk-neutral density of S_T then depends on three parameters (F, σ, T) and is given by the lognormal density

$$\psi(x|F,\sigma,T) = \frac{1}{x\sigma\sqrt{2\pi T}} e^{-\frac{1}{2}\left(\frac{\log(x)-\left[\log(F)-\frac{1}{2}\sigma^2 T\right]}{\sigma\sqrt{T}}\right)^2}. \quad (8)$$

Similarly, a risk-neutral, lognormal density from the HAR-RV model is given by replacing $\sigma\sqrt{T}$ by a term $\sqrt{\widehat{RV}_{t,t+h}}$ to give:

$$\psi\left(x\middle|F,\widehat{RV}_{t,t+h}\right) = \frac{1}{x\sqrt{2\pi\widehat{RV}_{t,t+h}}} e^{-\frac{1}{2}\left(\frac{\log(x)-\left[\log(F)-\frac{1}{2}\widehat{RV}_{t,t+h}\right]}{\sqrt{\widehat{RV}_{t,t+h}}}\right)^2}. \quad (9)$$

The quantity $\widehat{RV}_{t,t+h}$ is calculated from (5) and the stated bias correction, with the horizon h (measured in trading days) matching the calendar time T (measured in years). The lognormal assumption for the distribution of S_T in (9) can again be motivated by the empirical evidence in Andersen et al. (2001).

2.4 Nonparametric transformations

The risk-neutral, Q-densities are not satisfactory specifications of the real-world densities. One reason is that the Q-variance obtained from option prices is usually higher than the real-world variance, because there is a negative volatility risk premium (Carr and Wu, 2009). Consequently there are fewer observations than predicted in the tails of the Q-densities. A second reason is that the equity risk premium is, by definition, absent from all the risk-neutral densities. Hence it is necessary to use some technique to transform risk-neutral densities into real-world densities. We reviewed parametric and nonparametric transformation methodologies in the introduction. We apply the nonparametric method of Shackleton et al. (2010) because firstly it makes less assumptions than all alternatives and secondly it relies on standard density estimation theory, originally presented by Silverman (1986). The method is also applied and described by Høg and Tsiaras (2011) and Ivanova and Gutiérrez (2014).

The nonparametric calibration method relies on learning from past outcomes how best to change measure from risk-neutral to real-world. It is well-known that the cumulative probabilities u of observed prices p are uniformly distributed for correctly specified densities (e.g. Diebold et al., 1998; Elliott and Timmermann, 2016). Observed deviations from uniformity can be exploited to transform risk-neutral densities, to obtain better descriptions of real-world outcomes.

The nonparametric transformation calibration function for a selected horizon h, and a selected risk-neutral method, is determined by a set of $n-h$ cumulative, risk-neutral probabilities

$$u_{s+h} = F_{Q,s,T}(p_{s+h}|\Theta_s), \qquad 1 \leq s \leq n-h, \tag{10}$$

with T (years) matching h (trading days). Here $F_{Q,s,T}$ is the cumulative distribution function (c.d.f.) of the stock price p_{s+h}, with Θ_s a vector of density parameters. The values of the variables u for the Heston model are obtained by fast and accurate numerical integration, while u for the HAR-RV and Black-Scholes models is simply a cumulative probability of the normal distribution.[4] We note that the terms u are often called probability integral transform (PIT) values in the forecasting literature.

Let $\varphi()$ and $\Phi()$ represent the density and the c.d.f. of the standard normal distribution. We transform the PIT values u_i, whose domain is from 0 to 1, to new variables $y_i = \Phi^{-1}(u_i)$, and then estimate a nonparametric kernel density from the set $\{y_{h+1}, y_{h+2}, \dots, y_n\}$. We use a Gaussian kernel with bandwidth B to obtain the kernel density and c.d.f.:

---

4 When calculating densities and variables u, we replace the risk-neutral expectation F in (8) and (9) by the synthetic futures price on day s for a future transaction at time $s+h$. We evaluate the options-on-futures version of the Heston model, for which q equals r and the initial asset price is the synthetic futures price.

$$\hat{h}_T(y) = \frac{1}{(n-h)B} \sum_{i=h+1}^{n} \varphi\left(\frac{y - y_i}{B}\right),$$

$$\hat{H}_T(y) = \frac{1}{n-h} \sum_{i=h+1}^{n} \Phi\left(\frac{y - y_i}{B}\right). \quad (11)$$

We apply the standard bandwidth formula given by $B = 0.9\sigma_y/(n-h)^{0.2}$, where σ_y is the standard deviation of the terms y_i. The empirical calibration function, which is the estimated real-world c.d.f. of the risk-neutral PIT values, is then

$$\hat{C}_T(u) = \hat{H}_T\left(\Phi^{-1}(u)\right) \quad (12)$$

and its derivative is the calibration density

$$\hat{c}_T(u) = \frac{d}{du}\hat{C}_T(u) = \frac{\hat{h}_T(y)}{\varphi(y)}, \quad (13)$$

with $y = \Phi^{-1}(u)$.

At time t, the risk-neutral c.d.f. of p_{t+h}, denoted $F_{Q,T}(x)$, is transformed into the real-world c.d.f. of p_{t+h}, denoted $F_{P,T}(x)$, by the equation:

$$F_{P,T}(x) = \hat{C}_T\left(F_{Q,T}(x)\right). \quad (14)$$

Calculus then shows that the transformation of the risk-neutral density into the real-world density, from $f_{Q,T}(x)$ to $f_{P,T}(x)$, is given by:

$$f_{P,T}(x) = \frac{f_{Q,T}(x)\hat{h}_T(y)}{\varphi(y)}, \quad (15)$$

with $y = \Phi^{-1}(F_{Q,T}(x))$.

2.5 Parameter estimation

The densities are all evaluated out-of-sample because all the parameter values are obtained ex ante, i.e. the values at time *t* are estimated from the information available at time *t*. For the HAR variances we estimate all parameters from regressions over five-year windows. For

Black-Scholes lognormal densities, we use the nearest-the-money, nearest-to-expiry, option implied volatility.

For the Heston model, we estimate the risk-neutral parameters of the asset price dynamics every day. On each day, we estimate the initial variance V_t, the rate of reversion κ_t, the unconditional expectation of stochastic variance θ_t, the volatility of volatility σ_t, and the correlation ρ_t between the two Wiener processes. Suppose we use N_t European, call[5] option contracts on day t, denoted by $i = 1, \dots, N_t$, and the market prices are $c_{t,i}$, for strike prices $K_{t,i}$, and expiry times $T_{t,i}$. We let $p_{t,i}$ denote a futures price for the asset, calculated for a synthetic futures contract which expires in $T_{t,i}$ years. Then we calibrate the five risk-neutral Heston parameters by minimizing the total squared errors in

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

with $c(.)$ the Heston solution for the European call option price given in (3).[6,7]

2.6 Econometric methods

Elliott and Timmermann (2016, Chapter 18) provide a recent survey of methods for evaluating density forecasts in economics research. We apply these methods to stock prices.

2.6.1 Ranking forecasts

There are several ways to rank density forecasts, and we will use the standard log-likelihood criterion previously employed by Bao et. al (2007), Liu et. al (2007) and Shackleton et al.

5 Some call prices are derived from put prices and put-call parity, as discussed in Section 3.2.

6 Christoffersen and Jacobs (2004) conclude that the squared pricing error is a “good general-purpose loss function in option valuation applications”. Christoffersen et al. (2010) also employed it in a study of S&P 500 dynamics.

7 We assume the same price dynamics and parameter values apply to the set of contemporaneous futures contracts. The Appendix explains why this assumption is reasonable.

(2010). For a given horizon h, assume method m gives densities $f_{m,t}(x)$ at times $i, \ldots, j$ for the asset price at times $i+h, \ldots, j+h$. Our goal is to find the method which maximizes the out-of-sample log-likelihood of observed asset prices, which for method m is given by

$$L_m = \sum_{t=i}^{j} \log\left(f_{m,t}(p_{t+h})\right). \tag{17}$$

To compare any two different methods we apply a version of the log-likelihood ratio test in Amisano and Giacomini (2007). The null hypothesis states that two different density forecasting methods m and n have equal expected log-likelihood. The test is based on the log-likelihood differences

$$d_t = \log\left(f_{m,t}(p_{t+h})\right) - \log\left(f_{n,t}(p_{t+h})\right), \qquad i \leq t \leq j. \tag{18}$$

Amisano and Giacomini (2007) in their application follow Diebold and Mariano (1995) and add the assumption that the differences are uncorrelated; thus they ignore all covariance terms when estimating the variance of the sample mean $\bar{d}$. Hence the basic AG test statistic is

$$t_{i,j} = \frac{\bar{d}}{s_d/\sqrt{(j-i+1)}} = \frac{L_m - L_n}{s_d\sqrt{(j-i+1)}}. \tag{19}$$

This statistic follows an asymptotic standard normal distribution, where $\bar{d}$ is the mean and s_d is the standard deviation of the terms d_t. When $h > 1$ the forecasts overlap and it is plausible to expect some autocorrelation in the differences up to lag $h-1$. A Newey-West adjustment should then be made, replacing the estimated variance of $\bar{d}$ by

$$\frac{s_d^2}{n}[1 + 2\omega_1\hat{\rho}_1 + \cdots + 2\omega_{h-1}\hat{\rho}_{h-1}] \tag{20}$$

where the sample autocorrelations are $\hat{\rho}_\tau = \mathrm{cor}(d_t, d_{t+\tau})$. We use the standard set of weights given by $\omega_\tau = (h-\tau)/h$, $1 \leq \tau \leq h-1$.

2.6.2 Diagnostic tests

Appropriate diagnostic tests for density forecasts use PIT values, defined for a method m by the cumulative probabilities $u_{t+h} = F_{m,t}(p_{t+h})$. These PIT values are independent and uniformly distributed for non-overlapping, correctly specified densities.

The Kolmogorov-Smirnov (KS) test assesses uniformity by using the maximum difference between the empirical and theoretical cumulative functions. For forecasts made at times $i \leq t \leq j$, the sample c.d.f. of $\{u_{i+h}, \dots, u_{j+h}\}$, evaluated at u, is the proportion of values less than or equal to u, i.e.

$$\tilde{C}(u) = \frac{1}{j - i + 1} \sum_{t=i+h}^{j+h} S(u - u_t) \tag{21}$$

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

$$KS = \sup_{0 \leq u \leq 1} \left|\tilde{C}(u) - u\right|. \tag{22}$$

Although widely applied, we need to be cautious when interpreting KS statistics, as the KS test checks for uniformity under the i.i.d. assumption rather than testing i.i.d. and uniformity jointly.

Berkowitz (2001) proposed the BK test, which has the advantage that it tests independence and uniformity jointly; applications include Clements and Smith (2000), Clements (2004), Guidolin and Timmermann (2005), Shackleton et al. (2010) and Høg and Tsiaras (2011).

The BK method transforms the observations u_{t+h} to new variables $y_{t+h} = \Phi^{-1}(u_{t+h})$. The null hypothesis of the original test is that the values of y are i.i.d. and follow a standard normal distribution, against the alternative hypothesis that y is a stationary, Gaussian, AR(1) process with no restrictions on the mean, variance and autoregressive parameters. However, when $h > 1$ we have overlapping forecasts and must therefore expect correlated forecast errors, specifically $\text{cor}(y_t, y_{t+\tau}) > 0$ for $1 \leq \tau \leq h - 1$. To implement the test we therefore estimate

$$y_t - \mu = \rho(y_{t-h} - \mu) + \varepsilon_t. \tag{23}$$

Then the null hypothesis is $\mu = 0$, $\rho = 0$, and $\sigma^2 = \text{var}(\varepsilon_t) = 1$. The log-likelihood ratio test statistic (LR3) is

$$LR_3 = 2(L_1 - L_0) = 2\big(L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho}) - L(0,1,0)\big). \tag{24}$$

Here hats denote maximum-likelihood values, L_0 and L_1 are the maximum log-likelihoods for the null and alternative hypotheses, and the test statistic has an asymptotic χ_3^2 distribution. One disadvantage of the BK test is that models cannot be easily compared if they are all accepted or rejected. The AG test, which we discussed before, compares the log-likelihoods between models and may resolve this problem.

3. Data

3.1 Option data

We investigate a majority of the Dow Jones Industrial Average (DJIA) stocks for 10 years from 1st January 2003 to 31st December 2012. The data preparation and parameter estimation tasks are both time consuming[8]. Consequently we only report results for the 17 stocks which have straightforward and complete data. We find the 17 stocks are sufficient to determine the most accurate density forecasting method, because the comparisons reported in Section 4.4 show the same method clearly outperforms all others regardless of the forecast horizon chosen. Table 1 lists the stocks studied, which were all DJIA constituents at the end of our sample period.

The option data are obtained from Ivy DB OptionMetrics, which includes price information for all U.S. listed equity options, based on daily closing quotes at the CBOE. The

8 The estimation of the parameters for the Heston model is particularly time demanding.

OptionMetrics database also includes information about end-of-day security prices and zero-coupon interest rate curves. The security price file provides the closing price for each security on each day from CRSP. We calculate the interest rate corresponding to each option's expiry by linear interpolation of the two closest zero-coupon rates supplied by Ivy DB OptionMetrics.

3.2 Option prices

To filter option price records, we follow the criteria of Carr and Wu (2003, 2009 and 2010) and Huang and Wu (2004). We delete an option record when the bid price is zero, negative or more than the ask price. We eliminate all data for options which have maturity either less than eight calendar days or more than one year.

All the equity options are American. OptionMetrics provides implied volatilities, calculated from binomial trees which incorporate dividends and permit early exercise. We use equivalent European option prices defined by assuming the European and American implied volatilities are equal. This method assumes the early exercise premium can be obtained from constant volatility pricing models. The assumption is particularly reasonable for out-of-the-money options which have small early exercise premia.

European call and put prices for the same strike and maturity theoretically contain the same information. Either the call option or the put option will be out-of-the-money (OTM), or under rare circumstances both are at-the-money (ATM). Options are ATM when the strike price equals the stock price $(S = K)$, calls are OTM when $S < K$ and puts are OTM when $S > K$. We choose to only use the information given by the prices of OTM and ATM options, because in-the-money options are less liquid and have higher early exercise premia. We use put-call parity to obtain equivalent European call prices from the European OTM put prices.

The Black-Scholes lognormal densities are defined by using the implied volatility of the nearest-the-money, nearest-to-expiry contract, for which $|S - K|$ is nearer zero than for all other contemporaneous strikes.

3.3 IBM example

We use IBM to illustrate our data and results; its market capitalization is near the median and average across all DJIA stocks during the sample period. A total of 109,111 IBM option prices are investigated in our sample period. The average number of option prices used per day is 44, consisting of 19 OTM calls and 25 OTM puts. Table 2 summarizes the quantity, moneyness and maturity of the option contracts.

3.4 Futures prices

We calculate synthetic futures prices, for futures which have the same expiry dates as the options, from the usual no-arbitrage equation. Let D denote the present value of all the dividends expected until the option expiry time T. The no-arbitrage relationship between the current spot price S and the futures price F is then

$$F = e^{rT}(S - D). \quad (25)$$

3.5 High-frequency stock prices

We use the transaction prices of DJIA 30 Index stocks for ten years during the period between 1st January 1998 and 31st December 2012, obtained from pricedata.com. The prices provided are the last prices in one-minute intervals. After an inspection of the high-frequency data, we find a number of problematic days which do not have complete trading records. We set the price equal to that for the previous minute when there is a missing record, and we delete a day when there are more than 40 consecutive missing prices. The days deleted are usually close to holidays such as New Year's Day, Easter, Independence Day, Thanksgiving Day and

Christmas.

Between 2003 and 2012, 17 days are deleted because of missing high-frequency prices and these days usually only have prices for half a day. There are also 8 days with unsatisfactory option price data. All 25 days are deleted from the high-frequency and option files leaving a sample of 2488 days for each firm for the ten-year period ending on 31st December 2012.

The stocks are traded for six-and-a-half-hours, from 9:30 EST to 16:00 EST. We calculate realized variances from 5-minute returns. This popular method is motivated by Bandi and Russell (2006) who show that the 5-minute frequency can provide a satisfactory trade-off between maximizing the accuracy of volatility estimates and minimizing the bias from microstructure effects. Recently, Liu et al. (2015) compare hundreds of estimators of asset price variation for many asset classes and find that it is difficult to outperform 5-minute realized variance, particularly when forecasts of future variation are compared.

As usual, returns are changes in log prices. We have 77 5-minute intraday returns for each day after deleting the data in the first five minutes to avoid any opening effects. The realized variance for day t is the sum of the squares of the 5-minute returns $r_{t,i}$:

$$RV_t = \sum_{i=1}^{77} r_{t,i}^2. \quad (26)$$

However, this calculation of realized variance is downward biased as a measurement of close-to-close volatility over a 24-hour period. This is inevitable because we only use the information during the trading period, so the variation overnight (from close-to-open) is excluded. We thus need to multiply forecasts from the HAR-RV model by a scaling factor. The denominator of the scaling factor is the sum of the squares of the 5-minute returns representing the open market period, while the numerator is the sum of the squares of the daily returns representing open and closed market periods. We use a rolling window for the scaling factor, hence if we forecast the future realized variance on day t, then we use the information about returns up to and including day t to calculate

$$\widehat{RV}_{t,t+h}\left(\frac{\sum_{i=1}^{t} r_t^2}{\sum_{i=1}^{t}\sum_{j=1}^{77} r_{t,j}^2}\right). \tag{27}$$

This quantity replaces $\widehat{RV}_{t,t+h}$ in (9) when the high-frequency, lognormal densities are evaluated.

4. Empirical results

4.1 Heston risk-neutral parameters

Table 3 shows the summary statistics for risk-neutral parameters, obtained for each day in our sample period, firstly for IBM and secondly across all stocks. The risk-neutral parameters minimize the mean squared error (MSE) of option prices on each day.

For IBM, our median estimate of the stochastic variance θ is 0.3457, equivalent to an annualized volatility level of 58.80%. The mean estimate of the rate of reversion κ is 1.6861, for which the half-life parameter of the variance process is then about 5 months. The median estimate of the volatility of volatility parameter σ, which controls the kurtosis of returns, is 0.8617. Also the median estimate of the correlation ρ is -0.6652, consistent with estimates in the literature.

4.2 Examples of risk-neutral density forecasts

Illustrative one-day ahead Heston, Black-Scholes and HAR densities for IBM, calculated at the beginning of the option dataset on January 2nd 2003, are shown in Figure 1. The Heston density is negatively skewed while the lognormal, Black-Scholes density is slightly positively skewed. The lognormal HAR density is seen to have less variance than the Heston and

Black-Scholes densities. The one-month ahead densities for IBM calculated on the same day are shown in Figure 2 and they display similar properties.

4.3 Examples of cumulative probabilities and nonparametric transformations

From the one-day ahead risk-neutral cumulative distribution functions and the stock prices p_{t+1}, we find the observed risk-neutral probabilities $u_{t+1} = F_{Q,t}(p_{t+1})$ are not consistent with uniform probabilities. This is expected because risk-neutral distributions ignore the risk premia incorporated into real-world distributions.

The sample cumulative probabilities $\tilde{C}(u)$ are calculated using (21). The deviations between the sample c.d.f. and a uniform c.d.f., namely $\tilde{C}(u) - u$, are plotted in Figure 3 for IBM from 2003 to 2012, for the three sets of risk-neutral, one-day-ahead forecasts. For the Heston model we can observe from the figure that there are relatively few observations u close to either zero or one; only 5.1% of the variables u are below 0.1 and only 7.3% of them are above 0.9. The KS test statistic is the maximum value of $|\tilde{C}(u) - u|$, which is equal to 7.1%, hence the null hypothesis of a uniform distribution is rejected at the 0.01% significance level. The shape of the Heston curve may be explained by the fact that the historical volatility is lower than the risk-neutral volatility, hence the risk-neutral probabilities of large price changes exceed the real-world probabilities. The corresponding plots for IBM for one-day-ahead forecasts obtained from Black-Scholes and HAR lognormal densities are also shown in Figure 3.

The nonparametric estimate of the density of the risk-neutral probability u_{t+1} is given by (13). This calibration density $\hat{c}(u)$ multiplies the next-day, risk-neutral density $f_{Q,t}(x)$ to produce the next-day, real-world density $f_{P,t}(x)$, with $u = F_{Q,t}(x)$, from (11), (13) and (15).

The full-period calibration densities $\hat{c}(u)$, for one-day ahead Heston, Black-Scholes and

HAR lognormal forecasts are shown in Figure 4; these densities use the values of u for all 10 years from 2003 to 2012. It is seen that the option-based calibration densities are more than 1 when u is between 0.25 and 0.75, while they are less than 1 in the left and right tails. The real-world densities then have higher peaks and thinner tails than the risk-neutral densities and consequently the transformation from risk-neutral to real-world reduces the variance. The high-frequency HAR calibration density is much flatter and its main effect is to adjust the probabilities of extreme values.

The purpose of the calibration transformation is to create real-world densities which have uniformly distributed observed probabilities u_{t+1}. The differences $\tilde{C}(u) - u$ after applying the nonparametric calibration method for one-day ahead forecasts from Heston, Black-Scholes and HAR lognormal densities are also shown in Figure 3. The differences for these real-world densities are near zero for all values of u and therefore their cumulative probabilities are almost uniform, unlike the results for the risk-neutral densities. Comparable figures and results are obtained for longer horizon density forecasts.

4.4 Log-likelihood comparisons

Table 4 presents the out-of-sample, log-likelihoods for IBM, and the averages across all seventeen stocks from 2003 to 2012, for our three risk-neutral and three real-world forecasting methods. The density forecasts are for four horizons, namely one day, one week (5 trading days), two weeks (10) and one month (22).[9] Overlapping forecasts are evaluated for horizons exceeding one day. The log-likelihood of the untransformed HAR model is selected to define the benchmark level. The log-likelihoods of the other five density forecasting methods in excess of the benchmark level are summarized in the table.

For IBM stock, the transformed real-world densities derived from the lognormal

[9] For a horizon h trading days, we set $T = h/252$ to calculate option implied densities.

Black-Scholes model give the highest log-likelihoods for all four horizons ranging from one day to one month; the Black-Scholes method is also best for the risk-neutral densities for three horizons. The HAR model and the Heston model give similar log-likelihoods for all four horizons after applying transformations. The log-likelihoods after nonparametric transformations are higher than those under the risk-neutral measure for all methods and horizons, and the differences range from 66.3 to 192.8.

Similarly, for the averages across seventeen stocks, the lognormal Black-Scholes model gives the highest log-likelihoods for all four horizons, both for untransformed risk-neutral and transformed real-world densities. The HAR model produces higher average log-likelihoods than the Heston model for almost all horizons both before and after applying transformations, with the exception of the risk-neutral densities for the one week horizon. The log-likelihoods after applying the nonparametric transformation are always higher than those under the risk-neutral measure for all stocks, methods and horizons, and the average differences vary between 85.8 and 202.0.

Table 5 lists the number of times that each method has the highest log-likelihood for the selected forecast horizon across seventeen stocks. For transformed real-world densities, the lognormal Black-Scholes model gives the highest log-likelihoods for 59 out of 68 combinations from seventeen stocks and four horizons. The lognormal Black-Scholes model also gives the highest log-likelihoods 51 times for untransformed risk-neutral densities. The HAR model and the Heston model give the highest log-likelihoods for a similar number of times for risk-neutral densities, while the HAR model obtains the highest log-likelihoods more often than the Heston model for transformed real-world densities.

Table 6 summarizes significant differences between methods using the Amisano and Giacomini (AG) test. It shows the number of times that the row method provides statistically better forecasts than the column method at the 5% significance level for all seventeen stocks when the Newey-West adjustment is made to the estimated variance of $\bar{d}$ and 20 autocorrelations are used. For all four horizons, the nonparametric transformation of the Black-Scholes lognormal method has the largest number of times that it is statistically better than the other five density forecasting methods; furthermore, no method is ever statistically better than this top-ranked method. The number of times that each method is statistically better than the remaining methods decreases as the forecast horizon increases from one day to one week, two weeks and one month. The Newey-West adjustments are important and necessary for the AG test when the horizon is more than one day; without the adjustment for autocorrelation among the likelihood differences many more test results would appear to be significant. The nonparametric methods are more often significantly better than the parametric methods. At the one-day horizon, there are $17 \times 9 = 153$ possible comparisons between P and Q distributions of which 124 have P significantly better than Q.

Table 7 summarizes the AG test statistics for IBM. At the one day horizon, two of the AG test statistics are insignificant at the 5% level when the best method, transformed Black-Scholes lognormal, is compared to the five alternatives; the AG test statistics equal 0.33 and 0.99 for tests against transformed HAR and transformed Heston methods. The AG test results show that the best method for the one week horizon is significantly better than one of the remaining five methods at the 5% level, and the best method is statistically better than two methods at the 5% level for the two weeks horizon, while the best method has no significant differences at the longest, one month, horizon.

4.5 Diagnostic tests

The KS statistic tests if the densities are correctly specified under the i.i.d. assumption. Table 8 summarizes the p-values for the KS test for the six density forecasting methods and the four horizons for IBM. All the risk-neutral measure p-values are less than 0.5% and hence reject the null hypothesis at the 0.5% significance level, which can be explained by mis-specified risk-neutral densities which have higher variance than real-world densities. The untransformed HAR densities are also mis-specified, as they are conditionally normal. All sets of densities obtained by nonparametric transformations have satisfactory p-values greater than 50%.

Table 9 gives the number of times that the null hypothesis is rejected at the 5% significance level for the KS test across the seventeen stocks. All the nonparametric transformations pass the KS test while the null hypothesis is rejected for almost all risk neutral and untransformed cases when the significance level is 5%.

The Berkowitz LR3 statistic, for horizon equal to h days, tests the null hypothesis that the variables $y_i = \Phi^{-1}(u_i)$ follow a standard normal distribution and are independent of y_{i-h}. Table 10 presents the LR3 test statistic, and the estimates of the variance and AR parameters for the six density forecasting methods and the four horizons for IBM.

For IBM stock, the MLEs of the autoregressive parameters are between -0.01 and 0.01 for the one-day horizon, hence there is no significant evidence of time-series dependence. However, the MLEs for the one-week horizon range between -0.04 and -0.08, thus four of them reject the null hypothesis that the autoregressive parameter is zero at the 5% significance level. The longer two-weeks and one-month horizons also provide no evidence of dependent observations. The MLEs of the variance parameter are near one for correctly specified densities. The low estimates for one-day Black-Scholes and Heston forecasts under the Q measure can be explained by the fact that the risk-neutral standard deviations are on average higher than the historical standard deviations.

The LR3 test statistic is significant at the 5% level when it exceeds 7.81. Table 10 indicates that the null hypothesis is rejected for all IBM risk-neutral forecasts and all one-week forecasts. The null hypothesis is accepted for all real-world forecasts for one day, two-weeks and one-month horizons. The significant values of the LR3 test statistic are attributed to the negative estimates of the AR parameter for the one-week horizon and the mis-specified risk-neutral densities which have higher variances than the real-world levels.

Table 11 shows the number of times that the LR3 test rejects the null hypothesis at the 5% significance level for all seventeen stocks. The majority of the transformed distributions pass the LR3 test while the null hypothesis is rejected for almost all risk neutral specifications at the 5% significance level. The number of times that the null is rejected at the 5% level are similar across different horizons.

5. Conclusions

We compare density forecasts for the prices of DJIA stocks, obtained from 5-minute high-frequency returns and daily option prices by using Heston, lognormal Black-Scholes, lognormal HAR-RV and transformed densities. Our comparison criterion is the out-of-sample, log-likelihood of observed stock prices. For the 68 combinations from 17 stocks for 4 forecast horizons, the transformed lognormal Black-Scholes model gives the highest log-likelihoods for 59 combinations during the 10-year period from 2003 to 2012. The HAR-RV model and the Heston model have similar forecast accuracy for different horizons, either before or after applying a transformation which enhances the densities.

Our methodology follows that of Shackleton et al. (2010) but obtains some conclusions for DJIA stocks which are different to those presented in their study of the S&P 500 index. Firstly, we find that density forecasts obtained from option prices outperform high-frequency forecasts for all forecast horizons. This contrasts with high-frequency forecasts being superior for the S&P 500 index at the one-day horizon, although they are inferior to option forecasts at horizons of two or more weeks. Secondly, we find that simple option methods relying on

Black-Scholes implied volatilities outperform sophisticated methods which apply more advanced option pricing formulae. This contrasts with superior index forecasts from the Heston formula. Options are traded for many index strikes during a single day but for far fewer strikes for individual firms. We attribute the relatively unsatisfactory performance of the Heston model for our firms to the lower liquidity of their out-of-the-money options.

We use a nonparametric transformation to transform the risk-neutral densities into real-world densities. The log-likelihoods after the nonparametric transformation are always higher than those under the risk-neutral measure, for all methods and horizons. The nonparametric transformation also gives better diagnostic test results. Hence central banks, risk managers and other decision makers should not merely look at risk-neutral densities, but should also obtain more accurate predictions by using risk transformations applied to risk-neutral densities.

Appendix. Assumptions about prices, dividends and options

There are no jumps in the Heston dynamics, (1) and (2), so they are not applicable to any stock which pays dividends. We apply the Heston dynamics instead to synthetic futures prices which do not jump when dividends are paid. We need to assume all futures prices have the same dynamics, by which we mean that (1) and (2) apply to all contemporaneous futures prices with identical parameter values for all contracts. We can use the same dynamics for all futures by making simple dividend assumptions, outlined below; this is easy for continuous dividends but more complicated for discrete dividends.

We assume futures and options contracts expire at time T_1, and there is a dividend at time τ_1 between times 0 and T_1. The second expiry time for futures and options is T_2 and there is another dividend at τ_2 between times T_1 and T_2. We denote the futures price at t

for delivery at T to be $F_{t,T}$.

Our discussion refers to multiplicative dividend factors a_1, a_2, ..., which do not need to be calculated. We assume, at time t before time τ_i, that the expected dividends are

$$E_t[\text{dividend at } \tau_i] = a_1 e^{r(\tau_1 - t)} S_t \qquad i = 1, \qquad t < \tau_1,$$

$$= a_2(1 - a_1) e^{r(\tau_2 - t)} S_t \qquad i = 2, \qquad t < \tau_2,$$

$$= a_3(1 - a_1)(1 - a_2) e^{r(\tau_3 - t)} S_t \qquad i = 3, \qquad t < \tau_3,$$

etc. We assume futures prices are set by no-arbitrage conditions, so

$$F_{t,T} = e^{r(T-t)}[S_t - \text{PV(expected dividends from } t \text{ to } T)].$$

Then for the first contract

$$F_{t,T_1} = e^{r(T_1 - t)}\left[S_t - e^{-r(\tau_1 - t)} a_1 e^{r(\tau_1 - t)} S_t\right] = (1 - a_1) e^{r(T_1 - t)} S_t \qquad 0 \le t < \tau_1,$$

$$= e^{r(T_1 - t)} S_t \qquad \tau_1 \le t \le T_1,$$

so that

$$\log\left(F_{t,T_1}/S_t\right) = \log(1 - a_1) + r(T_1 - t) \qquad 0 \le t < \tau_1,$$

$$= r(T_1 - t) \qquad \tau_1 \le t \le T_1.$$

Thus

$$d\left(\log F_{t,T_1}\right) = d(\log S_t) - r dt, \qquad t \ne \tau_1.$$

Also S_t jumps down by $a_1 S_{\tau_1}$ at time $t = \tau_1$, but F_{t,T_1} does not jump at $t = \tau_1$.

Similarly, for the second contract

$$F_{t,T_2} = e^{r(T_2 - t)}\left[S_t - e^{-r(\tau_1 - t)} a_1 e^{r(\tau_1 - t)} S_t - e^{-r(\tau_2 - t)} a_2 (1 - a_1) e^{r(\tau_2 - t)} S_t\right]$$

$$= e^{r(T_2 - t)}(1 - a_1)(1 - a_2) S_t \qquad 0 \le t < \tau_1,$$

$$= e^{r(T_2 - t)}(1 - a_2) S_t \qquad \tau_1 \le t < \tau_2,$$

$$= e^{r(T_2 - t)} S_t \qquad \tau_2 \le t \le T_2.$$

Hence we have

$$d\left(\log F_{t,T_2}\right) = d(\log S_t) - r dt \qquad t \ne \tau_1, \tau_2,$$

$$= d\left(\log F_{t,T_1}\right) \qquad 0 \le t \le T_1.$$

And we also have

$$\frac{F_{t,T_2}}{F_{t,T_1}} = e^{r(T_2 - T_1)}(1 - a_2) \qquad 0 \le t \le T_1.$$

We see that the synthetic futures contracts have the same dynamics.

We estimate the Heston parameters from the prices $c_{i,j}$ of European options which expire at T_1, T_2, …, T_N, and have strike prices $K_{i,j}$, with $1 \le i \le N$ and $1 \le j \le n_i$. Our target is to estimate the Heston parameters θ as:

$$\hat{\theta} = \arg\min_{\theta} \sum_i \sum_j \left[c_{i,j} - c_{Heston}\left(F_{0,T_i}, T_i, K_{i,j}, r, \theta\right)\right]^2.$$

At time 0 and for any future time τ, we can obtain the density of $S_\tau = F_{\tau,\tau}$ by evaluating the Heston density with initial price $F_{0,\tau}$ and parameters $\hat{\theta}$.

References

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

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

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

Andersen, T.G., T. Bollerslev and F.X. Diebold, 2007. Roughing it up: including jump components in measuring, modeling and forecasting asset return volatility. *Review of Economics and Statistics* 89, 701-720.

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

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

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

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

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

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

Black, F. and M. Scholes, 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637-654.

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

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

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

Busch, T., B.J. Christensen and M.Ø. Nielsen, 2011. The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics* 160, 48–57.

Carr, P. and L. Wu, 2003. Finite moment log stable process and option pricing. *Journal of Finance* 58, 753-777.

Carr, P. and L. Wu, 2009. Variance risk premiums. *Review of Financial Studies* 22, 1311-1341.

Carr, P. and L. Wu, 2010. Stock options and credit default swaps: a joint framework for valuation and estimation. *Journal of Financial Econometrics* 8, 409-449.

Chang, B.-Y., P. Christoffersen, K. Jacobs and G. Vainberg, 2012. Option-implied measures of equity risk. *Review of Finance* 16, 385-428.

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

Christoffersen, P., K. Jacobs and K. Mimouni, 2010. Models for S&P 500 dynamics: evidence from realized volatility, daily returns and option prices. *Review of Financial Studies* 23, 3141-3189.

Clements, M.P., 2004. Evaluating the Bank of England density forecasts of inflation. *Economic Journal* 114, 855-877.

Clements, M.P. and J. Smith, 2000. Evaluating density forecasts of linear and non-linear models: applications to output growth and unemployment. *Journal of Forecasting* 19, 255-276.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174-196.

Diebold, F.X., T.A. Gunther and A.S. Tay, 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863-883.

Diebold, F.X. and R.S. Mariano, 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253-263.

Duffie, D., J. Pan and K.J. Singleton, 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343-1476.

Dupoyet, S., 2006. Information content of cross-sectional option prices: a comparison of alternative currency option pricing models. *Journal of Futures Markets* 26, 33-59.

Elliott, G. and A. Timmermann, 2016. Economic Forecasting. Princeton University Press. Princeton.

Eraker, B., 2004. Do equity prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance* 59, 1367-1403.

Giot, P. and S. Laurent, 2007. The information content of implied volatility in light of the jump/continuous decomposition of realized volatility. *Journal of Futures Markets* 27, 337-359.

Granger, C.W.J. and P. Newbold, 1976. Forecasting Economic Time Series. Academic Press, New York.

Guidolin, M. and A. Timmermann, 2005. Economic implication of bull and bear regimes in UK stock and bond returns. *Economic Journal* 115, 111-143.

Gutiérrez J.M.P. and R. Vincent-Humphreys, 2012. A quantitative mirror on the Euribor market using implied probability density functions. *Eurasian Economic Review* 2, 1-31.

Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6,327– 343.

Høg, E. and L. Tsiaras, 2011. Density forecasts of crude-oil prices using option implied and ARCH-type models. *Journal of Futures Markets* 31, 727-754.

Huang, J. and L. Wu, 2004. Specification analysis of option pricing models based on time-changed Levy processes. *Journal of Finance* 59, 1405-1439.

Ivanova, V. and J.M.P. Gutiérrez, 2014. Interest rate forecasts, state price densities and risk premium from Euribor options. *Journal of Banking and Finance* 48, 210-223.

Jackwerth, J.C. and M. Rubinstein, 1996. Recovering probability distributions from option prices. *Journal of Finance* 51, 1611-1631.

Jiang, G.J. and Y.S. Tian, 2005. The model-free implied volatility and its information content. *Review of Financial Studies* 13, 433-451.

Kang, B.J. and T.S. Kim, 2006. Option-implied risk preferences: an extension to wider classes of utility functions. *Journal of Financial Markets* 9, 180-198.

Kempf, A., O. Korn and S. Sassning, 2015. Portfolio optimization using forward-looking information. *Review of Finance* 19, 467-490.

Lin, Y-N., N. Strong and X. Xu, 2001. Pricing FTSE-100 index options under stochastic volatility. *Journal of Futures Markets* 21, 197-211.

Liu, L.Y., A.J. Patton and K. Sheppard, 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics* 187, 293-311.

Liu, X., M.B. Shackleton, S.J. Taylor and X. Xu, 2007. Closed-form transformations from risk-neutral to real-world densities. *Journal of Banking and Finance* 31, 1501-1520.

Madan, D.B. and F. Milne, 1994. Contingent claims valued and hedged by pricing and investing in a basis. *Mathematical Finance* 4, 223-245.

Martens, M. and J. Zein, 2004. Predicting financial volatility: High-frequency time-series forecast vis-à-vis implied volatility. *Journal of Futures Markets* 24, 1005-1028.

Pong, S.-H., M.B. Shackleton, S.J. Taylor and X. Xu, 2004. Forecasting currency volatility: a comparison of implied volatilities and AR(FI)MA models. *Journal of Banking and Finance* 28, 2541-2563.

Poon, S.-H. and C.W.J. Granger, 2003. Forecasting volatility in financial markets: a review. *Journal of Economic Literature* 41, 478-539.

Ritchey, R.J., 1990. Call option valuation for discrete normal mixtures. *Journal of Financial Research* 13, 285-296.

Shackleton, M.B., S.J. Taylor and P. Yu, 2010. A multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices. *Journal of Banking and Finance* 34, 2678-2693.

Silverman, B.W., 1986. Density Estimation for Statistical and Data Analysis. Chapman Hall.

Taylor, S.J., 2005. Asset Price Dynamics, Volatility, and Prediction. Princeton University Press. Princeton.

Taylor, S.J., P.K. Yadav and Y. Zhang, 2010. The information content of implied volatilities and model-free volatility expectations: evidence from options written on individual stocks. *Journal of Banking and Finance* 34, 871-881.

Trujillo-Barrera, A., P. Garcia and M. Mallory, 2012. Density forecasts of lean hog futures prices. Proceedings of the NCCC-134 conference on applied commodity price analysis, forecasting, and market risk management. St. Louis, MO.

[http://www.farmdoc.illinois.edu/nccc134].

Vergote O. and J.M.P. Gutiérrez, 2012. Interest rate expectations and uncertainty during ECB governing council days: evidence from intraday implied densities of 3-month Euribor. *Journal of Banking and Finance* 36, 2804-2823.

Yun, J., 2014. Out-of-sample density forecasts with affine jump diffusion models. *Journal of Banking and Finance* 47, 74-87.

Table 1

List of 17 DJIA constituent stocks studied.

Number	Company	Exchange	Symbol	Industry	Date Added
1	Alcoa	NYSE	AA	Aluminum	1959/6/1
2	American Express	NYSE	AXP	Consumer finance	1982/8/30
3	AT&T	NYSE	T	Telecommunication	1999/11/1
4	Boeing	NYSE	BA	Aerospace and defense	1987/3/12
5	Cisco Systems	NASDAQ	CSCO	Computer networking	2009/6/8
6	General Electric	NYSE	GE	Conglomerate	1907/11/7
7	Hewlett-Packard	NYSE	HPQ	Computers & technology	1997/3/17
8	The Home Depot	NYSE	HD	Home improvement retailer	1999/11/1
9	Intel	NASDAQ	INTC	Semiconductors	1999/11/1
10	IBM	NYSE	IBM	Computers & technology	1979/6/29
11	Johnson & Johnson	NYSE	JNJ	Pharmaceuticals	1997/3/17
12	JPMorgan Chase	NYSE	JPM	Banking	1991/5/6
13	McDonald's	NYSE	MCD	Fast food	1985/10/30
14	Merck	NYSE	MRK	Pharmaceuticals	1979/6/29
15	Pfizer	NYSE	PFE	Pharmaceuticals	2004/4/8
16	Wal-Mart	NYSE	WMT	Retail	1997/3/17
17	Walt Disney	NYSE	DIS	Broadcasting and entertainment	1991/5/6

Table 2

Summary statistics for IBM option data. Information about out-of-the-money (OTM) and at-the-money (ATM) options on IBM stock from 2003 to 2012.

	Total	Average per day	Maximum per day	Minimum per day
Calls	47709	19	46	6
Puts	61402	25	74	5
Total	109111	44	115	12

		Maturity			
		<1 month	Between 1 and 6 months	>6 months	Subtotal
Moneyness	S/K				
Deep OTM put	>1.05	6462	30100	13596	50158
		5.92%	27.59%	12.46%	45.97%
OTM put	1.01-1.05	2040	5123	1839	9002
		1.87%	4.70%	1.69%	8.25%
At/near the money	0.99-1.01	1049	2641	973	4663
		0.96%	2.42%	0.89%	4.27%
OTM call	0.95-0.99	2278	5733	2330	10341
		2.09%	5.25%	2.14%	9.48%
Deep OTM call	<0.95	3168	20393	11386	34947
		2.90%	18.69%	10.44%	32.03%
Subtotal		14997	63990	30124	109111
		13.74%	58.65%	27.61%	100.00%

Table 3

Summary statistics for risk-neutral calibrated parameters for IBM and across all stocks. Estimates are summarized for the risk-neutral dynamics (2). The parameters are estimated each day from 2003 to 2012, from the OTM and ATM options, by minimizing the MSE of the fitted option prices. We apply the constraints $0 \leq \kappa \leq 36$, $0 \leq \theta \leq 1$, $\sigma \geq 0$, $-1 \leq \rho \leq 1$, $0 \leq V_0 \leq 1$.

	κ	θ	σ	ρ	V_0
IBM					
Mean	1.6861	0.5042	1.2038	-0.6723	0.0653
Median	0.1661	0.3457	0.8617	-0.6652	0.0444
Standard deviation	3.6779	0.4201	2.1596	0.1051	0.0726
Averages across all firms					
Mean	3.0401	0.4037	1.9675	-0.6331	0.1081
Median	1.1136	0.2308	1.0267	-0.6305	0.0692
Standard deviation	5.2434	0.3594	5.6694	0.1462	0.1206

Table 4

Log-likelihoods for IBM and the average log-likelihoods across 17 stocks, from 2003 to 2012. The forecast horizons are 1, 5, 10 and 22 trading days ahead. Overlapping forecasts are evaluated for horizons beyond one day. The numbers shown are the log-likelihoods of the HAR untransformed density forecasts (0 for the average across 17 stocks) and the log-likelihoods of the other forecasts in excess of the HAR benchmark values. The letter Q defines untransformed and risk-neutral densities, while the letter P denotes nonparametric transformation of the Q densities defined by (15). The numbers in bold in each row refer to the best method, which has the highest log-likelihood for the selected forecast horizon.

Horizon	Forecasts	HAR		Black-Scholes		Heston	
		Q	P	Q	P	Q	P
IBM							
1 day	2487	-4312.5	124.1	33.0	**128.5**	-9.3	113.2
1 week	2483	-6419.1	157.3	100.1	**217.4**	100.9	167.2
2 weeks	2478	-7222.1	189.3	78.1	**270.9**	76.1	176.2
1 month	2466	-8232.5	179.9	77.2	**257.6**	65.7	151.1
Average							
1 day	2487	0	202.0	95.3	**213.3**	-12.2	152.1
1 week	2483	0	144.1	129.6	**215.4**	8.9	112.8
2 weeks	2478	0	126.6	58.4	**172.0**	-64.1	58.9
1 month	2466	0	149.0	56.5	**195.8**	-114.5	61.4

Table 5

Best methods. Each count is the frequency that the method has the highest log-likelihood for the selected forecast horizon across 17 stocks. Separate counts are shown for risk-neutral (Q) and transformed (P) densities. The log-likelihood always increases after transforming from Q to P, for all stocks, horizons and methods.

Horizon	Forecasts	Q			P		
		HAR	Black-Scholes	Heston	HAR	Black-Scholes	Heston
1 day	2487	1	14	2	4	12	1
1 week	2483	0	14	3	0	17	0
2 weeks	2478	2	13	2	1	16	0
1 month	2466	4	10	3	3	14	0
Total		7	51	10	8	59	1

Table 6
Amisano and Giacomini test results for overlapping forecasts when the Newey-West adjustment is made and 20 autocorrelations are used. The numbers are the times that the log-likelihood for the row method is statistically better than the column method at the 5% level for 17 stocks.

1 day	HAR-*Q*	HAR-*P*	BS-*Q*	BS-*P*	Heston-*Q*	Heston-*P*	Total
HAR-*Q*	/	0	0	0	2	1	3
HAR-*P*	15	/	11	0	15	3	44
BS-*Q*	10	0	/	0	10	2	22
BS-*P*	14	5	16	/	17	7	59
Heston-*Q*	3	0	0	0	/	0	3
Heston-*P*	13	1	7	0	16	/	37
1 week	**HAR-*Q***	**HAR-*P***	**BS-*Q***	**BS-*P***	**Heston-*Q***	**Heston-*P***	**Total**
HAR-*Q*	/	0	0	0	1	1	2
HAR-*P*	3	/	1	0	3	2	9
BS-*Q*	1	1	/	0	5	1	8
BS-*P*	4	6	7	/	11	5	33
Heston-*Q*	0	0	1	0	/	0	1
Heston-*P*	0	2	1	0	9	/	12
2 weeks	**HAR-*Q***	**HAR-*P***	**BS-*Q***	**BS-*P***	**Heston-*Q***	**Heston-*P***	**Total**
HAR-*Q*	/	0	0	0	1	1	2
HAR-*P*	0	/	0	0	3	2	5
BS-*Q*	0	0	/	0	2	1	3
BS-*P*	1	2	1	/	9	3	16
Heston-*Q*	0	0	0	0	/	0	0
Heston-*P*	0	0	0	0	6	/	6
1 month	**HAR-*Q***	**HAR-*P***	**BS-*Q***	**BS-*P***	**Heston-*Q***	**Heston-*P***	**Total**
HAR-*Q*	/	0	0	0	0	0	0
HAR-*P*	0	/	0	0	0	0	0
BS-*Q*	0	1	/	0	0	0	1
BS-*P*	0	1	0	/	0	0	1
Heston-*Q*	0	0	0	0	/	0	0
Heston-*P*	0	0	0	0	1	/	1

Table 7

Amisano and Giacomini test results for IBM overlapping forecasts when the Newey-West adjustment is made to the estimated variance of $\bar{d}$ and 20 autocorrelations are used. The null hypothesis states that two different density forecasting methods have equal expected log-likelihood. The numbers are the test statistics. Positive values imply the row method has higher log-likelihood than the column method. * indicates that the null hypothesis is rejected at the 5% significance level when $|t| > 1.96$.

1 day	HAR-*P*	BS-*Q*	BS-*P*	Heston-*Q*	Heston-*P*
HAR-*Q*	-3.17*	-1.01	-3.27*	0.26	-2.96*
HAR-*P*		2.98*	-0.33	3.68*	0.56
BS-*Q*			-4.05*	2.75*	-3.32*
BS-*P*				4.40*	0.99
Heston-*Q*					-4.34*

1 week	HAR-*P*	BS-*Q*	BS-*P*	Heston-*Q*	Heston-*P*
HAR-*Q*	-0.96	-0.83	-1.25	-0.61	-1.01
HAR-*P*		0.73	-1.73	1.06	-0.29
BS-*Q*			-1.69	-0.01	-1.00
BS-*P*				2.56*	1.83
Heston-*Q*					-2.38*

2 weeks	HAR-*P*	BS-*Q*	BS-*P*	Heston-*Q*	Heston-*P*
HAR-*Q*	-0.68	-0.54	-0.93	-0.26	-0.64
HAR-*P*		0.60	-1.58	1.61	0.24
BS-*Q*			-1.10	0.01	-0.60
BS-*P*				2.87*	2.10*
Heston-*Q*					-2.38*

1 month	HAR-*P*	BS-*Q*	BS-*P*	Heston-*Q*	Heston-*P*
HAR-*Q*	-0.53	-0.40	-0.69	-0.17	-0.41
HAR-*P*		0.46	-0.95	0.93	0.22
BS-*Q*			-0.82	0.04	-0.33
BS-*P*				1.73	1.08
Heston-*Q*					-1.34

Table 8

Kolmogorov-Smirnov test results for IBM overlapping forecasts. The numbers are the percentage p-values of the KS test for the null hypothesis that the terms u_t are uniformly distributed. The letter Q defines untransformed and risk-neutral densities, while the letter P denotes nonparametric transformation of the Q densities, defined by (15). * indicates that the p-value is greater than 50%. The null hypothesis is rejected at the α significance level when $p < \alpha$.

Horizon	Forecasts	HAR(%)		Black-Scholes(%)		Heston(%)	
		Q	P	Q	P	Q	P
1 day	2487	0.42	*	0.00	*	0.00	*
1 week	2483	0.01	*	0.00	*	0.00	*
2 weeks	2478	0.00	*	0.00	*	0.00	*
1 month	2466	0.00	*	0.00	*	0.00	*

Table 9

Kolmogorov-Smirnov test results for overlapping forecasts. The numbers are the frequencies that the null hypothesis is rejected at the 5% significance level for 17 stocks.

Forecast horizon	HAR		Black-Scholes		Heston	
	Q	P	Q	P	Q	P
1 day	13	0	17	0	17	0
1 week	16	0	16	0	16	0
2 weeks	14	0	17	0	16	0
1 month	17	0	17	0	16	0

Table 10

Berkowitz LR3 test results for IBM overlapping forecasts. The null hypothesis states that the variables $y_i = \Phi^{-1}(u_i)$ are Gaussian with mean 0, variance 1 and are independent of y_{i-h}. The tabulated numbers are the LR3 test statistic defined by (24), and the estimates of the variance and AR parameters. * indicates that the null hypothesis is rejected at the 5% significance level when LR3 $>$ 7.81.

Forecast horizon		HAR		Black-Scholes		Heston	
		Q	*P*	*Q*	*P*	*Q*	*P*
1 day	AR	-0.01	-0.01	0.01	0.00	0.01	0.00
	variance	1.17	0.97	0.79	0.97	0.78	0.97
	LR3	42.19*	1.74	74.23*	1.36	75.42*	1.47
1 week	AR	-0.04	-0.07	-0.06	-0.08	-0.05	-0.06
	variance	1.18	0.96	0.86	0.96	0.84	0.96
	LR3	50.06*	15.07*	44.06*	19.08*	44.69*	11.08*
2 weeks	AR	0.01	0.00	0.01	0.00	0.01	0.01
	variance	1.11	0.96	0.82	0.96	0.81	0.96
	LR3	30.61*	2.42	67.22*	2.51	56.32*	2.54
1 month	AR	0.01	-0.02	0.01	-0.02	-0.02	-0.01
	variance	1.12	0.96	0.86	0.96	0.90	0.96
	LR3	44.77*	3.42	62.91*	4.12	23.80*	2.64

Table 11

Berkowitz LR3 test results for overlapping forecasts. The numbers are the frequencies that the null hypothesis is rejected at the 5% significance level for 17 stocks.

Forecast horizon	HAR		Black-Scholes		Heston	
	Q	*P*	*Q*	*P*	*Q*	*P*
1 day	16	4	15	4	17	4
1 week	15	6	15	9	14	6
2 weeks	16	3	16	4	15	3
1 month	17	4	15	5	17	6

Figure 1. Black-Scholes, Heston and HAR one-day ahead density forecasts for IBM on January 2nd 2003.

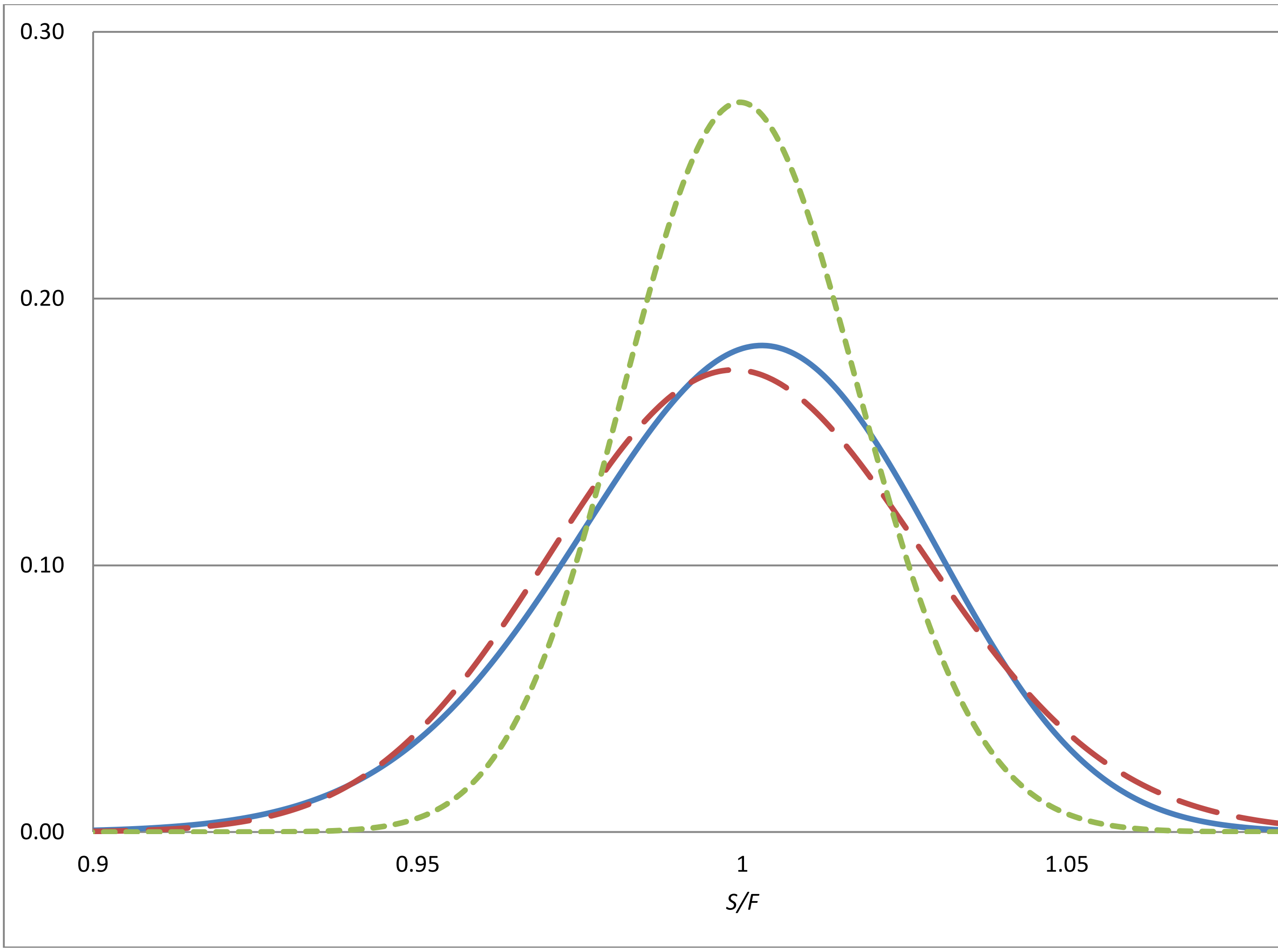

Figure 2. Black-Scholes, Heston and HAR one-month ahead density forecasts for IBM on January 2nd 2003.

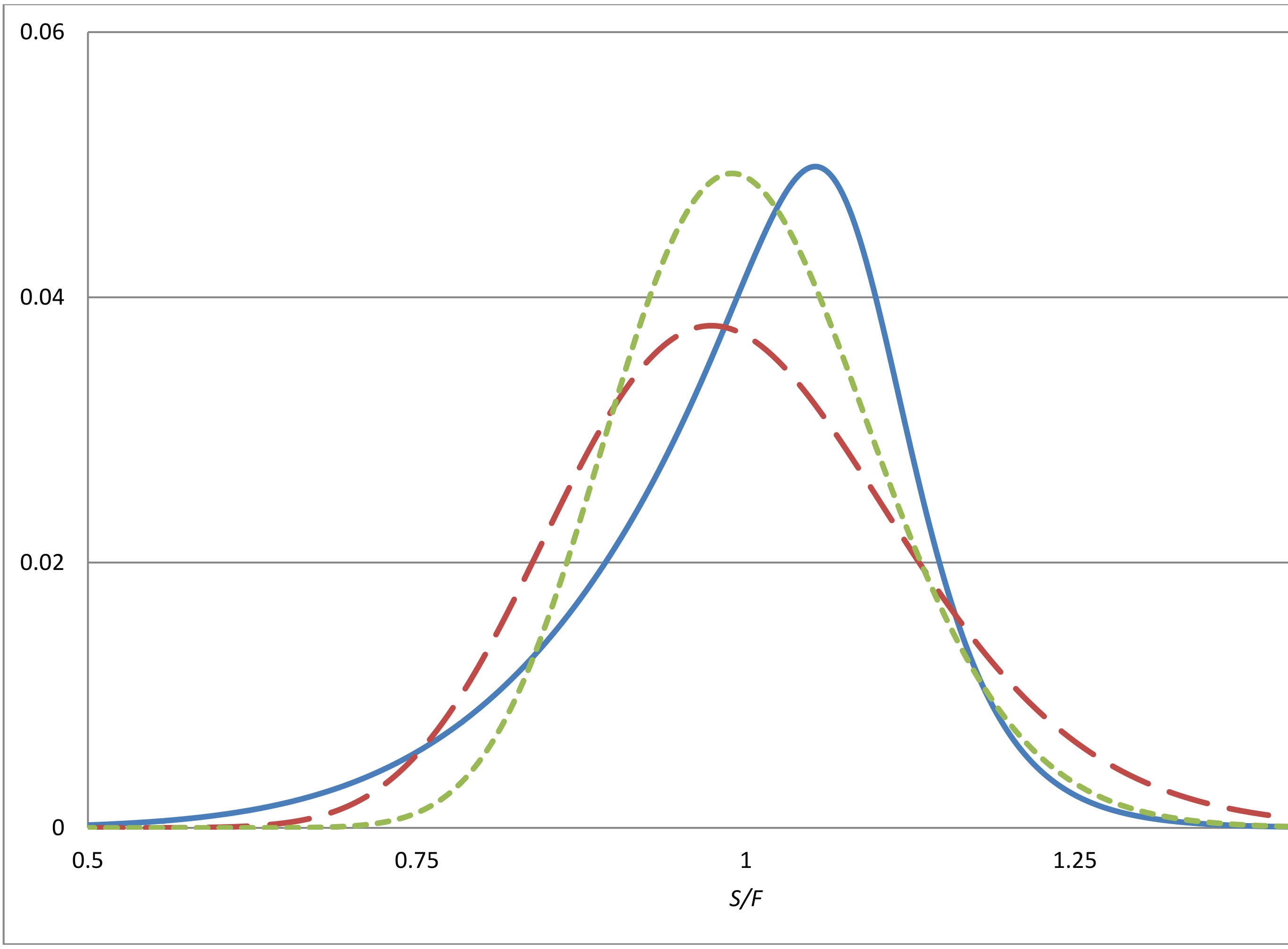

Figure 3. Function $\tilde{C}(u) - u$ for one-day ahead forecasts from the risk-neutral HAR, Black-Schol
nonparametric real-world transformations, for IBM.

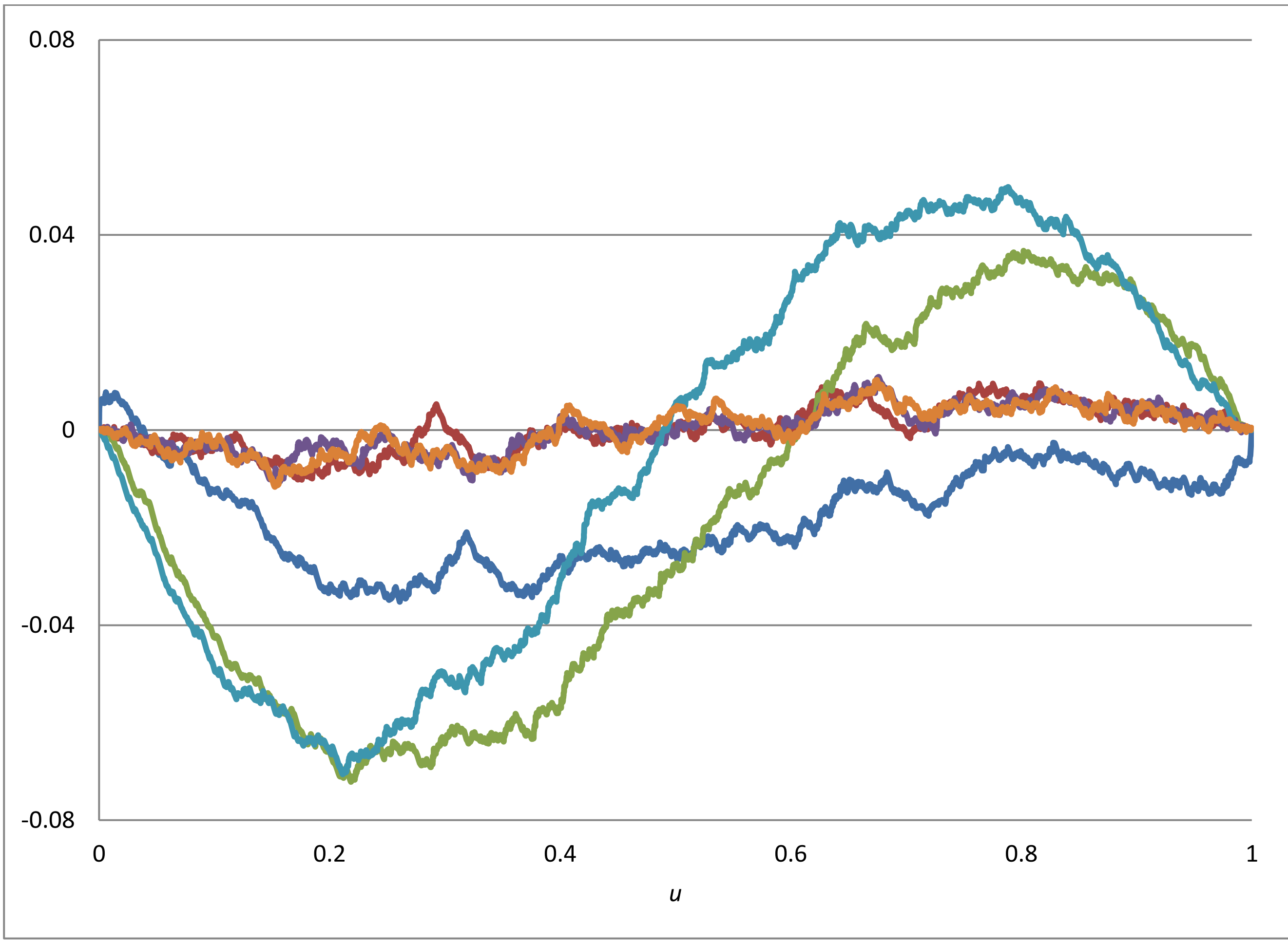

Figure 4. Nonparametric calibration densities $\hat{c}(u)$ estimated from one-day ahead HAR, Black-Scholes and