Density Forecasts of Crude-Oil Prices Using Option-Implied and ARCH-Type Models

Esben Høg\textsuperscript{1} Leonidas Tsiaras\textsuperscript{2}

July 16, 2009

15th International Conference on Computing in Economics and Finance
University of Technology, Sydney

\textsuperscript{1}Department of Mathematical Sciences, Aalborg University, Denmark
\textsuperscript{2}Aarhus School of Business and CREATE\textsuperscript{S}, Aarhus University, Denmark
Overview

Crude Oil Density Forecasts:

- Forecasts of the crude oil price probability distribution
- Necessary input to estimate risk measures such as VaR
- The standard approach is to compute them using time-series models that rely on past returns (for instance GARCH)
- Affected by factors that are difficult to incorporate in a time-series model

In This Paper:

- We compare GARCH-based forecasts with those obtained by option prices and calibration functions
- Our empirical results favor an option-based, non-parametric calibration, density forecast method
Why options?

Option Prices:
- Are determined by aggregate market expectations about the risk-neutral dynamics of the underlying asset until the maturity of the contract
- Given a set of observed option prices, option-implied expectations can be extracted

Informational Efficiency of Option-Implied Forecasts:
- Option-implied volatility has enjoyed great success in terms of forecasting (Poon and Granger, 2003)
- Option-implied betas\(^3\) and option-implied correlations\(^4\) are also found useful in terms of forecasting

---

Risk-Neutral (Forecast) Densities

- The risk-neutral density (RND, henceforth) can be inferred using a set of observed European option prices.
- Central banks routinely monitor RNDs in order to gauge market sentiment.

**Advantages:**
- Have forward-looking features by construction.
- Contain information that is absent from the record of historical returns.

**Disadvantages:**
- Risk-neutral densities refer to an artificial world where investors do not require compensations for bearing risks.
- Appropriate adjustments are required to correct these forecast densities and define "real-world" specifications.
Real-World Forecast Densities

Previous Research:

- Anagnostou-Basioudis et al. (2005) find that they perform similar to GARCH-based forecasts.
- Their risk-adjustment methods are however problematic and their results in-sample.
- Liu et al. (2007) prefer "real-world" densities over GARCH-based specifications but their results are also in-sample.
- Shackleton et al. (2007) provide the first out-of-sample evidence that "real-world" densities perform better for forecast horizons longer than one week.
Contributions (1)

**Crude Oil Forecast Densities:**
- The current literature has only focused on models that rely solely on past returns.
- Particularly, fat-tailed GARCH specifications are suggested in Giot and Laurent (2003), Hung, Lee and Liu (2008) and Asem, Costello and Gardner (2008) for constructing VaR forecasts.

We also construct forecast densities and VaR related quantities but use option prices instead of only GARCH-based estimates.
Contributions (2)

Accuracy of Option-Implied Forecasts:

- The only genuine evidence regarding the out-of-sample forecast performance of option-implied densities are provided by Shackleton et al (2007) for the case of the S&P500 index.

- The superior performance of option-implied densities may not hold for other, less liquid, assets.

Towards this end, we add further empirical evidence in the related literature.
Contributions (3)

**Piece-wise Evaluation of "Real-World" Densities:**

- End users usually have interests only in certain segments of the target density.
- A risk-adjustment method may improve the overall performance of the risk-neutral forecasts but still fail to correct forecasts for the region of interest.
- Thus, it may be inappropriate to generally recommend an option-based approach.

For this reason we pay special focus on testing for potential mis-specification in various density segments.
Historical Densities

- Historical forecast densities, i.e. forecast densities obtained using only past returns, are constructed using a standard GARCH(1,1) specification.

- In particular, mean and conditional variance of the returns are modeled as follows,

\[
  r_t = \mu + e_t, \\
  e_t = \sqrt{h_t} z_t, \quad z_t \sim i.i.d. \ D(0, 1) \\
  h_t = \omega + \alpha e_{t-1}^2 + \beta h_{t-1},
\]

with \( D(0, 1) \) being either the standard normal or the standardized Student-t probability density.

- The target densities are obtained by simulating the price process for the forecast horizon of interest.
The risk-neutral density (RND) of a particular asset at some future horizon $T$ can be estimated using a panel of $N$ European call option prices $\text{CALL}(K_1), \ldots, \text{CALL}(K_N)$ with strike prices $K_1, \ldots, K_N$ and maturity $T$.

We employ the flexible GB2 density to approximate the RND. The corresponding parameter vector is estimated by minimizing sum of squared price errors, i.e.

$$
\sum_{i=1}^{N} \left( \text{CALL}(K_i) - \text{CALL}_{GB2}(K_i|\theta_{GB2}) \right)^2 , \quad (4)
$$

where $\text{CALL}_{GB2}(K_i|\theta_{GB2})$ is the theoretical (fitted) price of a call option given by a GB2 density with a parameter vector $\theta_{GB2}$. 

Esben Høg, Leonidas Tsiaras

Crude-Oil Price Density Forecasting
From Risk-Neutral to Real-World Densities

Two alternative approaches can be used in order to convert risk-neutral densities into "real-world" specifications

**Economic Theory:**
- Several studies find that empirical pricing kernels are not consistent with standard utility theory for the S&P500
- Additional complications arise because our underlying asset is crude oil, so extra assumptions are needed

**Statistical Calibrations:**
- A more straightforward approach is to treat RNDs as mis-specified and attempt to properly re-calibrate them
- Two calibrations functions are employed in this paper: the Beta function as in Fackler and King (1990) and a non-parametric approach as in Diebold et al (1999) and Shackleton et al (2007)
Calibration Functions

- Let $f_Q(x)$ and $F_Q(x)$ denote the risk-neutral density and cumulative distribution function of the time $T$ price of the underlying asset $X_T$.

- Following Bunn (1984), David (1984) and Fackler and King (1990) also define the calibration function $C(u)$ as the real-world cumulative distribution of the random variable $U = F_Q(X_T)$. The real-world density, $f_P(x)$, and cumulative distribution function, $F_P(x)$, of $X_T$ are then:

\[ F_P(x) = C(F_Q(x)) \]  
\[ f_P(x) = \frac{dF_P(x)}{dx} = \frac{dC(F_Q(x))}{dx} = \frac{dC}{dF} \frac{dF}{dx} = c(F_Q(x))f_Q(x) \]
RWD Estimation: Parametric Calibration Function (1)

Parametric Specification:

- Flexible, two-parameter, specification with cumulative density, $C(u)$, and probability density function, $c(u)$, defined as:

$$
C(u|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \int_0^u t^{\alpha-1}(1 - t)^{\beta-1} \, dt \tag{7}
$$

$$
c(u|\alpha, \beta) = \frac{u^{\alpha-1}(1 - u)^{\beta-1}}{B(\alpha, \beta)}, \quad B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \tag{8}
$$
Every time a RND is available, a RWD specification can be generated through eq. (6) if the calibration function is known.

**Estimation Procedure:**
1) We collect all past risk-neutral forecast densities $f_{Q,i}$, given by the GB2 parameter vector $\theta_i$, and corresponding realizations of the target quantity $X_{i+1}$
2) We estimate the $\alpha$ and $\beta$ parameters of the Beta density by maximum likelihood, i.e. by finding the parameters that maximize the following quantity:

$$\log(L(X_1, X_2, \ldots, X_t) = \sum_{i=0}^{t-1} \log(f_{P,i}(X_{i+1}|\theta_i, \alpha, \beta)) \quad (9)$$
RWD Estimation: Non-Parametric Calibration Function

For the non-parametric estimation of the calibration function we follow Shackleton et al (2007)

**Estimation Procedure:**

1) We collect all past realizations of the variable \( u_i = F_{Q,i}(X_{i+1}) \)
2) We define a new series \( z_i = \Phi^{-1}(u_i) \) where \( \Phi(\cdot) \) is the inverse of the cumulative distribution function of the standard normal density
3) We fit a normal kernel in order to obtain the kernel density \( h(z) \), and empirical distribution function, \( H(z) \), of \( z_0, z_1, \ldots, z_{t-1} \)

The real-world cumulative and probability density function are then:

\[
F_P(x) = C(F_Q(x)) \quad \text{and} \quad f_P(x) = \frac{f_Q(x)h(z)}{\phi(z)} \quad , \quad (10)
\]

where \( \phi(\cdot) \) is the probability density function of the standard normal
Properties of Correctly Specified Forecast Densities

Let $f_t(u)$ denote a forecast density produced at time $t$ and $X_{t+1}$ the corresponding realization. Also define the following quantities:

$$\text{PIT}_t = \frac{X_{t+1}}{\int_{-\infty}^{\text{PIT}_t} f_t(u)du} \quad \text{and} \quad z_t = \Phi^{-1}(\text{PIT}_t), \quad (11)$$

where $\Phi^{-1}(\cdot)$ is the inverse standard normal cumulative distribution function.

As discussed in Diebold, Gunter and Tay (1998) and Berkowitz (2001), if $f(u)$ is correctly specified, the $\text{PIT}_t$ series should be i.i.d. $\text{Unif}(0,1)$ and the $z_t$ series should be i.i.d. $\text{N}(0,1)$. 

Esben Høg, Leonidas Tsiaras
Crude-Oil Price Density Forecasting
Goodness-of-fit tests (1)

We first test the independence and standard normality (or uniformity) assumption separately (Diebold et al, 1998)

**Standard normality assumption:**
- Kolmogorov-Smirnov test

**Independence assumption:**
- Following Diebold et al (1998), we investigate autocorrelation pattern in the first four moments of the $z_t$ series
- In particular, a test of no autocorrelation in the $k^{th}$ moment is conducted by regressing $(z_t - \bar{z}_t)^k$ on 6 and 12 lags, for $k = 1, 2, 3, 4$.
- A Lagrange Multiplier test is then conducted using the statistic $(T - l)R^2$, where $l$ is the number of lags
Goodness-of-fit tests (2)

- A joint test for independence and standard normality in the $z_t$ series is conducted using the LR3 test of Berkowitz (2001), i.e. by estimating the following model,

$$z_t - \mu = \rho(z_{t-1} - \mu) + \epsilon_t, \quad \epsilon \sim N(0, \sigma^2) \quad (12)$$

- The unknown parameter vector $\theta = [\mu, \sigma^2, \rho]$ can be estimated via maximum likelihood.

- A hypothesis that the $z_t$ series follows a uncorrelated Gaussian process with zero mean and unit variance against a Gaussian AR(1) alternative can be tested using a standard LR test.

- Specifically, the following statistic:

$$LR3 = -2(L(0, 1, 0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho}))$$

has a $\chi^2$ distribution with three degrees of freedom under the null.
Goodness-of-fit tests (3)

### Piecewise and interval density evaluation:

- Let $R_{i,t}$ denote a forecast density region of interest containing $p_{i,t}$ probability mass, i.e. $Pr(X_{t+1} \in R_{i,t}) = p_{i,t}$

- Let $HIT_{t+1} = 1$ if $X_{t+1} \in R_{i,t}$ and 0 otherwise.

- Assuming that $p_{i,t}$ is constant, then the $HIT_t$ series should be i.i.d. Bernoulli ($p_{i,t}$).

- We follow Patton (2006) and estimate a logit regression on the $HIT_t$ series with four explanatory variables, i.e. a constant and the number of previous hits in the last 1, 5 and 10 observations.

- A likelihood ratio test is then conducted to check whether the constant is equal to $p_{i,t}$ and that all other variables are insignificant.
The out-of-sample likelihood values can be used to rank competing forecast models (Bao et al., 2007; Shackleton et al., 2007).

In particular, given a sequence of $n$ ex-ante density forecasts $f_t(x)$ and $n$ ex-post realizations of the target quantity $X_{t+1}$ the out-of-sample log-likelihood (OLL) is defined as:

$$OLL = \sum_{t=0}^{n-1} \log(f_t(X_{t+1}))$$  \hspace{1cm} (13)

The correct forecast model has the highest expected log-likelihood.

The highest statistic indicates the specification that is closer to the true target density using the Kullback-Leibler Information Criterion as a loss function.
Data Issues: Futures Data

- New York Mercantile Exchange (NYMEX)
- January 1987-December 2006
- GARCH models are estimated using the 5 years of past data
- Returns are computed from the nearest expiration contract except when the time until expiration is less than four days, when we switch to the next contract
Data Issues: Options Data

- NYMEX, April 1990-December 2006
- American options written on the shortest delivery futures contract
- Barone-Adesi and Whaley (1987) formula to compute the prices of European options
- We select contracts with appr. 28 days to expire
- All risk-neutral densities are non-overlapping
- Three years of data reserved for initial calibrations, appr. 14 years left for forecasting
Some Remarks

Summary of competing forecast models:
- GARCH model with either standard normal (GARCH-N) or standardized Student-\(t\) innovations (GARCH-T)
- Risk-neutral density forecasts (RND)
- Real-world densities, obtained by RND forecasts adjusted using either a parametric (RWD-P) or a non-parametric (RWD-NP) calibration function

Summary of evaluation criteria:
1) G.O.F. Tests:
   - Separate tests of standard normality and independence
   - Joint tests of standard normality and independence
   - Mis-specification tests in particular density segments
2) Likelihood Criteria
   - Out-of-sample likelihood levels
Normality Tests

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Normality Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A. Descriptive Statistics</td>
<td>Mean</td>
</tr>
<tr>
<td>RND</td>
<td>0.181</td>
</tr>
<tr>
<td>GARCH-N</td>
<td>0.056</td>
</tr>
<tr>
<td>GARCH-T</td>
<td>0.011</td>
</tr>
<tr>
<td>RWD-P</td>
<td>0.06</td>
</tr>
<tr>
<td>RWD-NP</td>
<td>0.059</td>
</tr>
</tbody>
</table>

Panel B. Kolmogorov-Smirnov Test

<table>
<thead>
<tr>
<th>KS Statistic</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RND</td>
<td>0.139</td>
</tr>
<tr>
<td>GARCH-N</td>
<td>0.097</td>
</tr>
<tr>
<td>GARCH-T</td>
<td>0.076</td>
</tr>
<tr>
<td>RWD-P</td>
<td>0.093</td>
</tr>
<tr>
<td>RWD-NP</td>
<td>0.05</td>
</tr>
</tbody>
</table>

- If a method produces well specified forecasts, then $z_t \sim N(0, 1)$

Esben Høg, Leonidas Tsiaras
Independence Tests

TABLE 2
Autocorrelation Tests

<table>
<thead>
<tr>
<th></th>
<th>1st Moment</th>
<th>2nd Moment</th>
<th>3rd Moment</th>
<th>4th Moment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>RND</td>
<td>0.18</td>
<td>0.25</td>
<td>0.24</td>
<td>0.26</td>
</tr>
<tr>
<td>GARCH-N</td>
<td>0.27</td>
<td>0.3</td>
<td>0.25</td>
<td>0.3</td>
</tr>
<tr>
<td>GARCH-T</td>
<td>0.33</td>
<td>0.35</td>
<td>0.22</td>
<td>0.17</td>
</tr>
<tr>
<td>RWD-P</td>
<td>0.18</td>
<td>0.22</td>
<td>0.18</td>
<td>0.21</td>
</tr>
<tr>
<td>RWD-NP</td>
<td>0.19</td>
<td>0.14</td>
<td>0.11</td>
<td>0.13</td>
</tr>
</tbody>
</table>

NOTE: P-values of the LM tests corresponding to the null of no autocorrelation in the first four moments of the $z_t$ series.

Independence tests for the $z_t$ series produced by each of the forecast models.
Joint Tests

Joint tests of normality and independence of the $z_t$ series produced by each of the forecast models

- If the forecast model are correctly specified then we have that $\mu = 0$, $\sigma^2 = 1$ and $\rho = 0$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{\rho}$</th>
<th>LR3</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RND</td>
<td>0.19</td>
<td>0.91</td>
<td>-0.001</td>
<td>6.526</td>
<td>0.089</td>
</tr>
<tr>
<td>GARCH-N</td>
<td>0.062</td>
<td>0.908</td>
<td>-0.009</td>
<td>1.366</td>
<td>0.714</td>
</tr>
<tr>
<td>GARCH-T</td>
<td>0.026</td>
<td>0.897</td>
<td>-0.025</td>
<td>1.122</td>
<td>0.77</td>
</tr>
<tr>
<td>RWD-P</td>
<td>0.07</td>
<td>0.921</td>
<td>-0.01</td>
<td>1.336</td>
<td>0.721</td>
</tr>
<tr>
<td>RWD-NP</td>
<td>0.068</td>
<td>0.82</td>
<td>-0.03</td>
<td>3.879</td>
<td>0.275</td>
</tr>
<tr>
<td>Region</td>
<td>RND</td>
<td>GARCH-N</td>
<td>GARCH-T</td>
<td>RWD-P</td>
<td>RWD-NP</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----</td>
<td>---------</td>
<td>---------</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td>$PIT_t \in (0, 0.2)$</td>
<td>0.414</td>
<td>0.479</td>
<td>0.29</td>
<td>0.13</td>
<td>0.163</td>
</tr>
<tr>
<td>$PIT_t \in (0.2, 0.4)$</td>
<td>0.69</td>
<td>0.861</td>
<td>0.94</td>
<td>0.541</td>
<td>0.378</td>
</tr>
<tr>
<td>$PIT_t \in (0.4, 0.6)$</td>
<td>0.621</td>
<td>0.214</td>
<td>0.095</td>
<td>0.332</td>
<td>0.4</td>
</tr>
<tr>
<td>$PIT_t \in (0.6, 0.8)$</td>
<td>0.816</td>
<td>0.63</td>
<td>0.792</td>
<td>0.828</td>
<td>0.509</td>
</tr>
<tr>
<td>$PIT_t \in (0.8, 1)$</td>
<td>0.003</td>
<td>0.804</td>
<td>0.736</td>
<td>0.236</td>
<td>0.813</td>
</tr>
</tbody>
</table>

### Panel B. Unequal Density Mass

<table>
<thead>
<tr>
<th>Region</th>
<th>RND</th>
<th>GARCH-N</th>
<th>GARCH-T</th>
<th>RWD-P</th>
<th>RWD-NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PIT_t \in (0, 0.15)$</td>
<td>0.415</td>
<td>0.66</td>
<td>0.78</td>
<td>0.48</td>
<td>0.289</td>
</tr>
<tr>
<td>$PIT_t \in (0.15, 0.35)$</td>
<td>0.836</td>
<td>0.837</td>
<td>0.87</td>
<td>0.772</td>
<td>0.99</td>
</tr>
<tr>
<td>$PIT_t \in (0.35, 0.65)$</td>
<td>0.027</td>
<td>0.18</td>
<td>0.25</td>
<td>0.14</td>
<td>0.8</td>
</tr>
<tr>
<td>$PIT_t \in (0.65, 0.85)$</td>
<td>0.491</td>
<td>0.365</td>
<td>0.095</td>
<td>0.231</td>
<td>0.57</td>
</tr>
<tr>
<td>$PIT_t \in (0.85, 1)$</td>
<td>0.1</td>
<td>0.273</td>
<td>0.28</td>
<td>0.431</td>
<td>0.434</td>
</tr>
</tbody>
</table>

### Panel C. Tails and Center

<table>
<thead>
<tr>
<th>Region</th>
<th>RND</th>
<th>GARCH-N</th>
<th>GARCH-T</th>
<th>RWD-P</th>
<th>RWD-NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PIT_t \in (0, 0.25)$</td>
<td>0.724</td>
<td>0.727</td>
<td>0.76</td>
<td>0.289</td>
<td>0.23</td>
</tr>
<tr>
<td>$PIT_t \in (0.25, 0.75)$</td>
<td>0.043</td>
<td>0.165</td>
<td>0.242</td>
<td>0.047</td>
<td>0.632</td>
</tr>
<tr>
<td>$PIT_t \in (0.75, 1)$</td>
<td>0.001</td>
<td>0.053</td>
<td>0.097</td>
<td>0.014</td>
<td>0.417</td>
</tr>
<tr>
<td>$PIT_t \in (0, 0.25) \cup (0.75, 1)$</td>
<td>0.042</td>
<td>0.17</td>
<td>0.135</td>
<td>0.013</td>
<td>0.178</td>
</tr>
</tbody>
</table>

**TABLE 4**

HIT TESTS
Likelihood Comparisons

TABLE 5
Out-of-Sample Log-Likelihoods

<table>
<thead>
<tr>
<th></th>
<th>RND</th>
<th>GARCH-N</th>
<th>GARCH-T</th>
<th>RWD-P</th>
<th>RWD-NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period A</td>
<td>-154.4586</td>
<td>-153.2976</td>
<td>-152.0864</td>
<td>-155.1178</td>
<td>-152.7213</td>
</tr>
<tr>
<td>Period B</td>
<td>-224.2925</td>
<td>-221.8704</td>
<td>-221.6892</td>
<td>-222.5730</td>
<td>-219.1821</td>
</tr>
<tr>
<td>Entire Sample</td>
<td>-378.7511</td>
<td>-375.1680</td>
<td>-373.7756</td>
<td>-377.6908</td>
<td>-371.9035</td>
</tr>
</tbody>
</table>

-The highest statistic indicates the method that produces forecast densities that are closer to the true target density using the KLIC as a loss function.
Summary and Conclusions

**Forecast Setting:**
- Crude oil price densities, monthly horizon, 1993-2006 inclusive
- Comparisons between ARCH-type forecast methods and option-implied specifications

**Empirical Results:**
- Our preferred density forecast method is the RWD-NP specification
- Forecast densities based on the RWD-NP model rank favorably in goodness-of-fit tests
- They also have the highest out-of-sample log-likelihood values, which suggest the RWD-NP forecasts are closer to the true target density
- On the contrary, the parametric calibration function we employ fails to correct all sources of error


