The recovery theorem with application to risk management

Vaughan van Appel[1,2] and Eben Maré *[3]

[1]Department of Statistics, University of Johannesburg, P.O. Box 524, Auckland Park, 2006, South Africa, E-mail: vvanappel@gmail.com

[2]Department of Actuarial Science, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

[3]Department of Mathematics and Applied Mathematics, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa, E-mail: Eben.Mare@up.ac.za

January 4, 2020

Abstract

The forward-looking nature of option prices provides an appealing way to extract risk measures. In this paper, we extract forecast densities from option prices that can be used in forecasting risk measures. More specifically, we extract a real-world return density forecast, implied from option prices, using the recovery theorem. In addition, we backtest and compare the predictive power of this real-world return density forecast with a risk-neutral return density forecast, implied from option prices, and a simple historical simulation approach. In an empirical study, using the South African FTSE/JSE Top 40 index, we found that the extracted real-world density forecasts, using the recovery theorem, yield satisfying forecasts of risk measures.

Keywords: Density forecasting, recovery theorem, risk management, value at risk

1 Introduction

John Maynard Keynes said that "*successful investing is anticipating the anticipations of others*". In essence, financial derivative securities are forward-looking and essentially embed information about investors' beliefs about the distribution of asset returns (see e.g., Christoffersen et al., 2013; Hollstein et al., 2019; Dillschneider and Maurer, 2019). Investors, policy makers, and risk managers therefore look at market variables (or derivations thereof) aiming to gauge forecasts of economic variables or sentiment (or changes thereof) (see Bliss and Panigirtzoglou, 2004).

Financial derivative securities are frequently used to infer information. The prime example is the VIX index, which is derived from the prices of equity index options traded on the Chicago Board of Options Exchange (CBOE). This index reflects the market's view of 30-day index volatility and is used as a risk-sentiment gauge by investors. Bollerslev et al. (2009) showed that the difference between the VIX and the realised volatility on the S&P500

*Correspondence: Eben.Mare@up.ac.za

index carries significant explanatory power for future returns. Moreover, the VIX index is calculated using a model-free approach, illustrating the effectiveness of model-free approaches in the literature (see e.g., CBOE, 2009; Christoffersen et al., 2013). A more recent innovation by the CBOE is the SKEW index, which reflects the index option market's perceptions of so-called tail risk (see e.g., CBOE, 2011; Christoffersen et al., 2013).

The ability to accurately forecast future asset prices is an important and frequently studied problem in financial economics (see e.g., Bollerslev et al., 2009; Crisóstomo and Couso, 2018). The recent global financial crisis highlighted this problem, where many conventional financial theories were unable to realistically forecast risk measures. Recent studies have shown that option-implied moments, such as the VIX and SKEW, have predictive power for the realised variance (see e.g., Hollstein et al., 2019).

Forecasts of the option-implied return density can provide risk managers with more information than forecasts of the moments alone (see Barone-Adesi, 2016). Such measures of risk include Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR), which are two popular measures of risk used by financial practitioners and regulators, which is related to a quantile of the return distribution. More specifically, VaR is a single value summarising the potential loss of a financial asset (or portfolio). In percentage terms this corresponds to the α-th percentile of the asset return distribution and CVaR is a measure of tail-risk, which measures "how bad things could get" (see e.g., McNeil et al., 2005). That is,

$$VaR_{(1-\alpha)} = F^{-1}(\alpha) \tag{1.1}$$

and CVaR for the discrete case is defined as

$$\text{CVaR}_{(1-\alpha)} = \text{E}[R|R \leq \text{VaR}_{(1-\alpha)}], \tag{1.2}$$

where $1-\alpha$ is the confidence level, $F(R)$ the return forecasted cumulative distribution function, and $R = S_T/S_0$ the random variable representing the asset return from time zero to time T.

Many traditional strategies of measuring VaR rely on a parametric return density, such as the normal density, and past (historical) data to make market assumptions (see e.g., McNeil et al., 2005). In practice, financial returns exhibit skewness and kurtosis that are not captured in the normal assumption framework (see e.g., Cont, 2001). Consequently, this has rekindled great interest in fat-tail distributions (see e.g., Hull and White, 1998b). In contrast to using historical data, one can also make use of market quoted option prices to extract a forward-looking risk-neutral return density forecast (see e.g., Barone-Adesi, 2016; Breeden and Litzenberger, 1978). The purported forward-looking nature of option prices makes it conceptually better suited for forecasting than a historical scheme, especially during stressed economic environments.

In particular, historical simulation and risk-neutral methods are the most widely used methods in financial risk management, where most financial institutions prefer to use historical simulations to manage risk (see Pérignon and Smith, 2010). However, Christoffersen et al. (2013) and Crisóstomo and Couso (2018) found that methods based on option-implied information generally outperformed historical-based estimates. Similarly, Shackleton et al. (2010) compared the real-world option-implied densities to that of historical densities, where they found that the real-world option-implied forecasts for two- and four-week horizons were superior to that of the historical forecasts. The transformation from risk-neutral to real-world return densities have been studied in several papers (see e.g., Bakshi et al., 2003; Bliss and

Panigirtzoglou, 2004; Shackleton et al., 2010; Ross, 2015). More specifically, the recovery theorem, proposed by Ross (2015), is a model-free method that extracts a real-world return density forecast from option prices.

The aim of this paper is to (i) extract, backtest, and compare the real-world return density forecast model using the recovery theorem to various historical and risk-neutral density forecast models found in the literature; and (ii) backtest the tail of the return density forecast models for risk management purposes and show that one can extract reliable risk measures using option-implied data and the recovery theorem. Moreover, research into the forecasting ability of real-world return densities are scarce in the literature. Furthermore, it is likely that the entire return density forecast may be misspecified, but performs better in certain regions of the density, such as the tail, which is often more valuable to risk managers (see Berkowitz, 2001).

The remainder of this paper is structured as follows. Section 2 establishes methods for building return density forecast models. More specifically, we will construct two return density forecast models by historical simulation, three risk-neutral return density forecast models implied from option prices, and two real-world return density forecast models implied from option prices using the recovery theorem. Section 3 studies some commonly used backtesting approaches found in the literature. Section 4 analyses the forecasts in an empirical study using the South African FTSE/JSE Top 40 (Top40) index. In addition, the performance of the real-world return density forecast method will be backtested and benchmarked, using the Top40 index, against the two historical simulated return density forecast methods and the three risk-neutral return density forecast models, with a specific focus on risk management.

2 Extracting return densities

In principle, a good forecast density should coincide with the true return density of the asset or portfolio under study (see Knüppel, 2015). In this section, we describe some methods available for extracting the return density. More specifically, we discuss two historical simulation methods for extracting the return density forecast and five option-implied methods for extracting the return density forecast.

2.1 Historical simulation

Many securities have return distributions with so-called fat tails. Moreover, Cont (2001) presented some statistical stylized facts, which emerged in empirical studies of most asset returns. Furthermore, he showed that it is particularly difficult to reproduce many of these properties with a parametric model, requiring at least four parameters in the return distribution (i.e., a location parameter, scale parameter, a parameter describing the decay of the tails, and an asymmetry parameter to allow different behaviours between the tails). This occurrence has persuaded many risk managers to use historical simulation, rather than using a parametric model building approach, to extract the return density.

Historical simulation involves creating a database of the daily/weekly/monthly change in the asset over a period of time. For example, suppose that we have recorded 100 days of daily returns and we are interested in the 5th percentile of the daily return density (i.e., $\text{VaR}_{(0.95)}$). This would correspond to the 5th worst change out of the 100 days of asset value returns. This method will be referred to as the historical simulation method in this paper. The first drawback in using past data for simulation is that the forecast density will be slow to react to market shifts. Therefore, Hull and White (1998a) incorporated a volatility updating approach to adjust the historical database using the following exponential weighted moving average (EWMA) model:

$$\sigma_t^2 = \alpha\sigma_{t-1}^2 + (1-\alpha)R_{t-1}^2 \tag{2.1}$$

with $\alpha = 0.94$. Hence, suppose we are estimating the return density at time $N-1$ for time N. Let:

R_t : be the historical return on day t for the period covered by the historical sample ($t < N$)

$\hat{\sigma}_t^2$: be the historical EWMA estimate of the variance of the return for period t forecasted at the end of period $t-1$

$\hat{\sigma}_N^2$: is the EWMA estimate of the variance for period N. This estimate is made at the end of period $N-1$.

In the Hull and White (1998a) volatility updating approach, the original historical return data, R_t, is adjusted by multiplying the historical return data by the ratio of the current volatility to the volatility at the time of the observation, that is,

$$R_t^* = R_t\frac{\hat{\sigma}_N}{\hat{\sigma}_t}. \tag{2.2}$$

This method will be referred to as the historical-HW method for the remainder of this paper. The second drawback with historical simulation methods is that there may not be enough data available in the market to form a historical database, especially for new securities.

Since historical simulation requires a large database of past returns, which possibly consists of returns when the market was in a different economic environment, we next consider using a forward-looking approach to extract the return density forecast.

2.2 Risk-neutral densities

This method uses option prices to extract the return density function. Christoffersen et al. (2013) provided a description of situations when option-implied forecasts are likely to be most useful, such as, when the option market is highly liquid.

2.2.1 Model free risk-neutral density

Breeden and Litzenberger (1978) showed that the implied risk-neutral return density of a security can be extracted from a set of European-style option prices. For example, the value of a European call option at time $t = 0$, under the option-implied risk-neutral density of the underlying asset, $f(S_T)$, with expiration date T and strike x is calculated as follows:

$$\mathrm{C}(T,x) = e^{-rT}E[(S_T - x)^+] = e^{-rT}\int_x^\infty (S_T - x)f(S_T)\mathrm{d}S_T. \tag{2.3}$$

The cumulative distribution function (CDF), denoted by $F(x)$, can be obtained by taking the first order partial derivative of $\mathrm{C}(T,x)$ with respect to the strike x, i.e.,

$$\frac{\partial \mathrm{C}(T,x)}{\partial x} = -e^{-rT}\left[1 - F(x)\right]. \tag{2.4}$$

Rearranging (2.4), yields an expression for the implied risk-neutral CDF,

$$F(x) = 1 + e^{rT} \frac{\partial \mathrm{C}(T,x)}{\partial x}. \tag{2.5}$$

Thereafter, the conditional probability density function (PDF) is obtained by taking the partial derivative of (2.5) with respect to x, as follows:

$$f(x) = e^{rT} \frac{\partial^2 \mathrm{C}(T,x)}{\partial x^2}. \tag{2.6}$$

In practice, a continuum of traded strikes is not directly observed in the markets, especially in South Africa where option price data are sparse and noisy. Therefore, in this paper, we implemented the so-called stochastic volatility inspired (SVI) model to extract a dense implied volatility surface, which is then used to compute a dense set of call option prices across the full strike range for each expiry date (see Flint and Maré, 2017). This return PDF will be referred to as the risk-neutral density (RND) forecast model for the remainder of this paper.

2.2.2 Stochastic volatility models

In this section, we consider two stochastic volatility models to extract the risk-neutral return density, namely the Heston (1993) model and Bates (1996) model. The risk-neutral dynamics for the Heston model is given by:

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sqrt{V_t}S_t\mathrm{d}W_t^{(1)} \tag{2.7}$$

$$\mathrm{d}V_t = \kappa(\theta - V_t)\mathrm{d}t + \nu\sqrt{V_t}\mathrm{d}W_t^{(2)}, \tag{2.8}$$

where the parameter r represents the risk-free rate, κ models the speed of mean reversion of the variance, θ the long term variance, ν to volatility of variance, ρ the correlation between the two driving Brownian motions $W_t^{(1)}$ and $W_t^{(2)}$, S_0 the spot rate, and V_0 the spot variance.

It is well-known that adding jumps to the spot price process could improve the agreement between theoretical and observed option prices, especially in stressed markets (see e.g., Crisóstomo and Couso, 2018). Therefore, the Bates model is simply an extension of the Heston model with independent jumps added to the security price dynamics in (2.7), giving the following risk-neutral dynamics:

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sqrt{V_t}S_t\mathrm{d}W_t^{(1)} + (Y-1)S_t\mathrm{d}N_t \tag{2.9}$$

$$\mathrm{d}V_t = \kappa(\theta - V_t)\mathrm{d}t + \nu\sqrt{V_t}\mathrm{d}W_t^{(2)}, \tag{2.10}$$

where N_t is a Poisson process, which models the number of jumps with intensity $\lambda > 0$ and Y is the jump size distribution, which in this case is a log-normal distribution.

Heston (1993) and Bakshi and Madan (2000) provided analytical expressions for the characteristic function of $\log(S_T)$, which is then used to obtain the cumulative distribution function and risk-neutral density function of S_T, denoted by $F(x)$ and $f(x)$ respectively, for positive values of x (see e.g., Shackleton et al., 2010):

$$F(x) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{\exp(-iu\log(x))\psi(u)}{iu}\right]\mathrm{d}u \tag{2.11}$$

$$f(x) = \frac{1}{\pi x}\int_0^\infty \mathrm{Re}\left[\exp(-iu\log(x))\psi(u)\right]\mathrm{d}u, \tag{2.12}$$

where $\psi(u) = E\left[\exp(iu\log(S_T))\right]$ denotes the characteristic function of $\log(S_T)$.

In the next section, we transform the option-implied information to a real-world distribution. Shackleton et al. (2010) found in an empirical study that the real-world distribution improved forecasting performance for two- and four-week horizons.

2.3 Real-world densities using the recovery theorem

Real-world probabilities differ from risk-neutral probabilities in that investors require a premium that compensates them for carrying risk. The transformation from risk-neutral to real-world densities rely on assumptions (see e.g., Bliss and Panigirtzoglou, 2004; Shackleton et al., 2010; Ross, 2015; Dillschneider and Maurer, 2019). Moreover, Ross (2015) proposed a model-free method to recover the real-world transition matrix from a Markovian state variable S, under a particular set of assumptions, using market-based derivative prices. These assumptions are: (i) the transition state prices are strictly positive, (ii) the transition state prices are time-homogeneous, and (iii) the pricing kernel is transition independent. Firstly, he used the method proposed by Breeden and Litzenberger (1978) (see also Section 2.2.1) to construct a $n \times m$ state price matrix, S, by taking the second derivative with respect to the strike of a European call option at each tenor, t, i.e.,

$$S(t,x) = e^{rt}\frac{\partial^2 \mathrm{C}(t,x)}{\partial X^2}, \quad t = 1,\dots,m. \tag{2.13}$$

Numerically approximating (2.13) yields the forward-looking state price function for each tenor. Secondly, he constructs a $n \times n$ one period ahead irreducible time-homogeneous state transition probability matrix:

$$P_{i,j} = \Pr(S_{t+1} = j | S_t = i), \quad t = 1,\dots,m-1, \tag{2.14}$$

where the elements of P can easily be estimated by solving the following system of equations:

$$S_{t+1}^\top = S_t^\top P, \quad t = 1,2,\dots,m-1. \tag{2.15}$$

Intuitively, $P_{i,j}$ denotes the contingent price of a security that pays out one unit of currency if the security moves from state i to state j in one period, which is known as the contingent forward prices of a security. If one denotes $A = S_{i,j}^\top$, where $1 \le i \le n$ and $1 \le j \le m-1$, and $B = S_{i,j+1}^\top$, then (2.15) can be rewritten as an ordinary least squares (OLS) problem, as follows:

$$P = \arg\min_P \|AP - B\|_2^2 \tag{2.16}$$

$$\text{subject to} \quad S_1 = P_{i_0} \tag{2.17}$$

$$P_{i,j} \ge 0 \quad (i,j = 1,\dots,n), \tag{2.18}$$

where $\|\cdot\|_2$ denotes the Euclidean norm. Since S_1 is the one period ahead state price vector and P is a one period state transition matrix, we have by definition a constraint (2.17), where i_0 is the current state (normally defined at the centre row of the transition matrix P, i.e., $i_0 = (n+1)/2$). After estimating P, using standard optimisation techniques, we can extract the real-world return distribution, f, from P. The transition kernel, ψ in Ross's framework is defined as the ratio price per unit of probability, i.e.,

$$\psi_{i,j} = \frac{P_{i,j}}{f_{i,j}}. \tag{2.19}$$

Equation (2.19) is commonly recognised as the Radon-Nikodym derivative. Intuitively, Ross (2015) solves the two unknown quantities in (2.19) by assuming that the pricing kernel is transition independent and by using the Perron-Frobenius theorem to extract a unique positive eigenvector and eigenvalue. Thereafter, the elements of f can be calculated. We refer the interested reader to Ross (2015) for a more detailed representation of the recovery theorem and to Flint and Maré (2017) for details on a practical implementation.

However, accurately estimating P has proven to be difficult in the literature (see e.g., Kiriu and Hibiki, 2019; Sanford, 2018; Van Appel and Maré, 2018). Kiriu and Hibiki (2019) proposed a regularisation method with prior information to stabilise the estimation of P. For the prior information, $\bar{P}$, they suggest that $P_{i,j}$ should be similar to $P_{i+k,j+k}$ for all $k \leq \min(n-i, n-j)$. Furthermore, they estimated P, using a problem specific error function in an attempt to balance the relative gain in the objective function from each term in the regularised optimisation problem, as follows:

$$P = \underset{P\geq 0}{\arg\min} \|AP - B\|_2^2 + \zeta \left\|P - \bar{P}\right\|_2^2 \tag{2.20}$$

$$= \underset{P\geq 0}{\arg\min}\, y^{\text{fit}}(\zeta) + \zeta y^{\text{reg}}(\zeta) \tag{2.21}$$

$$\text{subject to} \quad (2.17) \text{ and } (2.18), \tag{2.22}$$

where,

$$\bar{P} = \begin{bmatrix} \sum_{k=1}^{i_0} S_{k,1} & S_{i_0+1,1} & \cdots & S_{n-1,1} & S_{n,1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \sum_{k=1}^{2} S_{k,1} & S_{3,1} & \cdots & S_{i_0,1} & S_{i_0+1,1} & S_{i_0+2,1} & \cdots & S_{n,1} & 0 \\ S_{1,1} & S_{2,1} & \cdots & S_{i_0-1,1} & S_{i_0,1} & S_{i_0+1,1} & \cdots & S_{n-1,1} & S_{n,1} \\ 0 & S_{1,1} & \cdots & S_{i_0-2,1} & S_{i_0-1,1} & S_{i_0,1} & \cdots & S_{n-2,1} & \sum_{k=n-1}^{n} S_{k,1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & S_{1,1} & S_{2,1} & \cdots & S_{i_0-1,1} & \sum_{k=i_0}^{n} S_{k,1} \end{bmatrix} \tag{2.23}$$

and ζ can be chosen by minimising the problem specific function:

$$h(\zeta) = \frac{y^{\text{fit}}(\zeta) - y^{\text{fit}}(0)}{y^{\text{fit}}(\infty) - y^{\text{fit}}(0)} + \frac{y^{\text{reg}}(\zeta) - y^{\text{reg}}(\infty)}{y^{\text{reg}}(0) - y^{\text{reg}}(\infty)}. \tag{2.24}$$

The real-world return distribution, extracted using (2.20) in the recovery theorem, will be referred to as the RWD model for the remainder of this paper.

Next, Sanford (2018) extended (2.15) to a multivariate regression model by assuming that contingent state prices are solely defined by state levels but conditioned on the volatility. That is,

$$S_{t+1}^{\top} = S_t^{\top} P + \sigma_t^{(\text{IV})} \beta, \qquad t = 1, 2, \ldots, m-1, \tag{2.25}$$

where $\sigma_t^{(\text{IV})}$ is the implied volatility state at time t as it is the best representation of the market's future volatility state and β is the volatility transition matrix. In order to stabilise the estimation of P, Van Appel and Maré (2018) extended the optimisation problem above by adding the regularisation of prior information, as such,

$$P = \underset{P,\beta}{\arg\min} \left\| AP + \sigma^{(\mathrm{IV})}\beta - B \right\|_2^2 + \zeta \|P - \bar{P}\|_2^2, \tag{2.26}$$

$$\text{subject to} \quad (2.17), (2.18) \text{ and } \beta \geq 0, \tag{2.27}$$

where $\bar{P}$ is given in (2.23). The real-world return distribution, extracted using (2.26) in the recovery theorem, will be referred to as the RWD-M model for the remainder of this paper.

3 Verification of the density forecasts

The aim of this section is to introduce methods to verify the accuracy of the return density forecast models introduced in Section 2. In practice, it is highly unlikely that an optimal model will exist as the true distribution may be too complicated to be represented by a simple mathematical model, or might not be adequately represented over all economic periods. Therefore, each model can only be considered an approximation of the truth. In order to assess whether (i) the real-world return density forecast models approximate the truth better than the simple historical simulation or risk-neutral models; and (ii) under which circumstances it can approximate the truth better, we introduce some commonly used forecast evaluation tests found in the literature, with a specific application to risk management.

Interval forecasts such as VaR are based on the inverse distribution function,

$$\bar{y}_t = F^{-1}(\alpha), \tag{3.1}$$

where, for example, $\alpha = 0.05$ for the $\mathrm{VaR}_{(0.95)}$. Christoffersen (1998) asserted that the interval should be exceeded $\alpha\%$ of the time and the violations should be uncorrelated across time. Combining these properties, the hit function

$$\mathbb{I}_t = \begin{cases} 1 & \text{if violation accurs} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

should be an independent and identically distributed (i.i.d.) Bernoulli sequence with parameter α. In a VaR setting, the Bernoulli variable rarely takes on the value 1, requiring a large number of sample observations to test the density forecast. In contrast, Rosenblatt (1952) proposed a transformation of the observed realisations into a series of i.i.d. random variables as follows:

$$x_t = \int_{-\infty}^{y_t} \hat{f}_t(u)du = \hat{F}_t(y_t), \tag{3.3}$$

where y_t is the *ex-post* return realisation and $\hat{f}(\cdot)$ is the *ex-ante* return density forecast[1]. More specifically, he showed that x_t is i.i.d. uniform on $(0, 1)$. This procedure, also commonly known as the probability integral transform (PIT), allows for a wider variety of tests to be conducted. Furthermore, this result is valid irrespective of the underlying distribution of returns, y_t, and remains valid even when the forecast model, $\hat{F}(\cdot)$, changes over time. A series

[1]Since the RND, RWD and RWD-M are recovered on a discrete grid, where the future realised returns are not likely to fall on one of the state grid points, we linearly interpolate the recovered CDF's to obtain x_t (see Jackwerth and Menner, 2017).

Electronic copy available at: https://ssrn.com/abstract=3407209

of forecast evaluation tests using graphical displays were proposed by Diebold and Mariano (1995). In contrast, Berkowitz (2001) proposed a series of likelihood ratio (LR) tests for model evaluation by generating a sequence $z_t = \Phi^{-1}(x_t)$ for the given model, where $\Phi^{-1}(\cdot)$ denotes the inverse cumulative standard normal distribution function. Then, by definition, z_t should be independent across variables with standard normal distribution. This second transformation allows for convenient tests that are associated with the Gaussian likelihood function. In particular, Berkowitz (2001) jointly assesses the mean (μ), variance (σ^2), and serial correlation (ρ) by testing the null hypothesis that z_t are i.i.d. $N(0,1)$ distributed against the following first-order autoregressive model with mean and variance other than $(0,1)$:

$$z_t - \mu = \rho(z_{t-1} - \mu) + \epsilon_t. \tag{3.4}$$

The log-likelihood function of (3.4) is often seen in statistics and is reproduced below for convenience (see Berkowitz, 2001):

$$\begin{aligned}\ell(\mu, \sigma^2, \rho|z) = &-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\rho^2}\right) - \frac{(z_1 - \mu/(1-\rho))^2}{2\sigma^2/(1-\rho^2)} \\ &- \frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - \sum_{t=2}^{T}\left[\frac{(z_t - \mu - \rho z_{t-1})^2}{2\sigma^2}\right],\end{aligned} \tag{3.5}$$

where σ^2 is the variance of ϵ_t.

Firstly, Berkowitz (2001) uses (3.5) to test for independence by considering the following LR test statistic:

$$LR_{\text{ind}} = -2(\ell(\hat{\mu}, \hat{\sigma}^2, 0) - \ell(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})). \tag{3.6}$$

Under the null hypothesis, (3.6) is distributed $\chi^2(1)$. More specifically, this test statistic is a measure of the degree to which the data supports a non-zero persistent parameter. Secondly, he also tests the null hypothesis that not only are the observations independent, but also have mean and variance equal to 0 and 1 respectively, using the following LR test statistic:

$$LR = -2(\ell(0, 1, 0) - \ell(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})). \tag{3.7}$$

Under the null hypothesis, (3.7) is distributed $\chi^2(3)$. For multi-step-ahead forecasts, practitioners use the following test statistic instead (see Knüppel, 2015):

$$LR_{\text{MS}} = -2(\ell(0, \sqrt{1-\hat{\rho}^2}, 0) - \ell(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})), \tag{3.8}$$

which is distributed $\chi^2(2)$. More specifically, multi-step-ahead forecasts are complicated with serial correlation of the outcomes with respect to the density forecast. That is, for example, if the true return turns out to be higher than our one-month forecast from today, then it is likely that the true one-month return for the next week's forecast will also be higher than the forecasted return. Therefore, (3.8) is particularly useful for density forecast evaluation for practitioners.

It must be noted that a density forecast model may be falsely rejected as it does not forecast well for particular regions of the distribution. It is possible that a forecast model performs poorly in forecasting expected returns, but performs better in predicting a certain region of the distribution, such as the tail of the distribution. Thirdly, cognisant of this, Berkowitz (2001) introduced a LR test that intentionally ignores model failures in the interior of the distribution and compares the lower tail of the foretasted density with the observed density by truncating any observed values that fall outside the tail area. Let this cut-off point be $\text{VaR} = \Phi^{-1}(\alpha)$. The new variable of interest, z_t^*, is then defined as:

$$z_t^* = \begin{cases} \text{VaR} & \text{if} \quad z_t \geq \text{VaR} \\ z_t & \text{if} \quad z_t < \text{VaR}. \end{cases} \tag{3.9}$$

The log-likelihood function for the lower tail is given as (see Berkowitz, 2001):

$$\begin{aligned} \ell(\mu, \sigma | z^*) = & \sum_{z_t^* < \text{VaR}} \left(-\frac{1}{2} \log \left(2\pi\sigma^2 \right) - \frac{1}{2\sigma^2} (z_t^* - \mu)^2 \right) \\ & + \sum_{z_t^* = \text{VaR}} \log \left(1 - \Phi \left(\frac{\text{VaR} - \mu}{\sigma} \right) \right), \end{aligned} \tag{3.10}$$

where the first two terms represent the usual Gaussian likelihood of losses and the third term is a normalisation factor arising from the truncation. As before a LR test is constructed with null hypothesis $\mu = 0$ and $\sigma^2 = 1$ against the unrestricted alternative with mean and variance other than 0 and 1 respectively, i.e.,

$$LR_{\text{tail}} = -2(\ell(0,1) - \ell(\hat{\mu}, \hat{\sigma}^2)). \tag{3.11}$$

Under the null hypothesis that the model is correct, the test statistic is distributed $\chi^2(2)$. In addition, Berkowitz (2001) showed, by using a Monte Carlo simulation, that these LR tests are powerful, even for samples containing only as few as 100 observations.

In summary, a well-specified model should simultaneously pass as many statistical backtests as possible. Therefore, in Appendix A, we briefly list additional backtests, which form part of the MATLAB Risk Management Toolbox (2018).

4 Application

In this section, we used weekly Top40 option trade data and the Top40 index prices to construct weekly one-month return density forecasts for the Top40 index over the period 05 September 2005 to 15 January 2018, giving us a total of 646 weekly one-month return density forecasts. The Top40 index is particularly useful as the underlying risky asset, as it is a key risk factor in the economy and is amongst the most liquid traded derivative contracts in the South African market. In particular, Carr and Madan (2000) showed that a major financial market index, such as the Top40 index, could be used as a proxy to price options on individual stocks that are illiquid. This makes the Top40 index an important market factor to illustrate the accuracy of forecast models. The number of weeks that a return density forecast was made for each subset of the time series considered in this study is shown in Table 1.

Table 1: Real market data

Panel A: Monthly one-Month returns

Time Period	Label	Number of weeks (N)
Sep 2005 - Jan 2018	Full-period	130

Panel B: Weekly one-Month returns

Time Period	Label	Number of weeks (N)
Sep 2005 - Dec 2007	Pre-crisis	122
Jan 2008 - Dec 2009	Crisis	104
Jan 2010 - Jan 2018	Post-crisis	420
Sep 2005 - Jan 2018	Full-period	646

The extracted density forecast models in this study are: (i) historical simulation, (ii) historical simulation with volatility updating, (iii) model-free RND extracted from option prices, (iv) RND extracted using the Heston model, (v) RND extracted using the Bates model, (vi) RWD using the recovery theorem with (2.20), and (vii) RWD-M using the recovery theorem with (2.26).

For the historical simulation methods we used a five-year historical period and for the historical-HW approach we used the EWMA model with $\alpha = 0.94$ for the volatility updating process (see Hull and White, 1998a). For the risk-neutral densities (RND, Heston and Bates), we extracted a 50-150% moneyness range for the risk-neutral return density forecasts. Similarly, for the real-world forecast densities we constructed a 51×51 one-month ahead transition probability matrix, P, spanning a 50-150% moneyness range. Recall, the one-month ahead forecast from today's state will correspond to the centre row of f. The performance of these density forecasts are evaluated using the verification tests discussed in Section 3.

In testing the consistency between the *ex-ante* return density scheme and the observed return realisation, we firstly, used Rosenblatt's PIT to transform the realisation of returns to a series of i.i.d. uniform random variables. Thereafter, we made the second transformation, proposed by Berkowitz (2001), to a realisation of i.i.d. standard normal random variables. The empirical CDF versus the standard normal CDF for each method during the global financial crisis is shown in Figure 4.1, where it can be seen that the random variable z_t deviates from the standard normal distribution for both historical simulation methods.

In addition, the Kolmogorov-Smirnov (KS) normality test and Jacque-Bera (JB) test is carried out to test for normality of z_t for each time period considered in this study. The results of these tests are shown in Table 2. More specifically, Panel A shows the normality test results for the monthly one-month return density forecast and Panel B shows the normality test results for the weekly one-month return density forecast. The JB test assesses whether the random variable, z_t, has skewness and kurtosis matching the normal distribution, which is not assessed in Berkowitz's likelihood ratio tests. Considering the density forecasting methods, it is only the Historical-HW method and the RWD-M method that passed the KS and JB normality tests, at a 5% significance level, for all time periods considered in this study.

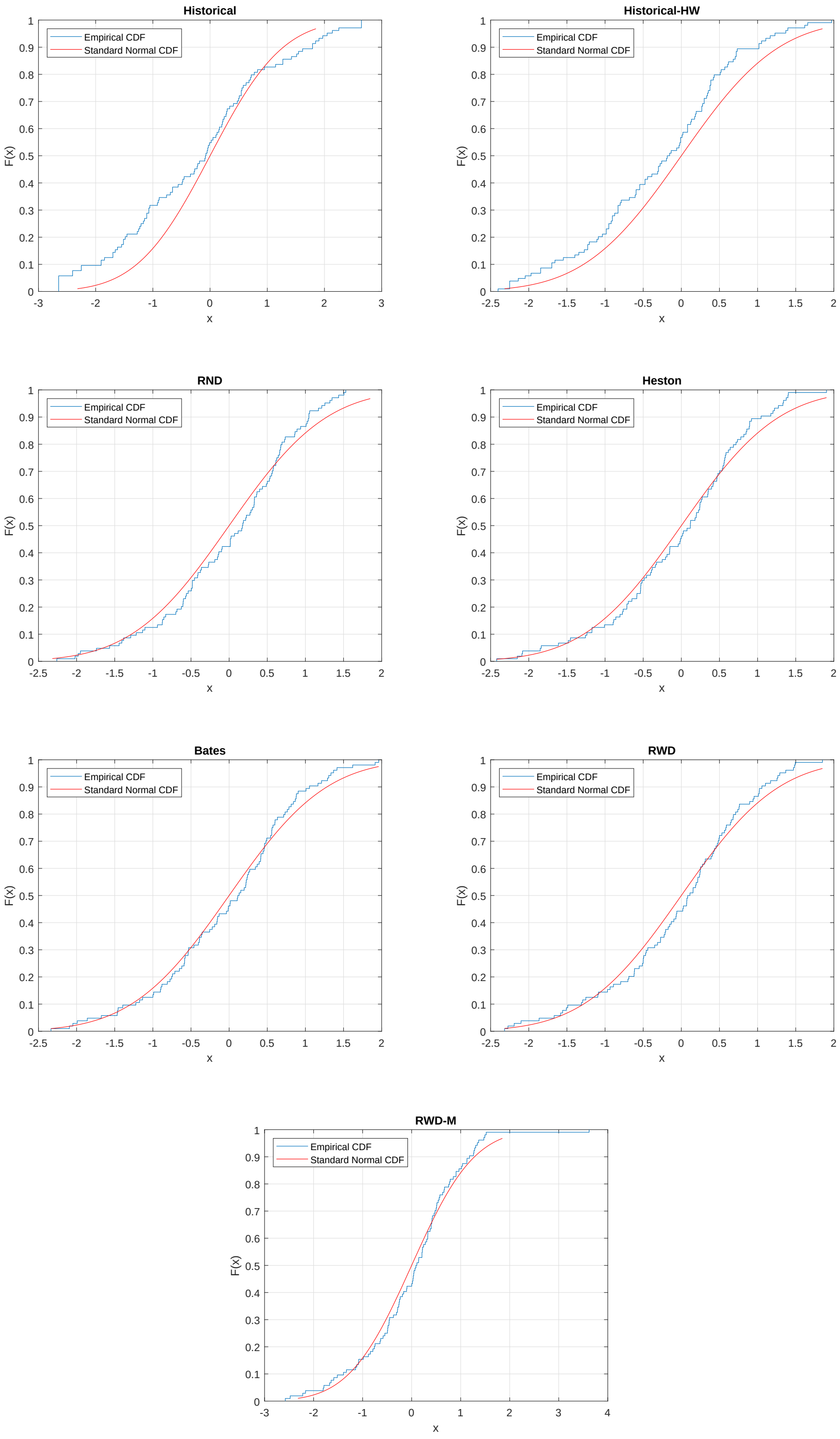


Figure 4.1: Crisis Period (Jan 2008 - Dec 2009): Empirical CDF and Normal CDF.

Table 2: Goodness-of-fit: Normality tests

Kolmogorov-Smirnov (KS) normality test

	Panel A: Monthly	**Panel B: Weekly**			
		p-values shown			
Method	**Full-period**	**Pre-crisis**	**Crisis**	**Post-crisis**	**Full-period**
Historical	0.685	0.365	0.004	0.136	0.323
Historical HW	0.853	0.119	0.050	0.680	0.667
RND	0.132	0.000	0.420	0.000	0.000
Heston	0.243	0.001	0.609	0.000	0.000
Bates	0.456	0.030	0.709	0.001	0.000
RWD	0.817	0.011	0.732	0.074	0.002
RWD-M	0.897	0.071	0.534	0.358	0.058

Jarque-Bera (JB) normality test

	Panel A: Monthly	**Panel B: Weekly**			
		p-values shown			
Method	**Full-period**	**Pre-crisis**	**Crisis**	**Post-crisis**	**Full-period**
Historical	0.500	0.172	0.372	0.397	0.500
Historical HW	0.500	0.107	0.426	0.286	0.496
RND	0.222	0.053	0.056	0.500	0.336
Heston	0.232	0.002	0.075	0.500	0.002
Bates	0.456	0.135	0.247	0.500	0.139
RWD	0.226	0.069	0.062	0.470	0.081
RWD-M	0.239	0.202	0.050	0.357	0.082

The results for Berkowitz's tests for the entire distribution is shown in Table 3. Since we are evaluating our density forecasts for more than one period ahead in Panel B, the evaluation is complicated by serial correlation of the outcomes with respect to the density forecast. Therefore, the density forecast evaluation in Panel B will be more distorted by the serial correlation of the outcomes than the density forecasts in Panel A. Due to the serial correlation, the LR_{MS} yielded the most accurate test results, where the historical-HW obtained acceptable density forecasts, at a 5% level of significance, for all time periods considered in this study. Interestingly the option-implied models (RND, Heston, Bates, RWD, and RWD-M) provided superior density forecasts during the global financial crisis, at a 5% level of significance, to the ordinary historical simulation method, which is a direct consequence of using forward-looking information rather than a historical database.

Models, which do not perform well in forecasting the entire return density, may perform better in forecasting the tail of the return density. Since risk managers are often more concerned about protection against extreme losses (i.e., the lower tail of the return density), the Berkowitz tail test is carried out for the $\mathrm{VaR}_{(0.95)}$ and $\mathrm{VaR}_{(0.90)}$. In particular, we failed to use the Berkowitz tail test for the $\mathrm{VaR}_{(0.99)}$ as we obtained no realised barrier hits for the option-implied densities over the sample period. This may be a direct consequence that options are often used for protection against large losses that may cause the option-implied densities to have a longer lower tail than what is normally expected by the spot market. The Berkowitz tail verification test results for the $\mathrm{VaR}_{(0.95)}$ and $\mathrm{VaR}_{(0.90)}$ forecasts are shown in Table 4.

The RWD-M model provided an acceptable fit, at a 5% level of significance, for the $\mathrm{VaR}_{(0.95)}$ and the $\mathrm{VaR}_{(0.90)}$ forecasts for all time periods considered in this study, where the historical-HW model provided acceptable VaR forecasts for all time periods, with exception to the $\text{VaR}_{(0.95)}$ forecast during the global financial crisis. The RND, Heston, and Bates models performed poorly during the post crisis and full period. Furthermore, the historical-HW and RWD-M are the more stable preforming models in this paper, outperforming the ordinary historical simulation, RND, Heston, Bates and RWD models for the Berkowitz tail test. In addition, the results for several backtests using the MATLAB Risk Management Toolbox (2018) are shown in Appendix A (see Tables 6 and 7).

In Figure 4.2a, the weekly historical Top40 index prices is shown and in Figure 4.2b the weekly one-month $\text{VaR}_{(0.95)}$ forecasts calculated for the Historical-HW, Heston, and RWD-M models are shown. It can be seen that during the financial crisis option-implied methods were quicker to react to market shifts than historical methods; thus making the option-implied methods more favourable during stressed economic times.

For comparison of shorter term VaR estimates, we also applied the commonly used square-root scaling law to the option-implied one-month VaR to obtain a one-week VaR forecast. This is done by multiplying the option-implied one-month VaR forecast by $1/\sqrt{5}$ (see e.g., McNeil et al., 2005). Furthermore, we also obtained the weekly one-week VaR forecast using the two historical simulation methods. We have chosen to use the one-week VaR measures as we used weekly option prices in our dataset, making it easy to compare. The backtest results for the scaled weekly one-week $\text{VaR}_{(0.95)}$ for the option-implied models, and the $\text{VaR}_{(0.95)}$ for the historical models using a weekly return database is shown in Appendix A (see Table 8). The results obtained are similar to that of the one-month VaR results where the return density forecasts obtained using option prices yielded better results than the historical simulation methods during the global financial crisis. In addition, the results for the weekly one-week $\text{VaR}_{(0.90)}$ is shown in Appendix A (see Table 9).

In Figure 4.2c, the weekly one-month $\text{CVaR}_{(0.95)}$ forecasts for the methods considered in this paper are shown. The option-implied CVaR forecasts were mostly above the historical CVaR forecasts. This indicates that the option-implied densities allocate more probability to significant losses than the historical densities. In addition, the option-implied CVaR estimates showed a significant increase over the global financial crisis period, whereas the historical methods lagged behind. Furthermore, the option-implied CVaR estimates also displayed an increase in CVaR during the period 2015-2017 when the index plateaued, indicating higher market uncertainty during the period, which was not captured by the historical simulation methods.

A challenging task for risk managers is to put in place the appropriate level of capital to cover unexpected losses. Unexpected loss is a measure of operational risk and is defined to be the difference between VaR and expected loss. In short, this is the required capital that a financial institution should have to cover unexpected losses corresponding to a desired confidence level. Figure 4.3 shows the evolution of the weekly one-month forecast of unexpected losses per Top40 index share for a 95% confidence level. Similar to the CVaR forecast, the option-implied models yielded larger unexpected loss forecasts than the historical methods. This will require financial institutions to carry more capital to cover unexpected losses under the option-implied models.

Table 3: Goodness of Fit: Berkowitz forecast density test

Panel A: Monthly one-month returns

p-values shown in parenthesis

Method	**Berkowitz**					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}$	LR_{ind}	LR	LR_{MS}
Monthly (Sep 2005 - Jan 2018)						
Historical	-0.09	1.13	-0.19	4.79 (0.0286)	7.26 (0.0639)	2.94 (0.2290)
Historical HW	-0.04	1.16	-0.20	5.19 (0.0227)	7.58 (0.0556)	2.56 (0.2774)
RND	0.17	0.71	-0.23	7.09 (0.0078)	14.44 (0.0024)	9.26 (0.0097)
Heston	0.06	0.62	-0.22	6.30 (0.0121)	16.93 (0.0007)	10.95 (0.0042)
Bates	0.07	0.70	-0.25	8.26 (0.0041)	13.61 (0.0035)	5.90 (0.0524)
RWD	0.02	0.72	-0.23	7.16 (0.0075)	11.28 (0.0103)	4.32 (0.1155)
RWD-M	-0.01	0.76	-0.24	7.48 (0.0062)	10.05 (0.0181)	2.76 (0.2517)

Panel B: Weekly one-month returns

p-values shown in parenthesis

Method	**Berkowitz**					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}$	LR_{ind}	LR	LR_{MS}
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)						
Historical	0.01	0.53	0.68	103.02 (0.0000)	74.63 (0.0000)	0.05 (0.9764)
Historical HW	0.03	0.61	0.65	88.99 (0.0000)	67.35 (0.0000)	0.37 (0.8316)
RND	0.11	0.39	0.60	75.35 (0.0000)	75.52 (0.0000)	15.31 (0.0005)
Heston	0.08	0.45	0.59	66.79 (0.0000)	63.25 (0.0000)	8.85 (0.0120)
Bates	0.08	0.61	0.59	68.57 (0.0000)	57.72 (0.0000)	1.53 (0.4644)
RWD	0.05	0.40	0.59	66.25 (0.0000)	66.56 (0.0000)	13.13 (0.0014)
RWD-M	0.04	0.42	0.59	64.52 (0.0000)	62.50 (0.0000)	10.40 (0.0055)
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)						
Historical	-0.08	1.00	0.66	83.53 (0.0000)	85.95 (0.0000)	21.76 (0.0000)
Historical HW	-0.08	0.43	0.73	130.43 (0.0000)	88.20 (0.0000)	1.96 (0.3738)
RND	0.00	0.30	0.76	146.88 (0.0000)	95.57 (0.0000)	4.98 (0.0827)
Heston	-0.01	0.33	0.75	140.88 (0.0000)	91.34 (0.0000)	3.17 (0.2049)
Bates	-0.01	0.37	0.73	124.36 (0.0000)	83.49 (0.0000)	2.45 (0.2939)
RWD	-0.01	0.40	0.71	105.54 (0.0000)	74.60 (0.0000)	2.32 (0.3123)
RWD-M	-0.01	0.57	0.64	73.09 (0.0000)	55.15 (0.0000)	0.05 (0.9735)
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)						
Historical	-0.02	0.42	0.68	364.77 (0.0000)	273.44 (0.0000)	10.54 (0.0051)
Historical HW	0.00	0.59	0.67	339.99 (0.0000)	249.78 (0.0000)	1.02 (0.5992)
RND	0.05	0.34	0.71	436.95 (0.0000)	332.16 (0.0000)	31.33 (0.0000)
Heston	0.01	0.24	0.70	406.65 (0.0000)	375.84 (0.0000)	91.26 (0.0000)
Bates	0.01	0.27	0.70	399.90 (0.0000)	349.15 (0.0000)	68.70 (0.0000)
RWD	0.01	0.36	0.69	371.23 (0.0000)	293.54 (0.0000)	27.16 (0.0000)
RWD-M	0.00	0.44	0.66	320.34 (0.0000)	250.96 (0.0000)	12.78 (0.0017)
Weekly one-Month returns: Sep 2005 - Jan 2018						
Historical	-0.02	0.54	0.68	552.41 (0.0000)	400.04 (0.0000)	0.43 (0.8084)
Historical HW	-0.01	0.57	0.68	553.86 (0.0000)	400.98 (0.0000)	1.11 (0.5738)
RND	0.05	0.34	0.70	650.03 (0.0000)	499.45 (0.0000)	47.43 (0.0000)
Heston	0.02	0.30	0.69	592.50 (0.0000)	506.51 (0.0000)	85.62 (0.0000)
Bates	0.02	0.36	0.68	534.42 (0.0000)	437.88 (0.0000)	49.42 (0.0000)
RWD	0.02	0.38	0.68	542.34 (0.0000)	433.59 (0.0000)	39.15 (0.0000)
RWD-M	0.01	0.46	0.65	460.81 (0.0000)	365.88 (0.0000)	17.76 (0.0001)

Table 4: Goodness-of-fit analysis: Berkowitz tail test

Panel A: Monthly one-month returns: Berkowitz tail

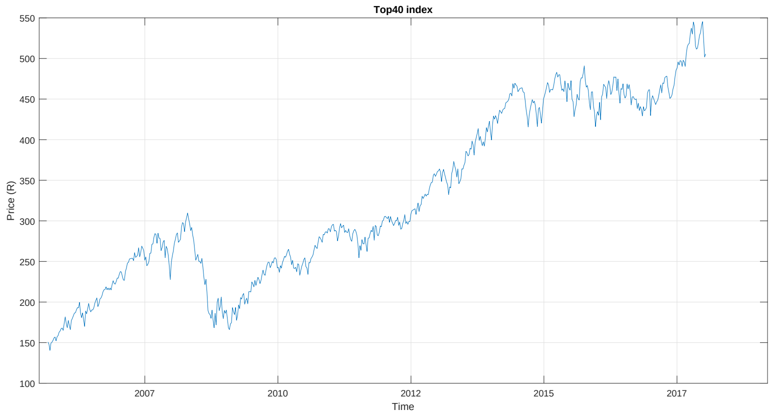
p-values shown in parenthesis

Method	VaR$_{(0.95)}$			VaR$_{(0.90)}$		
	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}
Monthly (Sep 2005 - Jan 2018)						
Historical	-0.73	0.38	2.74 (0.2547)	-0.06	1.20	1.94 (0.3791)
Historical HW	-0.31	0.75	0.46 (0.7963)	0.14	1.42	1.37 (0.5047)
RND	-1.23	0.04	8.21 (0.0165)	-0.30	0.41	4.64 (0.0984)
Heston	-0.77	0.24	2.93 (0.2309)	0.13	0.84	2.59 (0.2733)
Bates	-1.02	0.13	5.26 (0.0719)	0.00	0.69	2.99 (0.2239)
RWD	-0.81	0.24	3.16 (0.2064)	-0.20	0.62	1.27 (0.5298)
RWD-M	-0.65	0.38	1.78 (0.4099)	0.15	1.02	0.77 (0.6815)

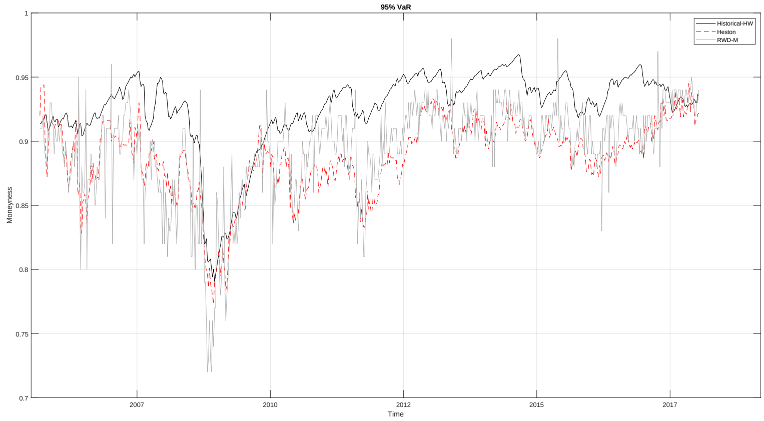
Panel B: Weekly one-month returns: Berkowitz tail

p-values shown in parenthesis

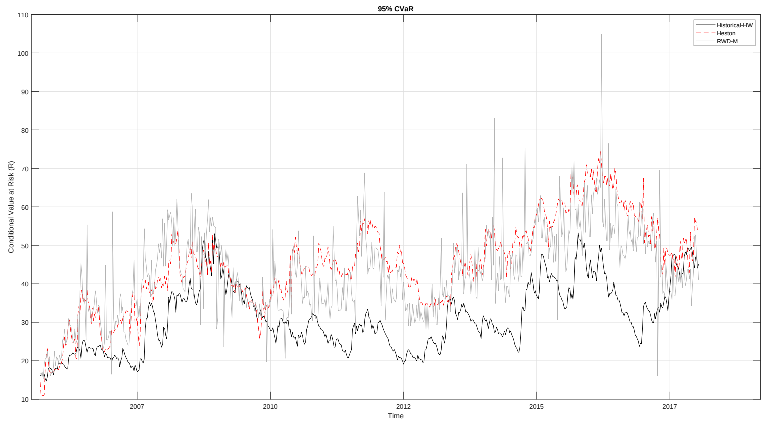
Method	VaR$_{(0.95)}$			VaR$_{(0.90)}$		
	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}	$\hat{\mu}$	$\hat{\sigma}^2$	LR_{tail}
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)						
Historical	-0.81	0.23	1.36 (0.5064)	0.14	1.31	0.53 (0.7687)
Historical HW	0.15	1.19	1.34 (0.5115)	0.40	1.70	1.33 (0.5132)
RND	-0.46	0.36	3.15 (0.2075)	0.45	1.00	5.97 (0.0506)
Heston	0.48	1.50	0.41 (0.8152)	1.56	2.99	5.75 (0.0564)
Bates	0.59	1.65	0.57 (0.7520)	1.60	3.06	5.97 (0.0505)
RWD	0.10	0.79	2.14 (0.3427)	0.96	1.68	5.79 (0.0553)
RWD-M	-0.42	0.44	1.81 (0.4045)	0.83	1.48	5.79 (0.0553)
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)						
Historical	-0.56	1.15	18.38 (0.0001)	-0.31	1.54	18.38 (0.0001)
Historical HW	-0.81	0.49	7.73 (0.0210)	-0.16	1.15	2.92 (0.2317)
RND	-0.38	0.58	0.42 (0.8089)	-0.15	0.75	0.29 (0.8653)
Heston	-0.22	0.83	0.19 (0.9102)	0.34	1.44	0.48 (0.7854)
Bates	-0.43	0.60	0.50 (0.7773)	0.00	0.99	0.01 (0.9956)
RWD	-0.20	0.85	0.18 (0.9163)	-0.15	0.91	0.30 (0.8600)
RWD-M	-0.16	0.99	0.65 (0.7216)	0.04	1.23	0.62 (0.7312)
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)						
Historical	-0.54	0.37	6.78 (0.0337)	0.25	0.58	4.85 (0.0883)
Historical HW	0.06	0.98	0.50 (0.7795)	-0.19	0.76	1.09 (0.5793)
RND	-0.41	0.26	27.28 (0.0000)	-0.07	0.41	37.31 (0.0000)
Heston	-0.68	0.16	28.39 (0.0000)	0.07	0.45	44.81 (0.0000)
Bates	-0.47	0.23	27.14 (0.0000)	-0.31	0.29	33.97 (0.0000)
RWD	-0.54	0.27	17.29 (0.0002)	-0.28	0.41	16.22 (0.0003)
RWD-M	-0.33	0.53	4.37 (0.1123)	-0.08	0.72	3.97 (0.1376)
Weekly one-Month returns: Sep 2005 - Jan 2018						
Historical	-0.17	0.92	1.81 (0.4040)	0.01	1.12	1.40 (0.4954)
Historical HW	-0.21	0.85	1.52 (0.4674)	-0.07	1.00	1.04 (0.5943)
RND	-0.15	0.50	22.26 (0.0000)	0.16	0.70	33.66 (0.0000)
Heston	0.44	1.07	13.71 (0.0011)	0.94	1.55	36.09 (0.0000)
Bates	0.39	1.02	13.90 (0.0010)	0.08	0.81	11.51 (0.0032)
RWD	-0.09	0.62	12.44 (0.0020)	-0.02	0.70	12.98 (0.0015)
RWD-M	-0.23	0.66	3.24 (0.1984)	0.09	0.93	4.22 (0.1212)



(a) Top 40 index price



(b) $\text{VaR}_{(0.95)}$



(c) $\text{CVaR}_{(0.95)}$

Figure 4.2: Comparison of the weekly Top40 index price with the forecasted weekly one-month $\text{VaR}_{(0.95)}$, and $\text{CVaR}_{(0.95)}$.

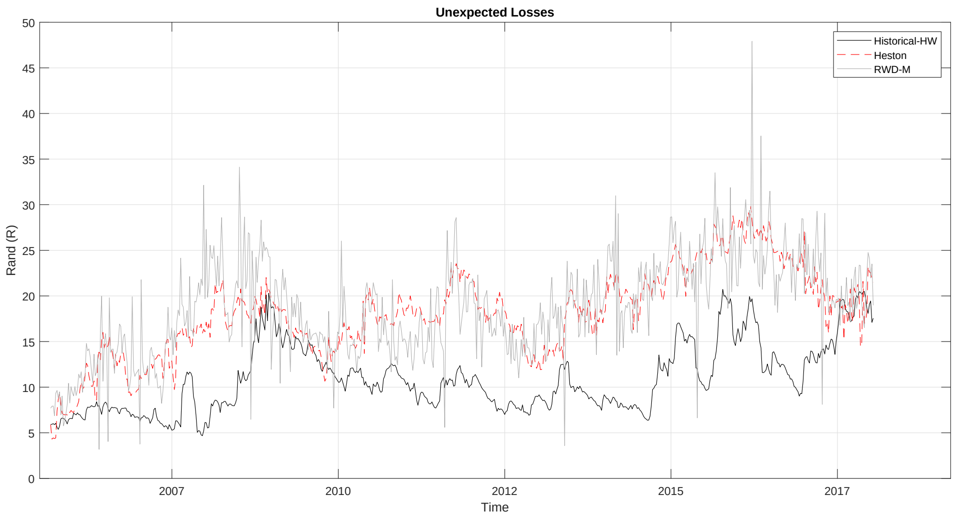


Figure 4.3: Weekly one-month forecasts of unexpected losses

In Table 5 the mean Sharpe ratio[2] and volatility is shown for each time period considered in this study. We notice that the real-world forward-looking Sharpe ratio is more sensitive and showed a considerable drop during the financial crisis period, where the other methods did not.

Table 5: Additional risk measures

Sharpe Ratio

	Panel A: Monthly	Panel B: Weekly			
Method	**Full-period**	**Pre-crisis**	**Crisis**	**Post-crisis**	**Full-period**
Historical	0.519	0.675	0.678	0.436	0.520
Historical HW	0.529	0.628	1.009	0.389	0.534
RND	-0.147	-0.110	-0.098	-0.168	-0.146
Heston	-0.137	-0.119	-0.103	-0.149	-0.136
Bates	-0.130	-0.110	-0.105	-0.142	-0.130
RWD	0.271	0.420	0.006	0.309	0.281
RWD-M	0.405	0.628	-0.040	0.504	0.440

Volatility

	Panel A: Monthly	Panel B: Weekly			
Method	**Full-period**	**Pre-crisis**	**Crisis**	**Post-crisis**	**Full-period**
Historical	0.183	0.191	0.209	0.175	0.183
Historical HW	0.169	0.183	0.291	0.137	0.170
RND	0.222	0.239	0.328	0.194	0.224
Heston	0.231	0.228	0.315	0.212	0.232
Bates	0.230	0.231	0.309	0.213	0.232
RWD	0.221	0.237	0.311	0.198	0.223
RWD-M	0.203	0.221	0.299	0.179	0.206

[2]The Sharpe ratio is calculated as the ratio of excess asset return above the risk-free rate to the standard deviation of the returns. The Sharpe ratio is a measure of risk-adjusted return and indicates how well the return of an asset compensates the investor for the risk taken.

5 Conclusion

In this study, we implemented seven methods for extracting the return density forecasts for the South African Top40 index with application to risk management. More specifically, two of these methods extracted the return density forecast using historical simulation, three methods extracted the risk-neutral return density forecast from option prices, and two methods used the recovery theorem proposed by Ross (2015) to extract the real-world return density forecast.

These methods were backtested and their performances over multiple time-periods were compared. Using a series of likelihood ratio tests, proposed by Berkowitz (2001), we found that no model proved to be reliable in extracting the entire return density forecast in all tests, where only the Historical-HW model proved to be reliable in extracting the entire return density forecast when Berkowitz's test was relaxed for serial correlation. However, it is naïve to expect that one can accurately extract the entire true market return density using a simple statistical model. A more realistic expectation is that only a specific region of the return density forecast is accurately extracted. Since risk managers are often more concerned with experiencing extreme losses, we used the Berkowitz tail test and other commonly used VaR backtests found in the literature to test whether the tail of the extracted real-world return density forecasts provided us with a more reliable VaR forecast than the historical simulation and risk-neutral VaR forecast.

In our study using the Top40 index, we found that the option-implied methods provided information about the potential losses in the Top40 index. More specifically, the extracted densities using option prices yielded superior VaR measures to the historical methods during the global financial crisis. Although the historical methods are well suited during normal economic periods, the real-world density forecasts can be an effective alternative during crisis periods. In addition, the RWD-M yielded more stable VaR forecasts over all time periods than the risk-neutral densities, making the recovery theorem useful in forecasting VaR. Moreover, using the option-implied densities will lead to overestimating the required risk capital during normal market conditions. Therefore, further research in optimally mixing the information obtained from risk-neutral, real-world and historical methods to obtain better risk forecasts can be valuable.

Acknowledgements

The authors thank Byran Taljaard, Emlyn Flint and Pierre Venter for their reading of the paper and comments received.

A Additional VaR backtesting results

In this section, we give a short description of the VaR backtests that is part of the MATLAB Risk Management Toolbox (2018).

- The traffic light (TL) test classifies the number of failures into three zones, namely, green, yellow, and red using a binomial distribution, $F(x|n, p)$ (see Basle Committee of Banking Supervision, 2011). In particular, the test computes the cumulative probability of observing up to x failures in n trails, with $p = \alpha$ and three zones:

– Green: $F(x|n,p) \leq 0.95$
– Yellow: $0.95 < F(x|n,p) \leq 0.9999$
– Red: $F(x|n,p) > 0.9999$.

This test is often used as a preliminary VaR accuracy check.

- The binomial (Bin) distribution test is an extension of Christoffersen (1998) Bernoulli test. It states that if $\mathbb{I}_t$ are i.i.d Bernoulli with parameter p, then the total number of failures, x, follows a binomial distribution with mean and variance equal to np and $np(1-p)$ respectively. Under the null hypothesis, $H_0 : p = \alpha$, the test statistic is approximated by

$$z = \frac{x - np}{\sqrt{np(1-p)}}, \tag{A.1}$$

which has a standard normal distribution.

- The proportion of failures (POF) test is a LR test proposed by Kupiec (1995). More specifically, the POF test determines whether the proportion of failures (i.e., number of failures divided by number of observations) denoted as $\hat{p}$ is consistent with the VaR confidence level. Under the null hypothesis, $H_0 : p = \alpha$, the LR test statistics is:

$$LR_{\text{POF}} = -2\log\left[(1-p)^{n-x}p^x\right] + 2\log\left[(1-\hat{p})^{n-x}(\hat{p})^x\right] \sim \chi^2(1). \tag{A.2}$$

- The time until first failure (TUFF) test, proposed by Kupiec (1995), is a LR test that measures the time until the first failure. Under the null hypothesis, $H_0 : p = 1/v$, where v is the time until the first failure in the sample, the LR test statistic is

$$LR_{\text{TUFF}} = -2\log\left[\frac{p(1-p)^{v-1}}{\hat{p}(1-\hat{p})^{v-1}}\right] \sim \chi^2(1). \tag{A.3}$$

The TUFF test is mostly used as a preliminary test to the POF test. Furthermore, it only considers the number of failures but not the time dynamics of the failures. The test also has been shown to have a low power in identifying poor VaR models.

- The conditional coverage independence (CCI) test, also known as the Markov test, assesses whether the probability of VaR failure for any given period is dependent on the outcome of the previous period (see Christoffersen, 1998). Using the indicator value in (3.2) and let $N_{i,j}$, $i = 0, 1$, $j = 0, 1$ be the number of periods in which state j occurred after state i occurred. Then let π_0 be the conditional probability of having a failure at time t, given that there was no failure at time $t-1$. Similarly, let π_1 be the conditional probability of having a failure at time t, given that there was a failure at time $t-1$. Under $H_0 : \pi_0 = \pi_1$, the LR test statistic is given as:

$$\begin{aligned} LR_{\text{CCI}} = &-2\log\left[(1-\pi)^{N_{00}+N_{01}}\pi^{N_{01}+N_{11}}\right] \\ &+ 2\log\left[(1-\pi_0)^{N_{00}}\pi_0^{N_{01}}(1-\pi_1)^{N_{10}}\pi_1^{N_{11}}\right] \sim \chi^2(1), \end{aligned} \tag{A.4}$$

where $\pi = \pi_0 + \pi_1$.

- The conditional coverage (CC) mixed test is a combination of the CCI test and the POF test. The CC test assesses whether the failures are independent and whether the correct failure rate is obtained (see Christoffersen, 1998). The LR test statistic is

$$LR_{\text{CC}} = LR_{\text{CCI}} + LR_{\text{POF}} \sim \chi^2(2). \tag{A.5}$$

  A VaR model must therefore satisfy both independence and the correct failure rate in this test, making this test appealing to practitioners.

- The time between failures independence (TBTI) test proposed by Haas (2001) is an extension of Kupiec's time until first failure (TUFF) test by not only testing the time until the first failure, but also the time between all failures. Under the null hypothesis, that failures are independent from each other, the LR test statistic is

$$LR_{TBFI} = \sum_{i=2}^{x} \left[-2\log\left(\frac{p(1-p)^{v_i-1}}{\hat{p}(1-\hat{p})^{v_i-1}} \right) \right] - 2\log\left[\frac{p(1-p)^{v-1}}{\hat{p}(1-\hat{p})^{v-1}} \right] \sim \chi^2(x), \tag{A.6}$$

  where v_i denotes the duration between the ith and $(i-1)$th failure, v the time until the first failure and x the number of failures in the sample.

- The time between failures (TBF) likelihood ratio test, introduces by Haas (2001), is a mixed LR test. Under the null hypothesis, that the correct failure rate is obtained and that the failures are independent, the test statistic is

$$LR_{\text{TBF}} = LR_{\text{POF}} + LR_{\text{TBFI}}. \tag{A.7}$$

  This test statistics is $\chi^2(x+1)$ distributed, where x is the number of failures. The advantage to this test is that it is robust, since it identifies problems in dependencies and the number of failures.

Using the MATLAB Risk Management Toolbox (2018), we show the backtest results obtained for the monthly one-month $\text{VaR}_{(0.95)}$ and $\text{VaR}_{(0.90)}$ in Tables 6 and 7, respectively. In addition, Table 9 shows the weekly one-week $\text{VaR}_{(0.90)}$.

Table 6: Goodness-of-fit: One-month $\text{VaR}_{(0.95)}$ backtests

Panel A: Monthly one-month returns

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Monthly (Sep 2005 - Jan 2018)								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	reject	reject	accept	reject	accept	reject	accept
Heston	green	accept	accept	accept	accept	accept	accept	accept
Bates	green	accept	accept	accept	accept	accept	accept	accept
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept

Panel B: Weekly one-month returns

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	accept	reject	reject	reject
RND	green	reject	reject	accept	reject	accept	reject	accept
Heston	green	accept	accept	accept	reject	reject	reject	reject
Bates	green	accept	accept	accept	reject	reject	reject	reject
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	reject	accept	accept	accept	accept	accept
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)								
Historical	yellow	reject	reject	accept	reject	reject	reject	reject
Historical HW	yellow	accept	accept	accept	reject	reject	reject	reject
RND	green	accept	accept	accept	reject	reject	reject	reject
Heston	green	accept	accept	accept	reject	reject	reject	reject
Bates	green	accept	accept	accept	reject	reject	reject	reject
RWD	green	accept	accept	accept	reject	reject	reject	reject
RWD-M	green	accept	accept	accept	reject	reject	reject	reject
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	reject	reject	reject	reject	accept	reject	reject
Heston	green	reject	reject	reject	reject	accept	reject	reject
Bates	green	reject	reject	accept	reject	accept	reject	reject
RWD	green	reject	reject	accept	reject	accept	reject	reject
RWD-M	green	reject	reject	accept	reject	accept	reject	accept
Weekly one-Month returns: Sep 2005 - Jan 2018								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	reject	reject	accept	reject	reject	reject	reject
Heston	green	reject	reject	accept	reject	reject	reject	reject
Bates	green	reject	reject	accept	reject	reject	reject	reject
RWD	green	reject	reject	accept	reject	reject	reject	reject
RWD-M	green	reject	reject	accept	reject	reject	reject	reject

Table 7: Goodness-of-fit: One-month $VaR_{(0.90)}$ backtests

Panel A: Monthly one-month returns

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Monthly (Sep 2005 - Jan 2018)								
Historical	green	accept	accept	accept	accept	accept	reject	reject
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	reject	reject	accept	accept	accept	accept	accept
Heston	green	accept	accept	accept	accept	accept	accept	accept
Bates	green	accept	accept	accept	accept	accept	accept	accept
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept

Panel B: Weekly one-month returns

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Weekly one-Month returns: Sep 2005 - Dec 2007 (Pre-Crisis)								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	reject	reject	accept	reject	reject	reject	reject
Heston	green	accept	reject	accept	reject	reject	reject	reject
Bates	green	accept	reject	accept	reject	reject	reject	reject
RWD	green	reject	reject	accept	reject	reject	reject	reject
RWD-M	green	reject	reject	accept	reject	reject	reject	reject
Weekly one-Month returns: Jan 2008 - Dec 2009 (Crisis)								
Historical	yellow	reject	reject	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	accept	accept	accept	reject	reject	reject	reject
Heston	green	accept	accept	accept	reject	reject	reject	reject
Bates	green	accept	accept	accept	reject	reject	reject	reject
RWD	green	accept	accept	accept	reject	reject	reject	reject
RWD-M	green	accept	accept	accept	reject	reject	reject	reject
Weekly one-Month returns: Jan 2010 - Jan 2018 (Post-Crisis)								
Historical	green	reject	reject	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	reject	reject	reject	reject	reject	reject	reject
Heston	green	reject	reject	reject	reject	reject	reject	reject
Bates	green	reject	reject	reject	reject	reject	reject	reject
RWD	green	reject	reject	reject	reject	reject	reject	reject
RWD-M	green	reject	reject	accept	reject	reject	reject	reject
Weekly one-Month returns: Sep 2005 - Jan 2018								
Historical	green	reject	reject	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	reject	reject	reject	reject
RND	green	reject	reject	accept	reject	reject	reject	reject
Heston	green	reject	reject	accept	reject	reject	reject	reject
Bates	green	reject	reject	accept	reject	reject	reject	reject
RWD	green	reject	reject	accept	reject	reject	reject	reject
RWD-M	green	reject	reject	accept	reject	reject	reject	reject

Table 8: Goodness-of-fit: Weekly one-week $\text{VaR}_{(0.95)}$ backtests

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Weekly one-week returns: Sep 2005 - Dec 2007 (Pre-Crisis)								
Historical	green	accept	accept	accept	accept	accept	accept	accept
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	accept	accept	accept	accept	accept	accept	accept
Heston	green	accept	accept	accept	accept	accept	accept	accept
Bates	green	accept	accept	accept	accept	accept	accept	accept
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept
Weekly one-week returns: Jan 2008 - Dec 2009 (Crisis)								
Historical	yellow	reject	reject	reject	accept	accept	reject	reject
Historical HW	green	accept	accept	reject	accept	accept	reject	reject
RND	green	accept	accept	accept	accept	accept	accept	accept
Heston	green	accept	accept	accept	accept	accept	accept	accept
Bates	green	accept	accept	accept	accept	accept	accept	accept
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept
Weekly one-week returns: Jan 2010 - Jan 2018 (Post-Crisis)								
Historical	green	reject	reject	accept	reject	accept	reject	accept
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	reject	reject	accept	reject	accept	reject	reject
Heston	green	reject	reject	accept	reject	accept	reject	reject
Bates	green	reject	reject	accept	reject	accept	reject	reject
RWD	green	reject	reject	accept	reject	accept	reject	accept
RWD-M	green	reject	reject	accept	reject	accept	accept	accept
Weekly one-week returns: Sep 2005 - Jan 2018								
Historical	green	accept	accept	accept	accept	accept	reject	reject
Historical HW	green	accept	accept	accept	accept	accept	accept	reject
RND	green	reject	reject	accept	reject	accept	reject	reject
Heston	green	reject	reject	accept	reject	accept	reject	reject
Bates	green	reject	reject	accept	reject	accept	reject	reject
RWD	green	reject	reject	accept	reject	accept	reject	accept
RWD-M	green	reject	reject	accept	reject	accept	accept	accept

Table 9: Goodness-of-fit: Weekly one-week $\mathrm{VaR}_{(0.90)}$ backtests

Method	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
Weekly one-week returns: Sep 2005 - Dec 2007 (Pre-Crisis)								
Historical	green	accept	accept	accept	accept	accept	accept	accept
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	accept	reject	accept	accept	accept	accept	accept
Heston	green	accept	accept	accept	accept	accept	accept	accept
Bates	green	accept	accept	accept	accept	accept	accept	accept
RWD	green	accept	reject	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept
Weekly one-week returns: Jan 2008 - Dec 2009 (Crisis)								
Historical	yellow	reject	reject	reject	reject	reject	reject	reject
Historical HW	green	accept	accept	reject	accept	accept	accept	accept
RND	green	accept	accept	accept	accept	accept	accept	accept
Heston	green	accept	accept	reject	accept	accept	accept	accept
Bates	green	accept	accept	reject	accept	accept	accept	accept
RWD	green	accept	accept	accept	accept	accept	accept	accept
RWD-M	green	accept	accept	accept	accept	accept	accept	accept
Weekly one-week returns: Jan 2010 - Jan 2018 (Post-Crisis)								
Historical	green	accept	accept	accept	accept	accept	reject	reject
Historical HW	green	accept	accept	accept	accept	accept	accept	accept
RND	green	reject	reject	accept	reject	accept	reject	reject
Heston	green	reject	reject	accept	reject	accept	reject	reject
Bates	green	reject	reject	accept	reject	accept	reject	reject
RWD	green	reject	reject	accept	reject	accept	reject	reject
RWD-M	accept	accept	accept	accept	accept	accept	accept	accept
Weekly one-week returns: Sep 2005 - Jan 2018								
Historical	green	accept	accept	accept	reject	reject	reject	reject
Historical HW	green	accept	accept	accept	accept	accept	accept	reject
RND	green	reject	reject	accept	reject	accept	reject	reject
Heston	green	reject	reject	accept	reject	accept	reject	reject
Bates	green	reject	reject	accept	reject	accept	reject	reject
RWD	green	reject	reject	accept	reject	accept	reject	accept
RWD-M	green	reject	reject	accept	accept	accept	accept	accept

References

Bakshi, G., Kapadia, N., and Madan, D. B. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies*, 16(1):101–143.

Bakshi, G. and Madan, D. B. (2000). Spanning and derivative-sequrity valuation. *Journal of Financial Economics*, 55(2):205–238.

Barone-Adesi, G. (2016). VaR and CVaR implied in option prices. *Journal of Risk and Financial Management*, 9(2):1–6.

Basle Committee of Banking Supervision (2011). *Supervisory Framework For The Use of "Backtesting" in Conjunction With The Internal Models Approach to Market Risk Capital Requirements*. Available at www.bis.org.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies*, 9(1):69–107.

Berkowitz, J. (2001). Testing density forecasts, with apllications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.

Bliss, R. R. and Panigirtzoglou, N. (2004). Option-implied risk aversion estimates. *The Journal of Finance*, 59(1):407–446.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies*, 22(11):4463–4492.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651.

Carr, P. and Madan, D. B. (2000). Factor models for option pricing. Unpublished paper.

CBOE (2009). The CBOE Skew index - VIX. Technical report, Chicago Board Options Exchange.

CBOE (2011). The CBOE Skew index - SKEW. Technical report, Chicago Board Options Exchange.

Christoffersen, P., Jacobs, K., and Chang, B. Y. (2013). Forecasting with option-implied information. *Handbook of Economic Forecasting*, 2:581–656.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

Crisóstomo, R. and Couso, L. (2018). Financial density forecasts: A comprehensive comparison of risk-neutral and historical schemes. *Journal of Forecasting*, 37(5):589–603.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuaracy. *Journal of Business & Economic Statistics*, 13:253–263.

Dillschneider, Y. and Maurer, R. (2019). Functional Ross recovery: Theoretical results and empirical tests. *Journal of Economic Dynamics & Control*, 108:103750.

Flint, E. and Maré, E. (2017). Estimating option-implied distributions in illiquid markets and implementing the Ross recovery theorem. *South African Actuarial Journal*, 17(1):1–28.

Haas, M. (2001). New methods in backtesting. Working Paper, Financial Engineering Research Center. URL: www.ime.usp.br/rvicente/risco/haas.pdf.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.

Hollstein, F., Prokopczuk, M., Tharann, B., and Simen, C. W. (2019). Predicting the equity market with option-implied variables. *The European Journal of Finance*, 25(10):937–965.

Hull, J. and White, A. (1998a). Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk*, 1(1):5–19.

Hull, J. and White, A. (1998b). Value at risk when daily changes in markets are not normally distributed. *Journal of Derivatives*, 5(3):9–19.

Jackwerth, J. C. and Menner, M. (2017). Does the Ross recovery theorem work empirically. Available at SSRN: https://ssrn.com/abstract=2960733 or http://dx.doi.org/10.2139/ssrn.2960733.

Kiriu, T. and Hibiki, N. (2019). Estimating forward looking distribution with the Ross recovery theorem. *Journal of the Operations Research Society of Japan*, 62(2):83–107.

Knüppel, M. (2015). Evaluating the calibration of multi-step-ahead density forecasts using raw moments. *Journal of Business & Economic Statistics*, 33(2):270–281.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3(2):73–84.

MATLAB Risk Management Toolbox (2018). MATLAB Risk Management Toolbox R2018b. Retrieved: https://www.mathworks.com/products/risk-management.html.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management: concepts, techniques, and tools*. Princeton University Press.

Pérignon, C. and Smith, D. R. (2010). The level and quality of Value-at-Risk disclosure by commercial banks. *Journal of Banking & Finance*, 34(2):362–377.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.

Ross, S. (2015). The recovery theorem. *The Journal of Finance*, 70(2):615–647.

Sanford, A. (2018). Recovery theorem with a multivariate Markov chain. Available at SSRN: https://ssrn.com/abstract=3247294 or http://dx.doi.org/10.2139/ssrn.3247294.

Shackleton, M. B., Taylor, S. J., and Yu, P. (2010). Multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices. *Journal of Banking & Finance*, 34(11):2678–2693.

Van Appel, V. and Maré, E. (2018). The Ross recovery theorem with a regularised multivariate Markov chain. *ORiON*, 34(2):133–155.