working paper

03 12

The Forecasting Performance of German Stock Option Densities

by Ben R. Craig, Ernst Glatzer,
Joachim G. Keller, and Martin Scheicher

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 03-12 November 2003

The Forecasting Performance of German Stock Option Densities

By Ben Craig, Ernst Glatzer, Joachim Keller and Martin Scheicher

In this paper we will be estimating risk-neutral densities (RND) for the largest euro area stock market (the index of which is the German DAX), reporting their statistical properties, and evaluating their forecasting performance. We have applied an innovative test procedure to a new, rich, and accurate data set. We have two main results. First, we have recorded strong negative skewness in the densities. Second, we find evidence for significant difference between the actual density and the risk-neutral density, leading to the conclusion that market participants were surprised by the extent of both the rise and the fall of the DAX.

Key words: option prices, risk-neutral density, density evaluation, overlapping data
JEL codes: C52, C22, G13, G15

Ben R. Craig is at the Federal Reserve Bank of Cleveland and may be reached at ben.craig@clev.frb.org or (216) 579-2061. Ernst Glatzer is at the Oesterreichische Nationalbank and may be reached at ernst.glatzer@oenb.co.at. Joachim Keller is at the Bundesbank and may be reached at joachim.keller@bundesbank.de. Martin Scheicher is at the Oesterreichische Nationalbank and may be reached at martin.scheicher@oenb.co.at.

1 Introduction

It is a well known fact that a sufficiently rich market in options will imply a state-price density for the underlying instrument. In the event of risk neutrality, these state-price densities correspond to the probabilities that the market assesses to its forecast state. This so-called "risk neutral density" (RND) describes the probabilities that market participants assign to all possible price levels of the underlying instrument. A detailed literature review is given by Jackwerth (1999); some recent studies include Bliss and Panigirtzoglou (2002), Coutant et al. (2001), Galati and Melick (2002), or Jondeau and Rockinger (2000). The entire risk neutral density offers a wide range of information about the price forecast. The volatility, or second moment of the forecast, measures its dispersion. Skewness, or the third moment, measures whether traders in the option markets expect large stock price increases or decreases to occur over the maturity of the corresponding contracts. Kurtosis, the fourth moment, indicates how frequently market participants expect extreme price changes of either sign to occur.

The purpose of our paper is to evaluate the forecasting performance of these option-implied densities for the German stock index, the DAX. We add to the existing literature on option-implied risk-neutral densities in two ways. First, our data concerns the DAX, a market that has not been analyzed before. This market is the largest options market within the euro area, and it is the stock market behind the largest economy in the euro area. Furthermore, our sample is quite comprehensive as it starts in December 1995. The data span the period before and after the boom period of the "new economy", a time of completely different expectations about the future of the economy. The data themselves show this to be a period of great volatility, more so than the data from US stock indices during this period. The data are measured well and provide perhaps the most precise measurements of an important European market during this time.

Secondly, the probability density functions based on option prices are derived under the assumption of risk neutrality. Therefore, there may be differences between the options-based estimates and the actual statistical density of returns. The literature contains many studies on the information content of the (risk-neutral) variance extracted from option prices. As the detailed review by Figlewski (1997) shows, there is valuable information in the implied volatility, which is absent in the variances estimated from the time series of stock returns. However, despite the growing use of the risk neutral density as an indicator of financial market uncertainty (see, for example, Söderlind and Svensson 1997), evidence on the overall informational content of this measure, a fit of goodness for the RNDs, is so far rather sparse. Therefore, we investigate how the realisations of the DAX index deviate from the probabilities estimated by market participants as these are represented by the risk-neutral densities. We are using a new test procedure, so far applied only for the FX market, that is able to detect whether the density forecasts extracted from option prices in a mixture of lognormals specification possess significant errors.

Our main results are as follows. Firstly, we report strong negative skewness in the risk-neutral density which indicates that the probability of a large decrease in stock prices exceeds the probability of a large increase. Our second result concerns the tests that are used to assess the accuracy of the forecasting densities. In line with Ait Sahalia et al. (2001), we document that the risk-neutral density for the S&P 500 deviates significantly from the density of actual outcomes. We, too, can reject the hypothesis of correct densities for most of the period. This is in contrast to Craig and Keller (2002) who report that these densities forecast exchange rates well. Hence, the results so far seem to be rather contradictory, though there is a major difference between the stock exchange and the FX market: Since a negative rational bubble ultimatly leads to a zero price, rational agents know that the bubble will eventually burst and therefore - by backward induction - nobody will pay a "bubble premium" in the first place. However, in the FX market, a positive rational bubble from the perspective of an agent of one currency area is necessarily a negative rational bubble from the perspective of an agent from the other currency area. Therefore, in contrast to the stock exchange, in the FX market the occurrence of a rational bubble can be ruled out.

However, the test begins to accept the same hypothesis toward the end of the period. We conjecture that this is because the market missed long-run tendencies in the market level. This omission was reflected in the poor forecast densities. In some sense, these missed long-run tendencies cancelled each other out, but this took a very long time.

The rest of this paper is organised as follows: The second section discusses the methods used to estimate the implied densities and describes our sample. In section three, we report the estimation results and then test the forecasting performance. Section four contains a summary of our main results and a conclusion.

2 Specification and Sample

When traders price options, they use forecasts of the probability of different asset prices for the period until the contract expires. The perception of market participants about the movement of the asset price, in particular the probability density until expiry, is thus incorporated into the market prices of the derivatives in the process of trading. Therefore, the observed prices of the options convey information about the market operators' assessment of the price process of the underlying instrument, in our case the DAX index.

Among practitioners, the seminal model of Black and Scholes (1973) is commonly used to price options. It assumes that the dynamics of the asset price follow a geometric Brownian motion (GBM). In this case, log returns follow a normal (Gaussian) density. Our specification of the density, the mixture of risk-neutral lognormals, is a generalisation of the Black and Scholes (1973) framework and was introduced by Melick and Thomas (1997).

$$\begin{aligned} \widehat{\pi}_{h,t}(.) &= \theta LogN(a_1, b_1, X_t) + (1-\theta)\, LogN(a_2, b_2, X_t) \\ a_i &= \ln X_t + \left(\mu_i - 0.5\sigma_i^2\right)(h) \\ b_i &= \sigma_i \sqrt{h} \end{aligned} \tag{1}$$

where

$X_t :=$ current price of the underlying asset,

$h = (\Theta - t) :=$ time to maturity of the option,

$\Theta =:$ expiry date of the option,

$\widehat{\pi}_{h,t}(.) =:$ extracted risk-neutral density from observed market prices at t, for realisations h days ahead,

$i =:$ index of state ($i = 1, 2$),

$\theta =:$ weight on the log-normal distributions ($0 \leq \theta \leq 1$),

μ_i, $\sigma_i =:$ mean and variance of the normal distributions,

$a_i, b_i =:$ location and dispersion parameters of the log-normal distributions.

This stochastic process is based on two states with different first and second moments, governed by the weights θ and $1 - \theta$. In each state, the stock price is lognormally distributed. The numerical estimation of the risk-neutral density is based on the nonlinear least squares method. So the squared pricing error, ie, the squared distance between the theoretical and observed prices is minimised with the data set for each trading day. Moreover, to asses to what extent a mixture of two lognormal distributions - instead of a simple lognormal distribution - reduces the squared pricing error, we consider the case of θ being 0 (or is 1).

Our sample consists of the daily trade statistics of the options and futures on the DAX index from 12. December 1995, to 15. May 2002, and comprises 1,613 trading days. Index options on Eurex are defined as European style contracts; hence we do not need to consider the impact of early exercise. Every day maturities of up to two years with at least five strike prices are traded. For contracts with a maturity of up to six months, there are at least nine strikes. The contract value is EUR 5 per index point. The minimum price movement is EUR 0.5. The payment of the option premium is due on the first trading day after the transaction has taken place. Options expire every third Friday in each month.

In order to eliminate the impact of a time-to-maturity effect from our estimates, we require a risk-neutral density with constant maturity. If this effect is neglected, the problem is that parameters change as the expiry date approaches; the volatility decreases with each time increment as the uncertainty about what the asset price will be on the day of maturity is reduced. This effect makes the evaluation of the density forecasting performance impossible, and hence we construct risk neutral densities with constant horizons by means of a linear interpolation.

Our interpolation is based on the put and call prices for all the strike prices sampled each trading day that mature within the next three months. As options expire each month for the next three months, we can always cover horizons of 42, 49 and 56 calendar days. For example, on a given day we construct the 56-day horizon by interpolating linearly between the options which expire in 40 and 70 days. Our choice is determined by the availability of data: We are unable to construct shorter horizons, because the first pillar of our interpolation would have a horizon shorter than two weeks. In this case, the prices become erratic due to the approach of the expiry date. Because the six-month horizon is not covered in every month on account of the quarterly cycle, we cannot construct constant maturity risk-neutral densities with longer horizons. For each trading day and for each horizon, we select those strikes that are available for both the next relevant expiry date and the relevant expiry month after that. We implement a linear interpolation of the Black-Scholes volatilities of these parallel strikes. We choose the volatilities and not the option prices as input and output because the narrower scale of the variance makes the interpolation more tractable (see Clews et al. (2000) for a similar methodology). After the interpolation across the set of parallel strikes, we transform the constant-maturity volatilities back into option prices and estimate the parameters of the mixture model.

To measure the risk-free rate we have used the FIBOR rate before the introduction of the euro and EURIBOR rates thereafter. We match the 42-, 49-, and 56-day horizons of the options with the maturity of the sampled interest rates by means of a linear interpolation.

For sound testing of the fit of the risk-neutral densities, data quality is a key concern. As a first step, we discarded all options with a price below EUR 0.5 and all those where the numerical routine failed to generate an implied volatility. Then, we also discarded all options with moneyness below 0.7 and above 1.25.

Based on the 1,613 daily data sets, we estimate the densities for the three maturity horizons with a rolling-window technique. For the price of the underlying instrument on each day, we take an estimate based on the DAX future. Depending on the state of the maturity cycle, there are three cases: a single future, DAX index and future, or two futures. These values enter as a starting point in the procedure to compute the at-the-money point. This point is obtained as the average of calls and puts for two strikes above and two strikes below the current value using the put/call parity.

2.1 Test procedures

We have applied two test procedures to evaluate the forecasting accuracy of the mixtures of our lognormal densities, where forecasting horizons of h $(= \Theta - t)$ days, with $h \in \{42, 49, 56\}$. Both tests rely on the so-called probability integral transformed realisations $z_{h,t}$. (In other words, each forecasting horizon, h, has a separate value $z_{h,t}$ associated with it). In general, $z_{h,t}$ is equal to the probability value of the estimated cumulative distribution function, $\widehat{\Pi}_{h,t}(.)$, h days ahead at the realization X_{t+h}, where X_{t+h} is the realization of the underlying asset on day $t + h$.

$$z_{h,t} = \int_{-\infty}^{X_{t+h}} \widehat{\pi}_{h,t}(u)\, du = \widehat{\Pi}_{h,t}(X_{t+h}) \qquad , \tag{2}$$

where $\widehat{\Pi}_{h,t}(.)$ is the estimated cumulative density for DAX realisations h days ahead, based on information available at t.

In the special case of the forecasting density being a mixture of lognormals, the integral transform equation equals

$$\begin{aligned} z_{h,t} &= \theta \int_{-\infty}^{\ln X_{t+h}} N(a_1, b_1, u)\, du + (1-\theta) \int_{-\infty}^{\ln X_{t+h}} N(a_2, b_2, u)\, du && (3) \\ &= \theta \int_{-\infty}^{\ln X_{t+h}} N\left(\frac{u-a_1}{b_1}\right) du + (1-\theta) \int_{-\infty}^{\ln X_{t+h}} N\left(\frac{u-a_2}{b_2}\right) du, && (4) \end{aligned}$$

where N is the standard normal distribution. [1]

The first test is based on deviations of the empirical density from the individual quantiles of the outcomes. It consists of testing whether the sequence of integral transforms $\{z_{h,1}, ...z_{h,t}, ...z_{h,T}\}$ is uniformly distributed $(\{z_{h,t}\}_{t=1,...,T} \sim U(0,1)$, with $h \in \{42, 49, 56\}$ and T being the last observation in the sample), as they would be if the true forecasting densities coincide with the estimated densities. This result is

[1] If X_t has a (mixture of) lognormal distribution, then the natural logarithm of that variable, $\ln X_t$, is (mixed) normally distributed.

robust to a time varying-behavior of the estimated densities, as long as the sequence of estimated densities consists of correct densities. This idea goes back to Fischer (1930) and Rosenblatt (1952) and has been recently applied by Diebold et al. (1998), Clements and Smith (2001), Craig and Keller (2002), and de Raaij and Raunig (2002). For a detailed survey about goodness-of-fit methods see Quesenberry (1986). For a graphical representation of the integral transformation see Craig and Keller (2002).

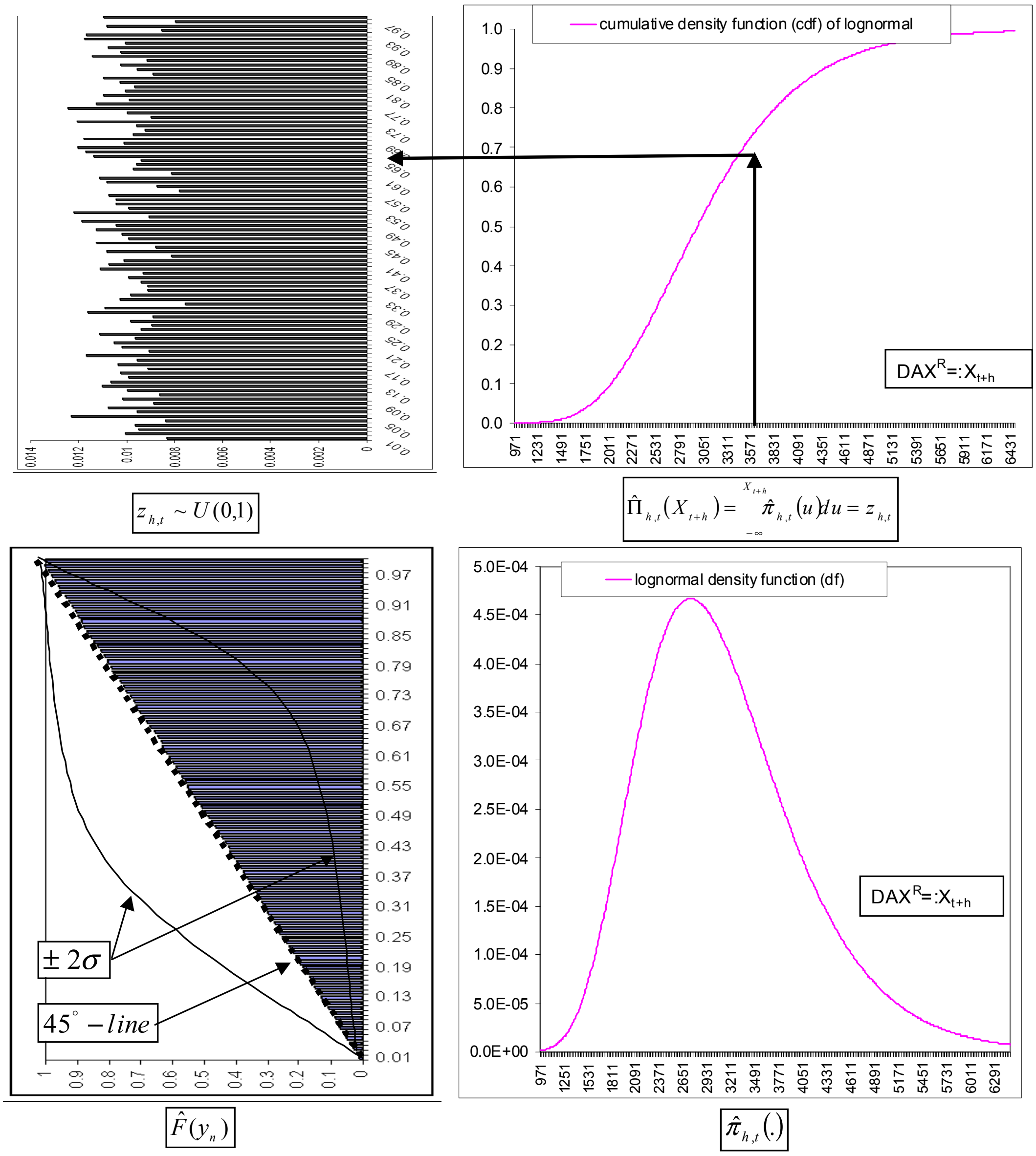
Figure 1 depicts the density function (*df*) and the cumulative density function (*cdf*) of a $U(0,1)$ distributed random variable. Under the null, a sequence of correct densities, and therefore of correctly transformed realizations, $\{z_{h,t}\}_{t=1,\ldots,T}$, can be represented by the *cdf* of the $U(0,1)$, which is equal to $F(y) = y$, for $0 \leq y \leq 1$ and yields the proportion of $z_{h,t}$'s that is less than y. Graphically, the null hypothesis corresponds to the 45 degree line that connects the origin of the upper graph with the upper right corner.

The empirical cumulative distribution function (*ecdf*), $\widehat{F}(y_n)$, is equal to the actual proportion of z_t's resulting from the estimated densities, which is less than y_n and is calculated (for different values of y_n in the interval $(0,1)$) as

$$\widehat{F}_h(y_n) \equiv \frac{\sum_{t=1}^{T} I(z_{h,t} \leq y_n)}{T}, \tag{5}$$

where $I(z_{h,t} \leq y_n)$ is the indicator function, equal to 1 if true and 0 otherwise. For simplifying notation we drop the subscript h of $\widehat{F}_h(y_n)$. If there are N separate values of y_n, the empirical proportion $\widehat{F}(y_n)$ can be represented graphically by the n^{th} bin of a total of N bins. If $\widehat{F}(y_n) \neq y_n$ the *ecdf* deviates from uniformity which violates the null hypothesis. To test whether the distance $\widehat{F}(y_n) - y_n \neq 0$ is significant at an individual y_n, we need to estimate the standard deviation of $\widehat{F}(y_n)$. Note however, that in the case where the sampling frequency (in our case daily) is higher than the forecasting horizon ($h \in \{42, 49, 56\}$), the forecast errors can overlap by up to h days and then follow a moving average process of order h. Therefore the confidence intervals for individual $\widehat{F}(y_n)$ have to be calculated by taking into account the time dependency up to the theoretical forecasting error overlap of h. The variance of an individual $\widehat{F}(y_n)$ equals:

Figure 1: The integral transformation



$$\widehat{var}(\widehat{F}(y_n)) = \widehat{\sigma}^2(\widehat{F}(y_n)) = \frac{1}{T}\left[\widehat{\gamma}^n(0) + 2\sum_{j=1}^{h}\left(1-\frac{j}{T}\right)\widehat{\gamma}^n(j)\right], \quad (6)$$

where the sample autocovariance of order j is defined by

$$\widehat{\gamma}^n(j) = \frac{1}{T}\sum_{t=j+1}^{T}\left(I^{n,t} - \overline{I^n}\right)\left(I^{n,t-j} - \overline{I^n}\right). \quad (7)$$

Under the null hypothesis, the ratio $t = (y_n - \widehat{F}(y_n))/\widehat{\sigma}(\widehat{F}(y_n))$ has a t distribution with $N-2$ degrees of freedom. These tests have the advantage of indicating where the outcome density is missing in a systematic way in the estimated forecasting density during the time period under review. For example, if the option implied densities had thicker tails in general than the "correct densities ", it would show up graphically as $\widehat{F}(y_n) < y_n$ for values of y_n close to 0 or 1 and as $\widehat{F}(y_n) > y_n$ for values close to 0.5. However, this test fails to account for the departures of $\widehat{F}(y_n)$ from y_n jointly for all N bins and is therefore not very powerful.

As a remedy for this shortcoming we propose a second, more powerful test based on the Cramer-von Mises statistic, which is defined as the squared distance of the *ecdf*, $\widehat{F}(y_n)$, from the null hypothesis and which is represented by the theoretical value $F(y) = y$.

$$\widehat{CvM} \equiv \int_0^1 (y_n - \widehat{F}(y_n))^2 d(y_n). \quad (8)$$

Unfortunately, this test relies on independence of the data, a condition which is clearly violated by the overlapping-forecasting error embedded in the $z_{h,t}$'s. To derive a test statistic for $\widehat{CvM}$ in the presence of data dependency we use the stationary bootstrap approach, suggested by Politis and Romano (1994), that is based on a resampling procedure, which uses samples of blocks of random lengths (for details see Craig and Keller (2002). To assess whether the distance from the 45 degree line is significant we calculate the bootstrapped distribution functions, $F_b(y_n)$, for each bootstrapped sample. One observation of the CvM_b statistic is calculated as

$$CvM_b \equiv \int_0^1 (y_n - F_b(y_n))^2 d(y_n). \quad (9)$$

We computed CvM_b for 100,000 replications and report a number, CvM_b, which is the proportion of bootstrapped distances, CvM_b, that are greater than $\widehat{CvM}$, the distance between our sample distribution function and the null, the uniform distribution function. Because the sample distribution function, $\widehat{CvM}$, and all bootstrapped sample distribution functions, CvM_b, are step functions, the sum expression in CvM_b is calculated directly. A value of CvM_b less than some critical value, α_0, rejects the hypothesis of $z_{h,1}, ...z_{h,t}, ..., z_{h,T}$ being drawn from a uniform distribution at the α_0 level.

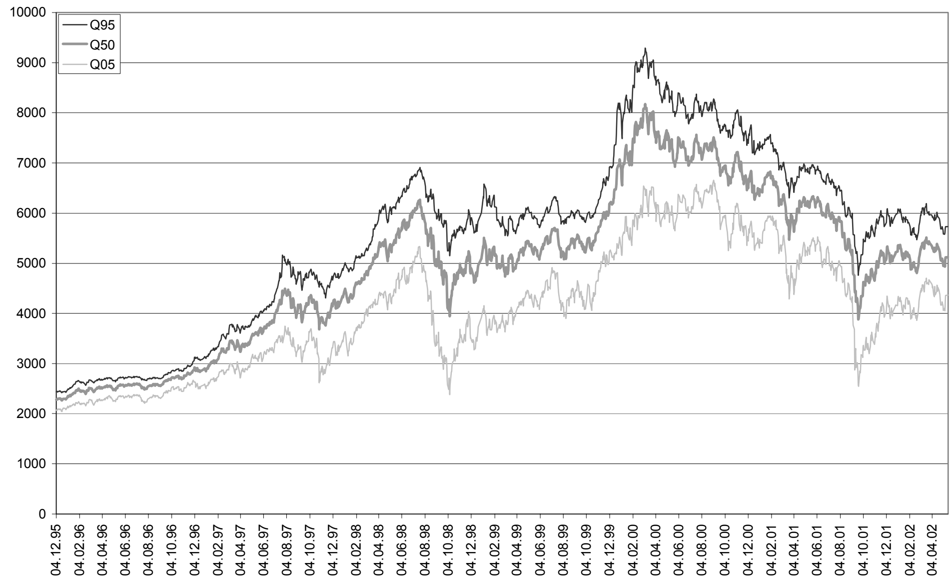
3 Empirical results

First we discuss the estimated risk-neutral densities, and then we study their performance. Figure 2 summarizes the information from the daily density estimates by plotting the median together with the 5 percent and 95 percent estimated percentiles for the 49-day horizon. Results for the 42 - and 56-day horizons are similar and have been omitted for reasons of space. During our sample period, the median of the estimated risk-neutral densities moved between 2,000 and 8,000 points, and the allocation of the probability mass between the centre and the tails changed. In 1996, the distance between the 5 percent and 95 percent percentiles was comparatively small so that the densities had more probability mass around the median. From December 1997 onwards, the width of the interval increased, corresponding to rising uncertainty about the future behavior of the German stock market. In the first half of 2000, the bands widened as market operators were expecting an ever larger range of possible values for the DAX index.

During the tumultuous periods of autumn 1998 and autumn 2001, the distance between the two tails of the distributions increased. Figure 2 also shows that the distance between the median and the upper and lower percentiles varies over time, and in some periods the lower percentile is more distant from the median. This asymmetry reflects the fact that the left and right tails of the risk-neutral distributions do not contain the same amount of probability mass, ie, the risk-neutral perception about directional moves differs according to the sign of the stock market move.

Figure 3 depicts the estimated risk-neutral densities in yearly intervals from May 1996 until the last estimate in May 2002. The graph clearly reflects the up and

Figure 2: Median and 5%, 95% RND confidence bands for mixture of lognormals



down movements of the German stock market. The shift in the allocation of the relative probability mass manifests itself quite clearly in the estimated densities. Because the probability has moved from the centre towards the tails, the densities have become flatter. This shift implies that a wider range of index values is now considered likely. Furthermore, the upward and subsequent downward movement in the probability of given index values is visible. The primary cause for this movement is the general fall in the value of stocks contained in the DAX index since the end of "irrational exuberance" in spring 2000. At that time, the right tail was above 9,000 points, and the probability of an index at 4,000 points was almost zero. In May 2002, the right tail was situated at 6,500 and the left fail at 3,500. Thus, the risk-neutral densities estimated for May 1999 and May 2002 are remarkably close, because within these three years, the DAX had lost all of its gains.

This view is corroborated by Figure 4 which depicts the same densities as above, now however centred at the (annualised) expected percentage change of the DAX index value. The mean of the percentage changes, the risk neutral discount factor, is close to 3%. This depiction facilitates the comparision of the densities, since the "width" of the densities does not depend upon the level of the index value.

Figure 3: Yearly RNDs

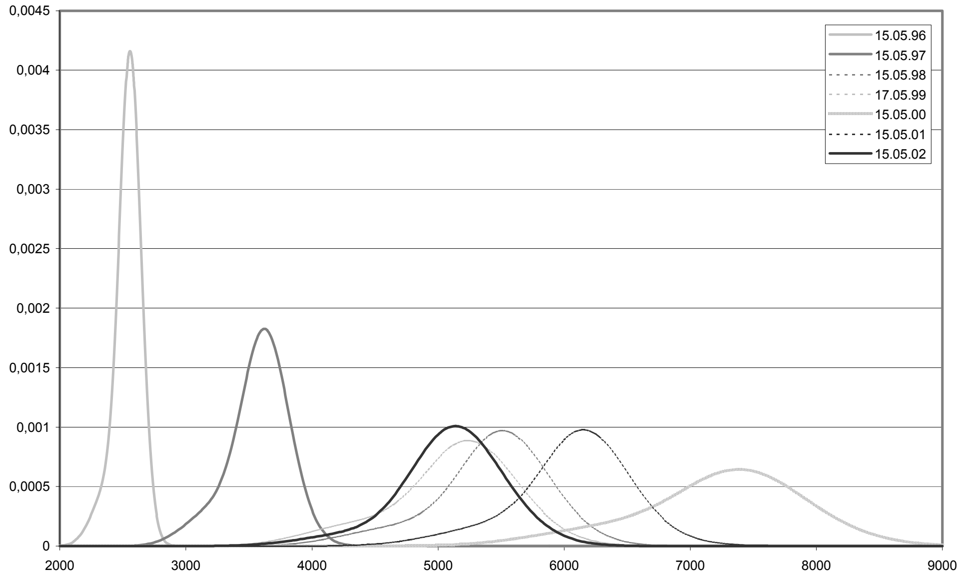
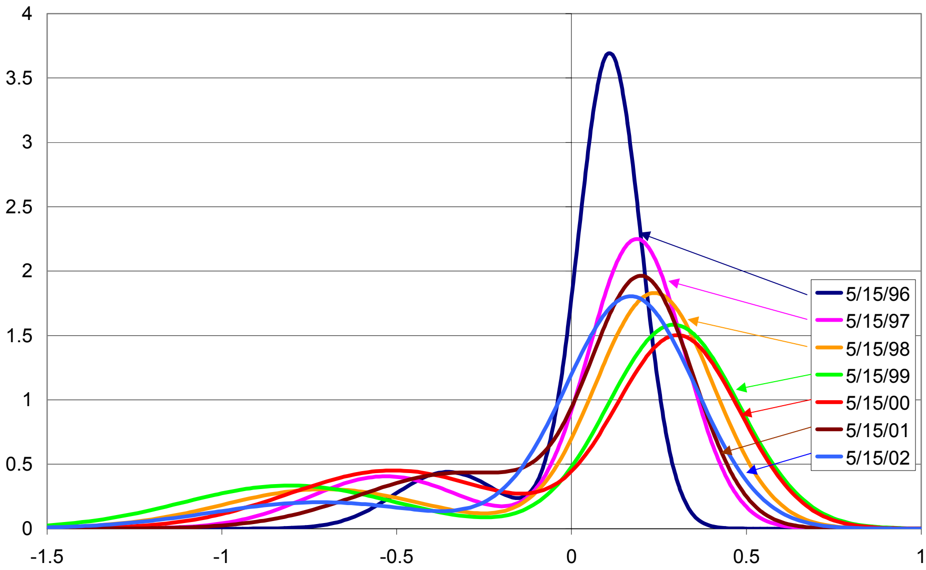


Figure 4: Yearly RNDs (percentage changes)



Another perspective on the information contained in the estimated density is given by the risk-neutral probability of certain percentage changes in the DAX index. A rise of the probability mass located in the lower tail of the density indicates that market participants expect the German stock market to become more prone to crashing. Figure 5 plots the level of the DAX index, the implied volatility (the VDAX)[2], and the risk-neutral probability of a fall of at least 10 percent. The Figure indicates a similar development of the probability and the volatility. For both series, the peaks coincide in 1998 and in 2001. In both measures, the highest level is recorded in October 1998, where the risk-neutral probability climbed to 32 percent and the volatility to 56 percent. Inspection of Figure 5 reveals that the downward movement on the German stock market since the first half of 2000 did not coincide with a permanent increase in the probability of further declines. The most recent episode, which is visible in Figure 5, concerns the increased uncertainty in the financial markets after the discovery of major accounting discrepancies at several large US firms, as was the case, for example, with the Enron Corp which filed voluntary petitions for Chapter 11 reorganisation in early December 2001. The discoveries led to growing doubts about the reliability of the information that investors receive about companies, and so we observe a clear rise both in the tail statistic and the implied volatility.

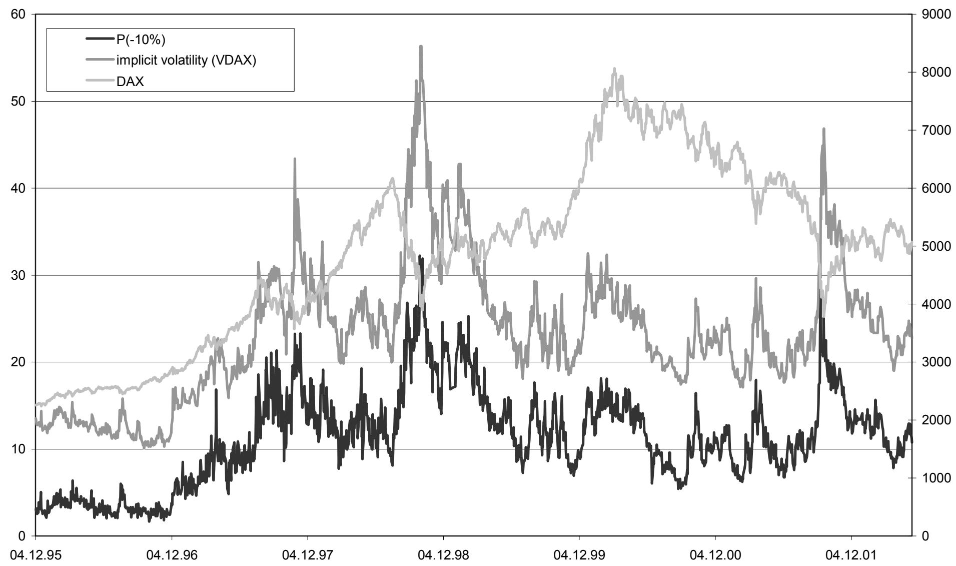
Of particular interest is the information contained in the moments of the risk neutral densities, summarised in Table 1. (row= moment of densities; column= descriptive statistics in time (t)). We measure skewness with the Pearson statistic, ie, $(mean - mode)/standard\ deviation$, because the commonly used skewness coefficient, the central third moment divided by the standard deviation, is very sensitive to the mass in the tails.

Table 1: Descriptive statistics of estimated moments

	Mean (in t)	Std (in t)	Med (in t)	Max (in t)	Min (in t)
Mean	4912.592	1506.317	5079.224	8089.984	2253.081
Std	447.8361	197.15	476.15	929.3824	100.1263
Skewness	-0.182735	0.068699	-0.178581	-0.011739	-0.560282
Kurtosis	0.882496	0.625993	0.794447	9.145827	-0.44906

[2] The VDAX is the DAX based volatility index and expresses the implied volatility of the DAX index, as expected by the forward market.

Figure 5: RNDs estimates for 10% tail statistics, VDAX (implied volatility) and DAX



We observe a pronounced negative asymmetry, so that the left tail is larger than the right. For the US stock market this observation has been documented by Jackwerth and Rubinstein (1996). The economic rationale behind the observation is that put options are used as hedging instruments to protect against large downward movements in stock prices. This demand by investors due to portfolio insurance strategies has increased the price of protection and therefore the left tail of the density receives more weight. In the next section, we will test whether the negative skewness in the risk-neutral density is compatible with the information in the actual (statistical) density.

3.1 Forecasting Performance

Our second test, based on the von Mises statistic, clearly does not reject the null when the entire sample is chosen. The results can be seen in Table 2. This is true over all forecast horizons and for both the lognormal ($\theta = 0$) and the mixture of lognormals (except in the case of the mixture of lognormals with a forecasting horizon of seven weeks, and even here the results only barely reject the null at the 5 percent level). Therefore, the tests seem to indicate that the option-based densities are good forecast densities.

Table 2: Von Mises statistic: p-values

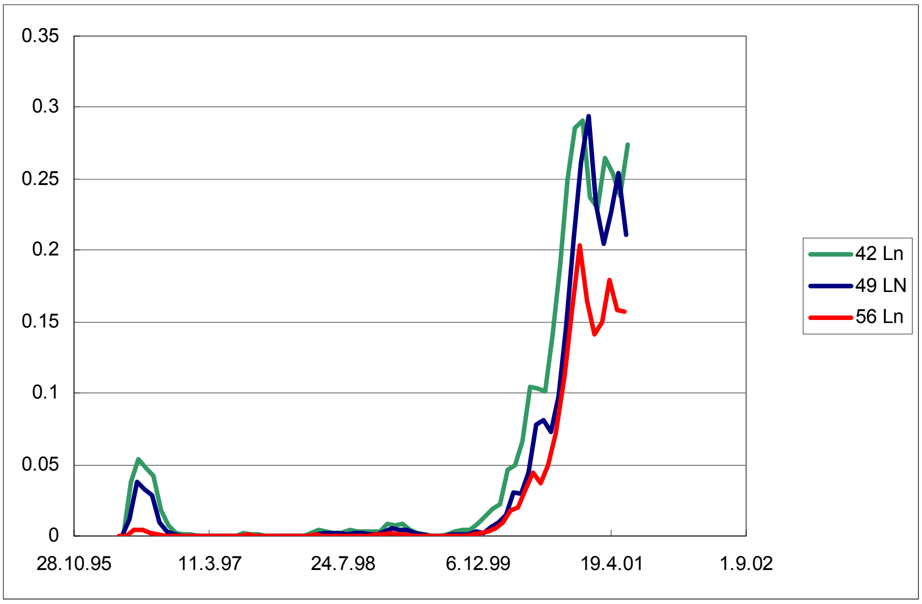
Horizon h in days*	42	49	56
Mixture of Lognormals			
Dec 1995 to Apr 2002	0.048	**0.062**	**0.056**
Dec 1995 to Mar 2000	0.000	0.000	0.000
Apr 2000 to Apr 2002	0.012	0.007	0.003
Lognormal			
Dec 1995 to Apr 2002	**0.263**	**0.243**	**0.172**
Dec 1995 to Mar 2000	0.009	0.009	0.001
Apr 2000 to Apr 2002	0.006	0.004	0.001

*Bold numbers indicate that the hypothesis of an accurate density can't be rejected.

The outcome is more complicated when we examine sub-periods of our sample period. When choosing sub-periods, we are careful to avoid choosing on the basis of outcomes, for this would induce a sample selection bias that would invalidate our results. For example, one should not choose to split the sample into one sub-sample that contained only those days where the prices increased and another in which prices decreased. This would slant the test against the null of the risk-neutral density being the same as the forecasting density. In the same sense one could not split the sample on the basis of where the prices were rising, generally, and then falling.

Yet, it is clear that a problem with the bootstrapped von Mises test lies in its lack of a test for the time series properties of the data. If our observations were independent, we could test the independence of the integral transforms $\{z_{h,1}, ...z_{h,t}, ...z_{h,T}\}$ (which would then be a series of independent uniform $U(0,1)$ random variables). However, it is our overlapping forecasting windows that cause the lack of independence and led us to use the bootstrapping approach in the first place. A test built

Figure 6: p-values for the RND (lognormal)

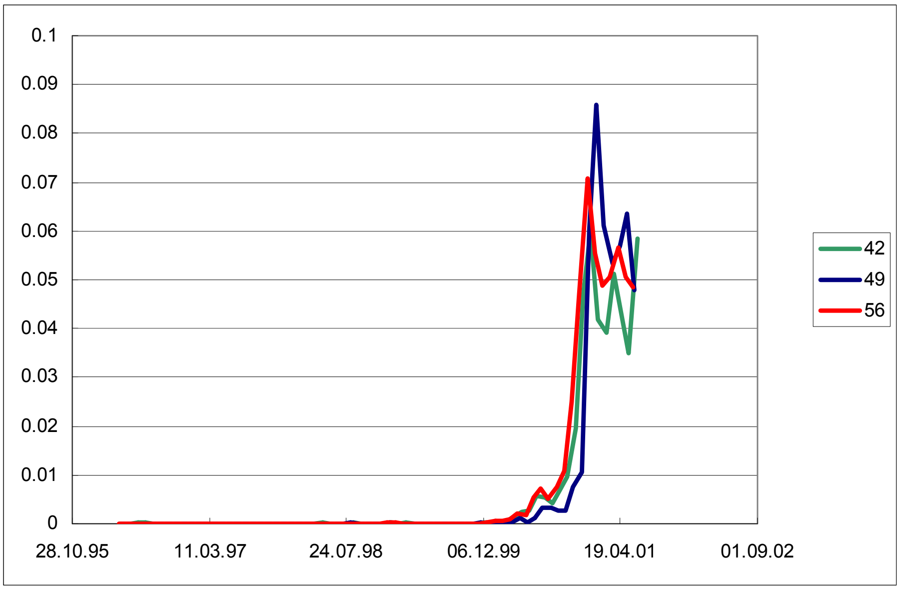


around just the non overlapping windows would give such a small number of observations that the test would have very poor power.

We chose to look at the sample as if the researcher were continually monitoring the forecasting performance of the risk-neutral densities from 100 days after the beginning of the time period to the end of our sample. In other words, we tested the forecasting properties of the densities 100 trading days after the first trading day of our sample, 120 days, and so forth right to the end of the sample period of more than 3,000 trading days. We computed p-values for each of the forecasting horizons and for both the lognormal and mixture of lognormal models. The results are graphed in Figures 6 and 7.

For the lognormal (Figure 6), the first four and a half years of the sample clearly reject the null of a good forecasting density. Suddenly, in May 2000 (for the 6-week horizon), July 2000 (7-week) and September 2000 (8-week), densities for all of the forecasting horizons are accepted as good forecasts by our tests. Similarly, the mixture of lognormals (Figure 7) is even more decisively rejected until the very beginning of 2001, where the test does not reject or only weakly rejects the densities for all forecast horizons. What is happening?

Figure 7: p-values for the RND (mixture of lognormals)



To better understand what may be at work, we divide the sample into the two periods, not to test the forecasting density, so much as to examine the failure of the densities as the market rose and to explore what happened as the market fell. The graphical results of the deviations of the empirical density from individual quantiles are shown in Figures 8 - 16. The nine Figures are organised in 3 blocks (Figures 8 to 10, 11 to 13 and 14 to 16), each of which depicts the results for different forecasting horizons (42; 49 and 56 days). Each individual Figure within a block shows the result for a different sample period. The first period covers the euphoric stock market period from December 1995 to March 2000, the second shows the period during which the bubble "burst" from April 2000 to March 2002, and the third spans the entire range of the sample. In each of the nine Figures we plot the actual *ecdf* of the integral-transformed realizations, $z_{h,t}$, against the theoretical *cdf*, which is represented by the dashed 45 degree line. The vertical axis represents the cumulative probabilities and the horizontal axis represents the bin number n, where n lies between 1 and $\overline{N}$, and here has been set at 200. Therefore, the n^{th} bar in the Figure represents the sum of the observed $z_{h,t}$'s (i.e. $\widehat{F}(y_n)$'s) that are equal to or less than $n/\overline{N}$. Under the null hypothesis the number of $z_{h,t}$'s in bin n is always

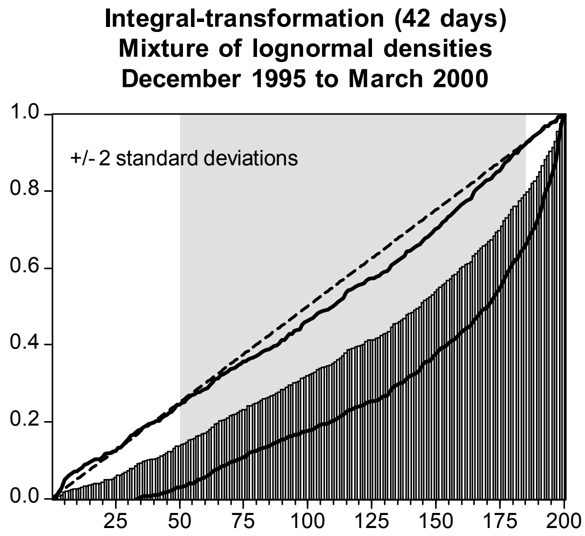


Figure 8

equal to $n/\overline{N}$. To assess whether deviations of the theoretical value are significant, a 95 percent confidence interval $\left(\pm 2\widehat{\sigma}^2(.)\right)$ is calculated according to equation (6) and drawn in the Figure (bold lines). The confidence intervals of these estimates narrow down relatively slowly because of the high moving-average time dependence (due to the overlapping data problem mentioned earlier) of the observations. The standard deviations are calculated around the estimated parameter $\widehat{F}(y_n)$, and the null hypothesis of correct forecasting densities are rejected if the corresponding point on the 45 degree line lies outside the lines.

In Figures 8 to 16 the areas of rejection of the hypothesis of correct densities are greyed out. The results vary highly in different sample periods. The uniform *cdf*'s corresponding to the *ecdf*'s of the first sub-sample (first column) are significantly below the uniform line for most of the bins and rise above the uniform line only at the end of the right side, indicating that the densities fail to capture the right frequency of positive returns during this time. Over the boom years market participants seemed to have had too pessimistic a view on the stockmarket, and subsequent positive realisations came as a surprise. This result holds regardless of the forecasting horizons of our densities.

Exactly the reverse is the case for the second sub-sample (second column). Here

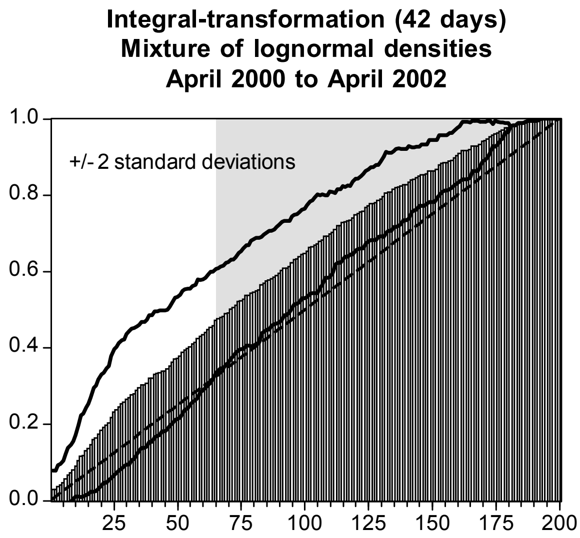


Figure 9

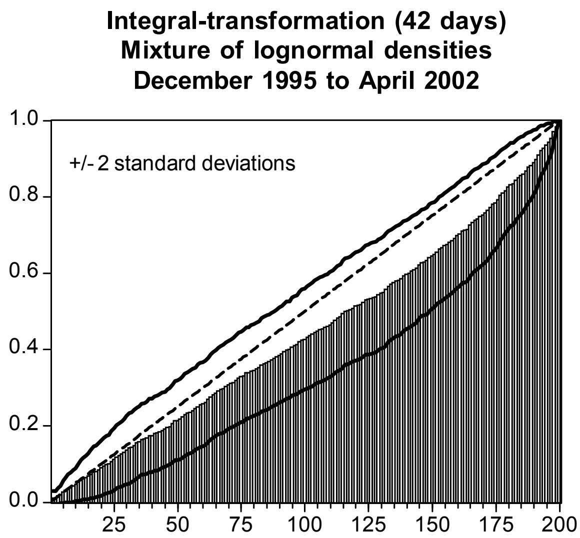


Figure 10

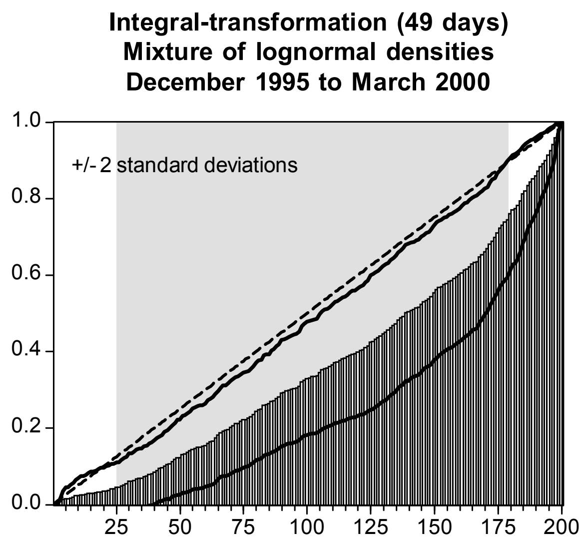


Figure 11

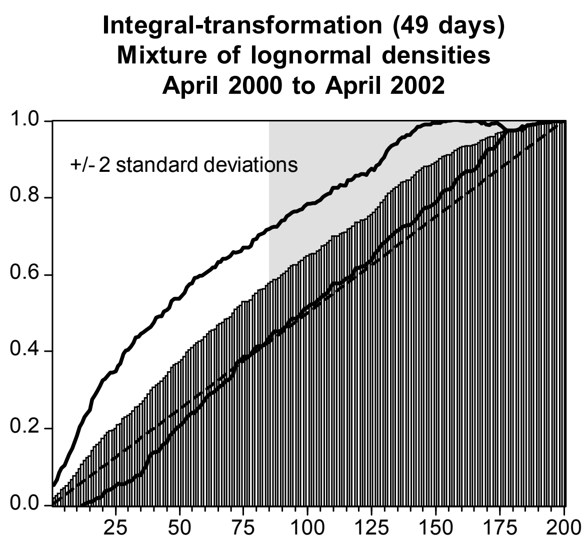


Figure 12

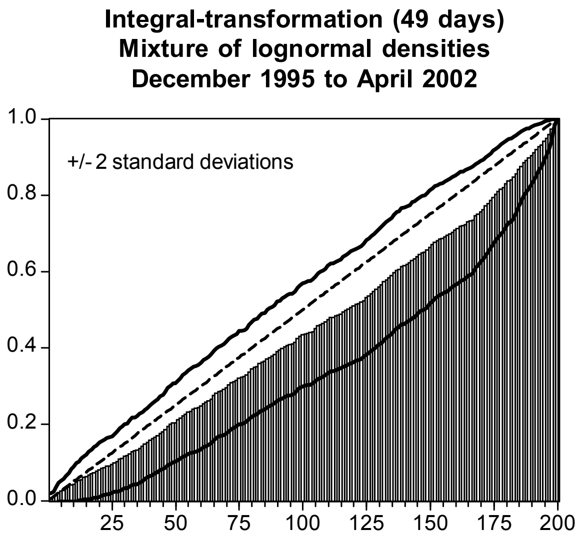


Figure 13

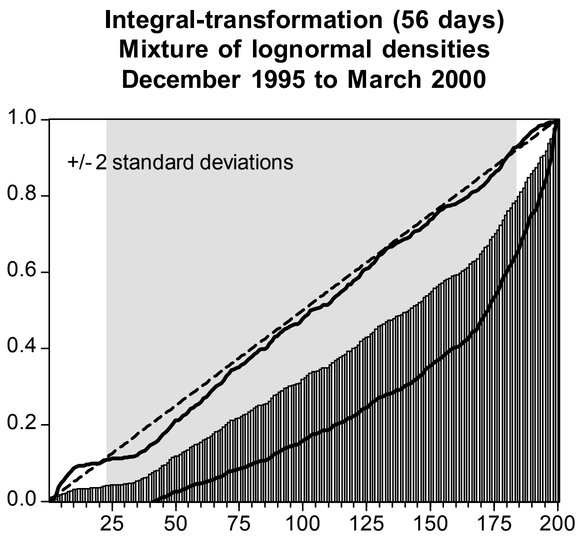


Figure 14

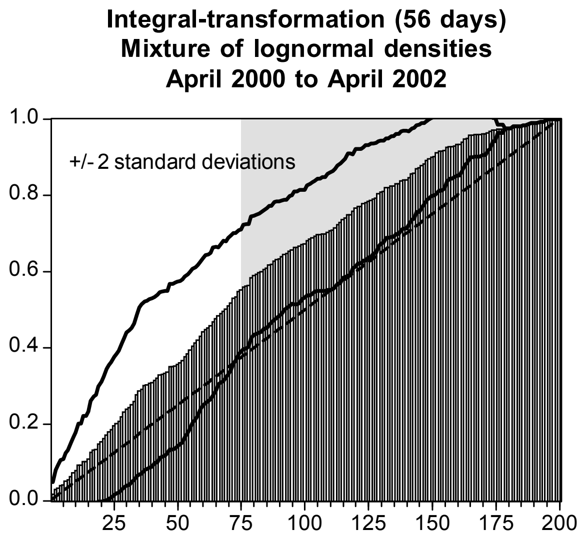


Figure 15

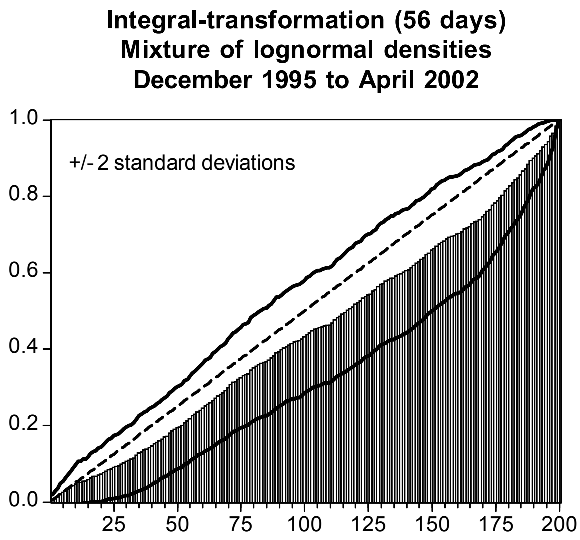


Figure 16

the uniform *cdf*'s are significantly above the uniform line at the extreme left side, but as one moves to the right, drop below it and stay there. Thus, for the second sub-sample, the lognormal forecast densities - again independent of the forecasting horizon - do not place enough probability mass at the left side. When the density predicts an outcome that is less likely than 12 percent, the actual realisations are so rare that we are not able to calculate standard deviations around these observations. In this period, agents were obviously surprised by the extent on the downturn of the German stock market. People systematically missed the trend of the market and made enduring forecasting errors, an observation challenging the generally accepted view of efficient markets.

When the entire sample is taken into account-/ borne in mind, things are different. We are not able to reject the hypothesis of correct forecasting densities for this period. Since the integral transformed realisations are not sorted chronologically, the errors of sub-samples 1 and 2 may cancel each other out. This is because the test statistic is based on the relative frequency of the realisations of $z_{h,t}$'s in specific bins. In sub-sample 1, the errors implied by the estimated density are reflected in an overproportionally high frequency of high $z_{h,t}$ realisations, whereas in sub-sample 2 the opposite is true. If the two sub-samples are merged, the relative frequency on both sides are declines, leading spuriously to a situation in which the null cannot be rejected. In the case of the whole sample (column 3), the line of bins still runs under the 45 degree line because the number of observations in sub-sample 1 with high $z_{h,t}$ realisations is higher than the number of observations in sub-sample 2 with low $z_{h,t}$ realisations. Having stated that, it is clear that under the null it should not be possible to reject the hypothesis of correct densities in any sub-sample, so that a test based on splitting the samples seems to be appropriate.

Note that the bootstrapped Cramer von Mises statistics, (CvM_b), (Table 2) give much the same intuition. We overwhelmingly reject the hypothesis of correct densities for both sub-samples, independent of the type of density we use (log-normal, mixtures of lognormals). But again, we have problems doing so on the basis of all the data with exception of the 42-day horizon forecast, which we reject at the 5 percent level.

It should also be noted that our results have little to do with the peso problem on the foreign exchange market or so-called small-sample bias. The small-sample bias hypothesis was originally put forward by Rogoff (1980) who argued that the finding

- based on t-statistics - that log forward exchange rates are biased predictors of the future log spot exchange rates is spurious. He based his assertions on the observation that convergence of the distribution of the slope parameters to normality is very slow, due to the fact that distributions of asset price changes often show non-normal features (such as fat tails and bi-modality). In this context, inference from t-test statistics lead the researcher to reject the hypothesis of unbiasedness too often. In other words, if we knew the small sample properties of the density of the estimated parameter, we would not be able to reject the null.

The case is different for our estimated densities. Our method explicitly takes into account the fact that asset prices are no normally distributed and assigns specific probabilities to the occurrence of rare events. The approach of a mixture of lognormal densities appears in particular to be appropriate for modelling discrete outcomes such as regime shifts or sudden changes in a trend. Under the null of the sequence of mixed lognormal densities being the true densities, we do not underestimate the probability of a type-I error, as one might in the case cited by Rogoff. Therefore the t-statistics of the individual $\widehat{F}(y_n)$ give a nonbiased indication of whether the densities fail and if so, where.

What seems to be happening, then, is that when the options market failed to provide good forecasting densities, it did so in a way that was entirely symmetrical. When it missed the dramatic rise early in the sample period, it placed too much of the weight of the density on the low side; when it missed the fall, it placed too much on the high side. That the market undershot the rapid rise and fall of the DAX is obvious, but what is interesting is the symmetry of the failure. When both densities are added together, their failures in a sense cancel each other out. The option density seems like a good forecast density only when seen over the entire sample period.

3.2 Trend adjusted tests

This suggests one further test to see if the difference in trend is the sole difficulty driving the Von Mises rejection of the option based risk neutral density as a forecast of future levels of the DAX. We check whether the de-trended future DAX is consistent with the de trended forecast implied by the options. In other words, for the two periods, the first up to 8. March 2000 where prices were generally rising rapidly, and the second from 8. March onwards where prices were falling we de-trend the actual outcomes of the DAX. We then compare these two series with the two series of forecast densities obtained from the options, with the density re-centered about zero by a simple translation of the density obtained by subtraction of the mean from all values of the density. Thus the series are corrected for any differences in means, being both normalised to a mean of zero. A typical result is displayed in Table 3.

Table 3: Von-Mises statistic: p-values
Trend adjusted densities

Horizon h in days*	42	49	56
Mixture of Lognormals			
Dec. 1995 to Mar. 2000	0.0055	0.0093	0.0191
Apr. 2000 to Apr. 2002	**0.5828**	**0.1988**	**0.1001**

*Bold numbers indicate that the hypothesis of an accurate density can't be rejected.

It is clear from the table that such an experiment has only limited Success at best. For the great run-up in stock values prior to April 2000, the densities do a poor job of forecasting, even when adjusted for the differences in expectation. However, after this period, during the crash, trend adjusting delivers densities that are quite consistent with the DAX outcomes. The main message of Table 3, however, is that a simple adjustment was not enough to repair the forecast densities implied by options during the heady period of increasing stock values. The densities were off in more ways than in just the mean.

4 Concluding remarks

This paper has analysed a method for measuring the market perception of the uncertainty about the future dynamics in the German stock market. Our sample extended from December 1995 to May 2002. We evaluated a mixture of lognormals. After discussing the estimation results, we described our tests of the performance of the DAX risk-neutral densities.

The tests revealed several things. Firstly, for the largest part of the sample period, the data strongly reject the hypothesis of options price densities being correct forecasting densities at the 6-week, 7-week and 8-week forecasting horizon. Secondly, the main reason for the rejection seems to be the future prices (the median of the densities) around which the densities are centred, since the sequence of future prices consistently miss the temporary, stochastic trends prevailing in the data. This was just as true during the upturn as during the downturn of the stock market. In this symmetry, the experience of the DAX options has proved to be special. Densities calculated for the US markets from options consistently mismeasure the outcomes on the downside. We found that the DAX densities mismeasured on the downside during the run-up in stock prices, but that in the subsequent period the densities consistently over-predicted the stock prices. Thirdly, since the future price, independent of the specific method applied (i.e., mixtures of lognormals, Shimko interpolation etc), is always the "centre" of the density, the rejection is not restricted to a mixture of lognormals approach. Fourth, if the data are purged of the trend (or first moment error), it turns out this was not sufficient to fix the densities during the period of rapidly rising prices, but during the price falls, higher moments are matched quite well. The densities are capturing the standard deviation (general level of uncertainty), the skewness (asymmetric up - and downward risks) and the kurtosis (extreme events) of the underlying stochastic process quite accurately. Interestingly enough, this was true only during the period where the DAX options over-predicted market performance, a phenomenon that is not typical of the experience of densities calculated from options on the S&P 500 index. Finally, when observed over the entire seven year period of our data, options densities did a good job of predicting the density of outcomes. Options provided poor densities when measured with respect to sub-periods within our data.

A natural question arising in this context is how to interpret the results with respect to the rationality of market participants. Given that the stock price process, possibly corrected by a drift rate that reflects the growth path of the underlying economy, is a martingale, it is not surprising that the futures process follows a martingale, too. This is because basic "no arbitrage" principles make the futures rate adjust so that it equals the spot price multiplied by the risk-free interest rate. The martingale hypothesis is empirically corroborated by the fact that it is impossible to reject the hypothesis of a unit root. Therefore, it makes perfect sense for the market participants to centre their densities around the futures prices.

On the other hand, stock markets around the world, besides Japan, have soared by historical standards to extreme high levels until recently. In contrast to this development, basic economic indicators, the fundamentals, have not soared to the same extent. Behind this backdrop, asset price increases seemed unwarranted given historical experience. However relatively few openly denounced this phenomenon as irrational. If the rational analysis of fundamentals does not justify the surge in stock prices, then what were the driving factors behind the speculative bubble? One explanation that recently gained a lot of support is that market participants do not always behave completely rationally but have bounded rationality, at least for longer horizons. However, one result of the bounded rationality literature of Hansen and Sargent (2003) is that for a wide variety of classes of bounded rationality, the effect on the market is to pose a simple mean shift in the pricing of risk. Our results are inconsistent with this: even adjusting for differences in means do not account for the failure in the option derived densities to forecast future prices. Something seriously more complicated is at work here.

In this sense the degree of rationality is linked to the planning horizons of the investors. When one looks at the performance of the option based densities over the entire eight year period, the densities perform remarkably well. Yet much of the literature about the recent bubbles have predicted the opposite conclusion. Whereas investment managers and the general public have a rational view of stock market prices in the short-term, they may be prone to make longer-term investment decisions that are driven by self-fulfilling, possibly irrational factors, such as exuberance, changes in attitudes towards savings, growing equity culture etc (for a survey, see RJ Shiller (2000)). If this is the case, market participants may accept that in a boom period, short-term expectations are consistently disappointed. It is an interesting, yet open, question as to how our empirical results from these two horizons could be reconciled by an agent-based model. Future research could point to this direction.

5 Bibliography

AIT-SHALIA, Y AND Y WANG AND F YARED (2001), "Do option markets correctly price the probabilities of movement of the underlying asset?", *Journal of Econometrics*, Vol 102, pp67–110.

BLACK, F AND W SCHOLES (1973), "The pricing of options and corporate liabilities", *Journal of Political Economy*, Vol 61, pp637-59.

BLISS, R AND N PANIGIRTZOGLOU (2002), " Testing the stability of implied probability density functions", *Journal of Banking & Finance*, Vol 26, pp381-422.

CLEWS, R AND N PANIGIRTZOGLOU AND J PROUDMAN (2000), "Recent developments in extracting information from options markets", Bank of England, *Quarterly Bulletin*, February 2000, pp50 – 60.

CLEMENTS, M P AND J SMITH [2001], "Evaluating Forecasts from SETAR Models of Exchange Rates", *Journal of International Money and Finance*, Vol 20, pp133-148.

COUTANT, S AND E JONDEAU AND M ROCKINGER (2001), "Reading PIBOR futures options smiles, The 1997 snap election", *Journal of Banking & Finance*, Vol 25, pp1957-1987.

CRAIG, B R AND J G KELLER (2002), "The Empirical Performance of Option Based Densities of Foreign Exchange", Deutsche Bundesbank *Discussion Paper* , No 07/02.

DE RAAIJ, G AND B RAUNIG (2002), "Evaluating Density Forecasts with an Application to Stock Market Returns ", Oesterreichische Nationalbank *Working Paper*, No 59.

DIEBOLD, F X AND T A GUNTHER AND A S TAY [1998], "Evaluating Density Forecasts with Applications to Financial Risk Management", *International Economic Review*, Vol 39, Issue 4, pp863-883.

FISCHER, R A [1930], "Inverse Probability", Proceedings of the Cambridge Philosophical Society, Vol 36, Issue 4, 528 - 535.

GALATI, G AND W MELICK (2002), "Central bank intervention and market expectations", BIS Papers No 10.

HANSEN, L P AND SARGENT, T (2003), "Robust Learning: A Quartett of Semi-Groups: Bounded Rationality and Risk", Unpublished mimeo.

HULL, J (2000), "Options, Futures, and other Derivatives ", fourth ed, Prentice

Hall.

JACKWERTH, J AND M RUBINSTEIN (1996), "Recovering Probability Distributions from Option Prices ", *Journal of Finance*, Vol 51, pp1611-1631.

JACKWERTH, J (1999), "Option-implied risk neutral distributions and implied binomial trees: A literature review ", *Journal of Derivatives*, pp66-92.

JONDEAU, E AND M ROCKINGER (2000), "Reading the smile: the message conveyed by methods which infer risk-neutral densities ", *Journal of International Money And Finance*, Vol 19, pp885-915.

LEHAR, A, M SCHEICHER AND C SCHITTENKOPF (2002), "GARCH vs. Stochastic Volatility: Option Pricing and Risk Management ", *Journal of Banking & Finance*, Vol 26, pp323-345.

MELICK, W R AND C P THOMAS (1997), "Recovering an asset's implied PDF from Option Prices: An Application to crude oil during the Gulf crisis ", *Journal of Financial and Quantitative Analysis*, March 1997, pp91-115.

POLITIS, D N AND J R ROMANO [1994], "The Stationary Bootstrap," *American Statistical Association*, Vol 89, Issue 428, pp1303-1313.

QUESENBERRY, C P [1986], *Some Transformation Methods in Goodness-of-Fit*, in: "Goodness-of-Fit Techniques", eds R B D'Agostino and M A Stephens, pp235-275.

ROGOFF, K (1980), "Tests of the martingale model of foreign exchange futures markets." In *Essays on expectations and eccahnge rate volatility.* Ph.D. dissertation, Massachustts Institute of Technology.

ROSENBLATT (1952), "Remarks on a Multivariate Transformation," *An. Math. Stat.* Vol 23, pp470-472.

SÖDERLIND, P AND L SVENSSON (1997), "New techniques to extract market information from financial instruments ", *Journal of Monetary Economics*, Vol 40, pp383-429.

SHILLER, R J (2000), "Irrational Exuberance", Broadways Book, New York.

VON MISES, R (1931), "Wahrscheinlichkeitsrechnung", Vienna, Leipzig.

WEINBERG S (2001), "Interpreting the volatility smile: An examination of the information content of option prices " Board of Governors of the Federal Reserve System, International Finance Discussion Paper 706.

6 Appendix

We are interested in the density of the integral transforms, $f_{h,t}(z_{h,t})$ at time $t+h$. From equation (2) we know that the integral transforms $z_{h,t} = \widehat{\Pi}_{h,t}(X_{t+h})$ where $\widehat{\Pi}_{t,h}(X_{t+h})$ is the estimated *cdf* (and therefore monotonic function) of X_{t+h} based on information at time t. Moreover, assume that the true density of X_{t+h} is $\pi_{h,t}(X_{t+h})$ at $t+h$.

- Then, since $\widehat{\Pi}_{h,t}(X_{t+h})$ is monotonic, the inverse transformation, $X_{t+h} = \widehat{\Pi}_{h,t}^{-1}(z_{t+h})$, exists.
- The Jacobian of the transformation is the absolute value of the determinant of the partial derivative $J = \left|\frac{\varrho X_{t+h}}{\varrho z_{h,t}}\right| = \left|\frac{\varrho \widehat{\Pi}_{h,t}^{-1}(z_{h,t})}{\varrho z_{h,t}}\right|$.
- Then the density $f_{h,t}(z_{h,t}) = \pi_{h,t}(X_{t+h}) \left|\frac{\varrho \widehat{\Pi}_{h,t}^{-1}(z_{h,t})}{\varrho z_{h,t}}\right| = \pi_{h,t}(\widehat{\Pi}_{h,t}^{-1}(z_{h,t})) \left|\frac{\varrho \widehat{\Pi}_{h,t}^{-1}(z_{h,t})}{\varrho z_{h,t}}\right|$.
- Inserting values for $z_{h,t}$ and $\widehat{\Pi}_{h,t}^{-1}(z_{h,t})$ in $\left|\frac{\varrho \widehat{\Pi}_{h,t}^{-1}(z_{h,t})}{\varrho z_{h,t}}\right|$ yields $\left|\frac{\varrho X_{h,t}}{\varrho \widehat{\Pi}_{h,t}(X_{h,t})}\right| = \frac{1}{\widehat{\pi}_{h,t}(X_{h,t})}$.
- Therefore, $f_{h,t}(z_{h,t}) = \frac{\pi_{h,t}(\widehat{\Pi}_{h,t}^{-1}(z_{h,t}))}{\widehat{\pi}_{h,t}(X_{h,t})} = \frac{\pi_{h,t}(X_{h,t})}{\widehat{\pi}_{h,t}(X_{h,t})}$.

  Since $\widehat{\pi}_{h,t}(X_{t+h})$ is the estimated density, and $\pi_{h,t}(X_{t+h})$ is the true density of X_{t+h}, $f_{h,t}(z_{h,t}) \sim U(0,1)$ if $\widehat{\pi}_{h,t}(X_{t+h}) = \pi_{h,t}(X_{t+h})$.

Federal Reserve Bank
of Cleveland
Research Department
P.O. Box 6387
Cleveland, OH 44101

Address Correction Requested:
Please send corrected mailing label to the
Federal Reserve Bank of Cleveland
Research Department
P.O. Box 6387
Cleveland, OH 44101

PRST STD
U.S. Postage Paid
Cleveland, OH
Permit No. 385