working paper

03 13

The Empirical Performance of Option-Based Densities of Foreign Exchange

by Ben R. Craig and Joachim G. Keller

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 03-13 November 2003

The Empirical Performance of Option-Based Densities of Foreign Exchange

By Ben R. Craig and Joachim G. Keller

In this paper, we calculate risk-neutral densities (RND) by estimating the daily diffusion process of the underlying futures contract for foreign exchange, based on the price of the American puts and calls reported on the Chicago Mercantile Exchange for the end of the day. Our quick and accurate method of calculating the prices of the American options uses higher-order lattices and smoothing of the option's value function at the boundaries to mitigate the nondifferentiability of the payoff boundary at expiration and the early exercise boundary. We estimate the diffusion process by minimizing the squared distance between the calculated prices and the observed prices in the data. We also test whether the densities provided from American options provide a good forecasting tool. We use a nonparametric test of the densities that depends on inverse probabilities. We modify the test to compensate for an inherent problem that arises from the time-series nature of the transformed variables when the forecasting windows overlap. We find that the densities based on the American option prices for foreign exchange do considerably well for the longer time horizons.

Key words: risk-neutral density, option prices, diffusion process
JEL codes: 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. Joachim G. Keller is at the Bundesbank and may be reached at joachim.keller@bundesbank.de. This working paper was previously presented at the regular joint research workshop of the Deutsche Bundesbank and the Oesterreichische Nationalbank. It is also available as a Deutsche Bundesbank Working Paper and as Oesterreichische Nationalbank Working Paper No. 60. The authors thank Christoph Fischer, Heinz Herrmann and Christian Upper as well as seminar participants at the Deutsche Bundesbank, at the joint research workshop of Deutsche Bundesbank and the Oesterreichische Nationalbank and at the "Joint Lunchtime Seminar" organized by the Center for Financial Studies, ECB, and Deutsche Bundesbank.

1 Introduction

It is well known that with complete markets, a sufficiently rich set of European options prices implies a state price density that one may interpret as a probability density over the price that underlies the derivative contract, if agents are risk neutral. In this case the state price density is called a risk neutral density. European options have been used to recover the risk neutral densities for a variety of prices and indices, including oil and the Standard and Poor's 500 index. The richest market for foreign exchange options present a difficulty in applying this theory, however. The most liquid foreign exchange options, sold on the Chicago Mercantile Exchange are American options based on a futures price. As is well known, this type of option have an early exercise feature that destroys the logic behind computing the risk neutral densities from European options. To see this, the European option price, $c_t(K, X, T-t)$, (in this case of a call option) at time, t, with a strike price, K, expiring at time T, in a one state model can be expressed as

$$c_t(K, X, T-t) = e^{-\rho(T-t)} \int_K^{\infty} (X_T - K)\pi_T(X)dX \tag{1}$$

where ρ is the discount rate, (here assumed constant) and $\pi_T(X)$ is the risk neutral density over the state space of X at the expiration date T. As pointed out by

Breeden and Litzenberger (1978), differentiation of this expression twice with respect to the strike price, K, gives the risk neutral density, $\pi_T(X)$ times a discount factor, $e^{-\rho(T-t)}$. The subsequent literature (e.g. Shimko (1993), Malz (1997), Jackwerth and Rubinstein (1996) and Stutzer (1996)) has concentrated on estimation of the density from noisy or, in the Malz case, extrapolated data on prices by using parametric distributions, mixtures of parametric distributions, or non-parametric smoothers to fit the second derivative of the option price function with respect to the strike price. Others, like Neuhaus (1995) do not rely on smoothing equations and calculate probabilities at and between strike prices. Once the risk neutral density is calculated, then it can be used to forecast the price of the underlying basis for the option, or it may be used to price other derivatives based on the same sequence.

With an American option based on a future price, the relationship in equation (1) breaks down. The expectation operator must take into account the early exercise boundary, which will differ for each option based on a different strike price, and differ by time to expiration for the same option. Under this regime, equation (1) is no longer true, and arguments which generate equation (1) from the theory of option pricing, such as application of Feynmann-Kac to the partial differential equation system defining the evolution of the option price no longer make sense. This leaves a researcher with two choices. One can use a thinner market, such as the European options offered by the Philadelphia exchange or use the European options prices where they are quoted by a single bank. Another possibility, explored in this paper, is to calculate the risk neutral densities from American option prices on the thickly traded market by using methods that are theoretically consistent with the early exercise option.

The method adopted in this paper to calculate the risk neutral density in this case is to first estimate the underlying process of the underlying futures contract for foreign exchange, based on the traded price of the American puts and calls reported for the end of the trading day. This estimated process implies a risk neutral density for each point of time in the future. In order to estimate the diffusion process we need methods of calculating the prices of American options that are fast and accurate. The numerical problems posed by American options are tough. We solve the pricing of American options by using higher order lattices combined with smoothing at the boundaries in order to mitigate the non-differentiability of both the payoff boundary at expiration and the early exercise boundary. By calculating the price of an American option quickly, we can estimate the diffusion process by minimizing the sum of the squares between the calculated prices and the observed prices in the data.

This paper also tests whether the densities provided from American options provide a good forecasting tool. We use a non-parametric test of the densities that depends on the inverse probability ideas of Fischer (1930) and others. A problem with the use of these tests in the past has been the time series nature of the transformed variables when the forecasting windows overlap. The inverse probability of the realized thirty day ahead spot at time, t, is correlated with the same corresponding number at time $t-1$, because the spot shares twenty-nine days of history. We modify the tests based on the inverse probability functions to account for this correlation between our random variables that are uniform under the null.

We find that the densities based on the American option markets for foreign exchange do quite well for the thirty to sixty day time horizon. Less sophisticated models of the diffusion process, such as the simple log normal Black-Scholes model, do less well than more sophisticated models in forecasting the one-hundred-eighty day horizon. However, all of the single state models described in this paper fail to match the data for short time horizons.

The plan of the paper is this: first we describe our data. The next section lays out the numerical methods we used to calculate the risk neutral densities implied by American option prices based on a futures contract. Next we describe the tests that we use to evaluate our implied densities, especially those that take into account the time series nature of the overlapping windows of the forecasts. Our results are detailed in the next section and are followed by a short section where we lay out some of the implications that may be drawn from our study.

2 The data

The American options are exchange-traded, approach a fixed expiration date and can be exercised before maturity. Our data are over two million transaction prices from the Chicago Mercantile Exchange (CME) for fifteen years of options based on the US dollar DM futures prices. The prices are close of day transactions, and they always represent prices which have been used in an exchange on that day. While these data are advantageous in that they represent the most liquid market for foreign exchange options, and they include more different strike prices each day than all other data sources combined, they have a major disadvantage: because of historical reasons, these are American style options based on an underlying future. Because of this there is a substantial incentive to exercise the option early. One can think of the underlying future as providing a continuous stream of "dividends" as the future price changes to reflect the known expected change of the foreign exchange. As is well known, an American style option on an underlying stock which provides a continuous stream of dividends does not always provide incentive to hold the option until its expiration date. For some values of the underlying price, a trader can do better by cashing in the option early. This provides a "boundary" of prices, under or over which (depending on whether the option is a call or a put) the trader always exercises the option before the expiration date. This early exercise boundary is something that we take account of in calculating our risk neutral densities.

In addition, some of the data are especially noisy. As a result we imposed some requirements which all our data had to meet. All options included in the data set had to have both a volume of exchange and an open interest that were positive on the trading day. In addition, because of the historical illiquidity in certain markets, other prices were excluded: options expiring within 10 days of the current trading date, options expiring more than 100 days from the current trading date, and options with strike prices that are greater than .05 in relative, time normalized moneyness. In other words, options are excluded if $\left|\frac{X_t-K}{K\sqrt{T-t}}\right| > .05$, with K being the strike price, X_t the actual futures rate and $\sqrt{T-t} = \sqrt{\tau}$ the normalizing time factor, which is the difference between expiration date T and the actual date t . This excludes those options in the extreme tails where prices are known to be driven more by illiquidity than by market expectations. The time period under investigation runs from January 25, 1984 to December 31, 1998. Days with traded options that did not include at least 8 different strike prices were excluded. This left us with 3900 separate trading days with which to estimate densities. The number of different options on the days where densities were estimated ran from a low of 8 to a high of 106. An average day included about 58 options prices that were usable. Note that all option prices that matched the above filters were used, even those that occasionally did not meet the arbitrage conditions implied by option theory. (In the two million data points this happened about 20 times). In the case of our estimation, these anomalies were considered part of the error term in the non-linear least squares technique.

3 Estimation of the Densities

Following Dumas et.al. (1998), our procedure is to estimate the parameters of a diffusion process in order to approximate the risk neutral density for each day. Thus we first calculate the instantaneous volatility of the spot, $\hat{\sigma}_t(X, \tau, \widehat{\beta})$, a function of the state of the exchange rate and of time to expiration τ of the contract. We estimate the diffusion function, $\hat{\sigma}_t(X, \tau, \widehat{\beta})$, parametrically, by minimizing with respect to a parameter vector $\widehat{\beta}$ the sum of the squared deviations of the observed option prices from the prices implied by $\hat{\sigma}_t(X, \tau, \widehat{\beta})$. This function is estimated separately for each day for which we have options price data. Each function implies a distinct risk neutral density for any period ahead for which one wishes to forecast.

As is usual when handling option prices, a trade off must be made between having a rich enough parameterization of $\hat{\sigma}_t(X, \tau, \widehat{\beta})$ to capture the details of the market's valuation of the risk and over fitting. Following the literature on fitting European options to single state diffusions, we fit four specifications of $\hat{\sigma}_t(X, \tau, \widehat{\beta})$ in this paper.

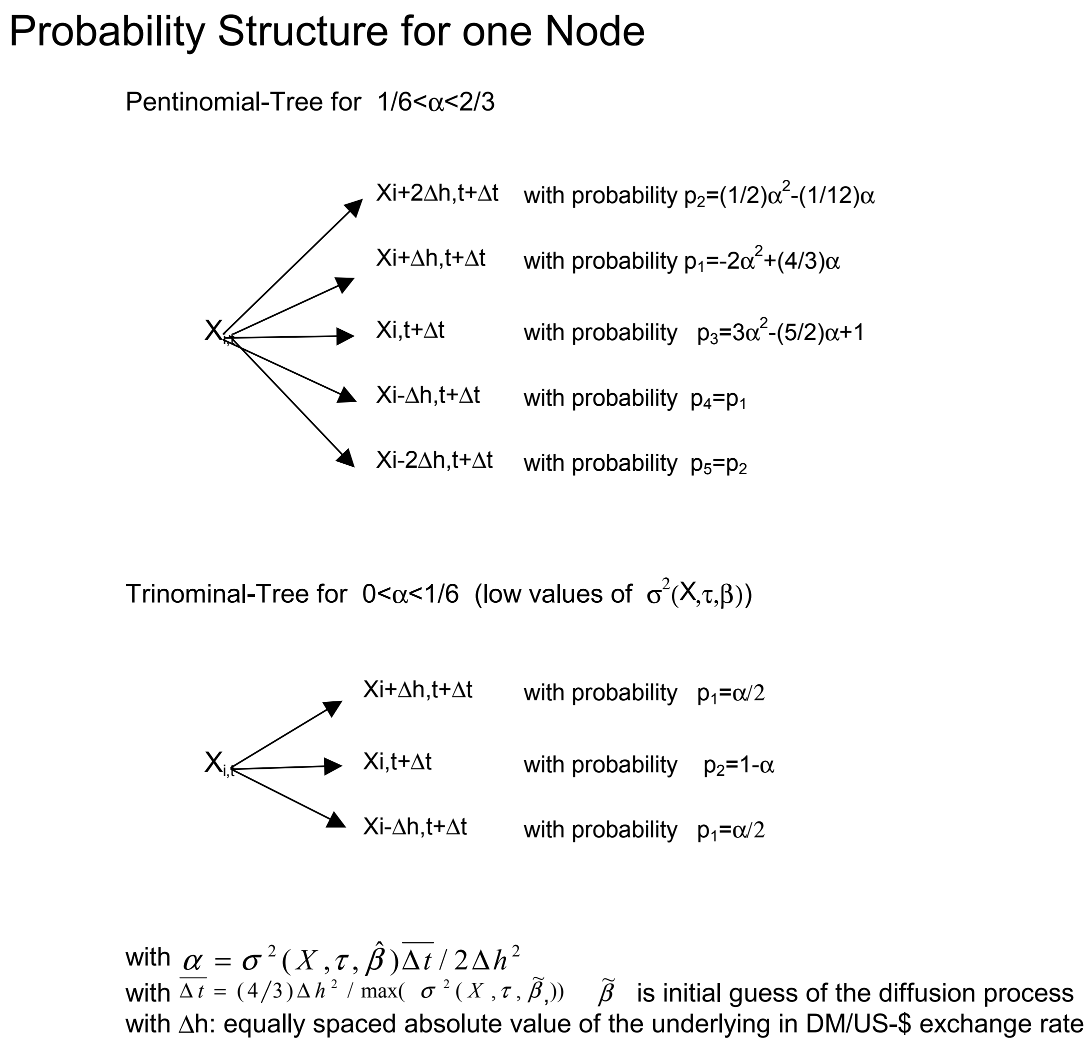
$$\begin{aligned} \hat{\sigma}_t(X, \tau, \widehat{\beta}) &= \beta_1 X, \\ &\quad \beta_0 + \beta_1 X, \\ &\quad \beta_0 + \beta_1 X + \beta_2 X^2, \\ &\quad \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 \end{aligned} \tag{2}$$

The first parameterization is the Black Scholes, log normal specification. The second adds a normal term which has the effect of allowing for thicker tails on the density. The third and fourth specifications are polynomial extensions to this which allow for the standard volatility "smile" and "sneer" often observed in foreign exchange options.

By estimating the diffusion process rather than the implied state space density for each expiration date, we allow for tests of forecast densities of a variety of horizons, not just the expiration dates for which we have option data. We obtain forecast ahead densities for one, seven, fourteen, thirty, ninety and one-hundred-eighty days ahead of the current information by using the separate estimates $\hat{\sigma}_t(X, \tau, \widehat{\beta})$ for each day, t. From these densities we acquire the series $\widehat{\Pi}_{\theta,t}(X_{t+\theta})$, which is the probability, given the estimated density at t that the θ ahead forecast is less than or equal to the observed θ ahead outcome, $X_{t+\theta}$. For clarification reasons, we drop the θ notation when we refer to an estimated density, so that $\widehat{\Pi}_{\theta,t}(X_{t+\theta}) = \widehat{\Pi}_t(X_t)$.

Estimation of the daily diffusions $\hat{\sigma}_t(X, \tau, \widehat{\beta})$ hinges on being able to calculate the price of a given option quickly and accurately, given an arbitrary function $\hat{\sigma}_t(X, \tau, \widehat{\beta})$. We accomplish this by using higher order lattice methods. Lattices are simply discretizations of both the time and the state space that allow one to compute the value function for each option directly. A binomial tree is a lattice with two branches. Our initial work with binomial lattices suggested that they did not converge quickly enough to provide accurate prices of the options. Therefore we use higher order lattices that hold the intervals of discretization of the state space and time constant and have more branches. In our case, we match the first five moments of the Brownian motion process assumed in our parameterization of $\hat{\sigma}_t(X, \tau, \widehat{\beta})$.

The probability weights for each branch are given in figure 1. They are derived



Probability Structure for one Node
Pentinomial-Tree for $1/6<\alpha<2/3$
$X_{i,t}$
$X_{i+2\Delta h,t+\Delta t}$ with probability $p_2=(1/2)\alpha^2-(1/12)\alpha$
$X_{i+\Delta h,t+\Delta t}$ with probability $p_1=-2\alpha^2+(4/3)\alpha$
$X_{i,t+\Delta t}$ with probability $p_3=3\alpha^2-(5/2)\alpha+1$
$X_{i-\Delta h,t+\Delta t}$ with probability $p_4=p_1$
$X_{i-2\Delta h,t+\Delta t}$ with probability $p_5=p_2$
Trinominal-Tree for $0<\alpha<1/6$ (low values of $\sigma^2(X,\tau,\beta)$)
$X_{i,t}$
$X_{i+\Delta h,t+\Delta t}$ with probability $p_1=\alpha/2$
$X_{i,t+\Delta t}$ with probability $p_2=1-\alpha$
$X_{i-\Delta h,t+\Delta t}$ with probability $p_1=\alpha/2$
with $\alpha=\sigma^2(X,\tau,\hat{\beta})\overline{\Delta t}/2\Delta h^2$
with $\overline{\Delta t}=(4/3)\Delta h^2/\max(\sigma^2(X,\tau,\tilde{\beta}))$ $\tilde{\beta}$ is initial guess of the diffusion process
with Δh: equally spaced absolute value of the underlying in DM/US-$ exchange rate

Figure 1:

by solving six equations in seven unknowns (the probabilities p_i, $\hat{\sigma}_t(X,\tau,\widehat{\beta})^2$, Δt), giving one degree of freedom, which we employ to set new state depending probabilities for each daily estimated diffusion process $\hat{\sigma}_t(X,\tau,\widehat{\beta})$. The system to solve is:

$$
\begin{aligned}
&p_1+p_2+p_3+p_4+p_5=1\\
&E\left[X(t+\Delta t)-X(t)\right]^1=p_1\Delta h+p_2 2\Delta h+p_3 0-p_4\Delta h-p_5 2\Delta h=0\\
&E\left[X(t+\Delta t)-X(t)\right]^2=p_1\Delta h^2+p_2(2\Delta h)^2+p_3 0+p_4\Delta h^2+p_5(2\Delta h)^2=\hat{\sigma}_t(X,\tau,\widehat{\beta})^2\Delta t\\
&E\left[X(t+\Delta t)-X(t)\right]^3=p_1\Delta h^3+p_2(2\Delta h)^3+p_3 0-p_4\Delta h^3-p_5(2\Delta h)^3=0\\
&E\left[X(t+\Delta t)-X(t)\right]^4=p_1\Delta h^4+p_2(2\Delta h)^4+p_3 0+p_4\Delta h^4+p_5(2\Delta h)^4=3\hat{\sigma}_t(X,\tau,\widehat{\beta})^4\Delta t^2\\
&E\left[X(t+\Delta t)-X(t)\right]^5=p_1\Delta h^5+p_2(2\Delta h)^5+p_3 0-p_4\Delta h^5-p_5(2\Delta h)^5=0
\end{aligned}
\tag{3}
$$

For either tree in figure 1, p_i depends upon $\alpha=\frac{\sigma_t(X,\tau,\widehat{\beta})^2\overline{\Delta t}}{2\Delta h^2}$. For the pentionomial tree the p_i are positive if and only if $\alpha\in\left[\frac{1}{6},\frac{2}{3}\right]$. For $\alpha\in\left[0,\frac{1}{6}\right]$, i.e. for c.p. small values of $\hat{\sigma}_t(X,\tau,\widehat{\beta})^2$, we reduce the pentionomial model to a trinomial model, by dropping the equations for the fourth and fifth moment and cutting the further branches (i.e., those branches with an increment of $2\Delta h$)[1].

The time step $\overline{\Delta t}$ is determined by the size of a chosen state space increment, Δh, a chosen value of α and the maximum $\widetilde{\sigma}_t(X,\tau,\widetilde{\beta})$ at the end of the lattice, given the initial guess $\widetilde{\beta}$ of the diffusion process on day t. A reasonable value of Δh proved to be 10^{-7}. This space increment yielded very acurate prices for European options for which an analytical solution exists. In our scheme we used a value for the time step of $\overline{\Delta t}=\frac{4}{3}\frac{\Delta h^2}{\max(\widetilde{\sigma}_t(X,\tau,\widetilde{\beta})^2)}$, which allowed the fourth moments to be matched for the largest part of the state space. This lays down the tree structure in terms of $\overline{\Delta t}$ and Δh for the whole estimation procedure for a trading day. To simplify the notation, we drop the bar and write $\overline{\Delta t}=\Delta t$. The probabilities are modified appropriately for fractional values of Δt when needed to place the lattice on those whole numbered

[1] In another version of this paper we circumvent the problem of "too small" values of $\hat{\sigma}_t(X,\tau,\widehat{\beta})^2$, and therefore of values of α below 1/6, by augmenting, if necessary, the state space increment Δh, so that the critical value of α is only reached by smaller values of $\hat{\sigma}_t(X,\tau,\widehat{\beta})^2$. In this case we have therefore an adaptive tree structure which allows for every value of $\hat{\sigma}_t(X,\tau,\widehat{\beta})^2$ to match the first 5 moments.

days when the options expire (so that the end condition can be set.) The value of the options on day t is calculated using the probabilities p_i of figure 1. The p_i change for each node, according to the diffusion process $\hat{\sigma}_t(X, \tau, \widehat{\beta})$ and the value of the state X, since α is a function of X.

In contrast to a scheme using approximations that calculate the value of the option only at the trading date, the early exercise boundary is easily incorporated within this framework by adding a maximization operator into the calculation of the discretized value functions at each node and at each time. Thus, for a call option, the value of the node at state X and time $t - \Delta t$, is

$$V(X, t - \Delta t) = \max\{e^{-\rho\Delta t} PV_{X_t}, Fp(X, t - \Delta t) - K\},$$

where

$$PV_{X_t} \equiv P_u(V(X+\Delta h,t)+V(X-\Delta h,t))+P_{2u}(V(X+2\Delta h,t)+ V(X-2\Delta h,t))+(1-2P_u-2P_{2u})V(X,t) \quad (4)$$

and where K is the strike price, $Fp(X, t - \Delta t)$ is the value of the underlying future price.

For a given diffusion, a higher order lattice approximates the value function of each option by using the higher order terms of a moment generating function for the true value function. The comparison of using a binomial tree to using a higher order approximation in evaluation a diffusion expectation is analogous to the comparison of using a sum of binomial variables to using the sum of multinomial variables that are close to the normal in evaluating a normal expectation. Because of central limit theorems, averaging the binomial outcomes does approximate the normal distribution, but it does so more slowly than the sum of variables drawn from a distribution closer to the normal.

The American option adds a complication to the calculation of a standard diffusion process. The argument above relies on the underlying true value function being smooth. This is a problem with options in general (because the value at the expiration date contains a point at the strike price where it is clearly non-differentiable) and with American options in particular (because the early exercise also creates a non-differentiability in the value function at the boundary). We handle this by using kernel smoothers. Thus, for a small distance around the early exercise boundary, in the neighborhood where $e^{-\rho\Delta t}PV_{X_t}$ and $Fp(X,t-\Delta t)-K$ are nearly equal, we use the value function, $\phi(e^{-\rho\Delta t}PV_{X_t})+(1-\phi)(Fp(X,t-\Delta t)-K)$ where ϕ is a many times differentiable kernel between 0 and 1, with the property that $\phi\to 1$ for values of $e^{-\rho\Delta t}PV_{X_t}$ that are 'much' greater than $Fp(X,t-\Delta t)-K$ and $\phi\to 0$ for $e^{-\rho\Delta t}PV_{Xt}$ that are 'much' smaller than $Fp(X,t-\Delta t)-K$. The kernel that we use is a Logistic cumulative distribution function, $\phi(\frac{e^{-\rho\Delta t}PV_{X_t}-(Fp(X,t-\Delta t)-K)}{\omega})$, where ω is the bandwidth. The bandwidth parameter ω, defines the term "'much' greater than" by determining how quickly $\phi(\frac{e^{-\rho\Delta t}PV_{X_t}-(Fp(X,t-\Delta t)-K)}{\omega})$ goes to one or zero for positive or negative values. Choosing ω too large over-smoothes in the sense that the underlying function evaluation is completely dominated by the smoothing function. Choosing ω too small does not solve the problem caused by non-differentiability for the higher order lattice. However, for a wide range of ω, calculation of the value of an option quickly converged to the theoretical true value where these were known. We report results for values of ω of .005 for the value function boundary and of .003 for the early exercise boundary. Although the kernel smoothing adds a lot of computation and complication even for small bandwidths, we find it makes a large difference in the calculated theoretical price of an option (and was much closer to the actual value of the option when we had a solution to compare our solution to.)

4 Evaluating density forecasts

Different methods of estimation lead to different forecasting densities, some of which necessarily must be wrong. The ranking of these non correct density forecasts is a difficult task. This is because a ranking depends on the often unknown individual loss function of agents, that may include more arguments than the first two moments. For example, decision makers with non symmetric expected loss indexes care about more than the mean and the variance of a distribution. Moreover, different agents have different loss functions, so that it is often impossible to find a ranking upon which all individuals agree unanimously. However, it can be shown, that the correct density is always preferred over false densities. Therefore, as a second best solution one tries to approximate the true density as good as possible.

To assess whether there is significant evidence whether the estimated densities coincide with the true densities at a first step we perform the probability integral transforms of the actual realizations. Under the hypothesis that the true densities functions correspond to our estimated densities the transformed realizations are uniformly distributed. To assess this property of the transformed realizations we suggest as a second step two different tests, based upon the distance of the observed distribution of the transformed random variables from the uniform distribution. This distance is in the L_2 topology, and was first suggested by Cramer in the 1920's. These tests are robust to time dependence in the data.

The basic univariate integral transformation theorem is due to Fischer (1930) and has been generalized for the multivariate case by Rosenblatt (1952). A thorough overview of transformation methods in Goodness-of-Fit techniques is given by Quesenberry (1986). Recently, Diebold et.al. (1998) apply this concept to time series, evaluating the densities implied by a MA(1)-t-GARCH(1,1) model. Clements and Smith (2001) use the probability integral transforms for evaluating the density forecast of a self-exciting threshold autoregressive model.

The basic idea is to evaluate a sequence of actual exchange rate realizations $\{X_t\}_{t=1}^N$ with respect to a sequence of densities $\{\widehat{\pi}_{\theta,t}\}_{t=1}^N$ $\left(= \{\widehat{\pi}_t\}_{t=1}^N\right)$ estimated at time t, with the information available at t. Again, the forecast horizon is θ. The probability integral transforms z_t correspond to the function values of the cumulative density functions, evaluated at $X_{t+\theta}$. For simplification $X_{t+\theta}$ is written as $X_{t.}$

$$z_t = \int_{-\infty}^{X_t} \widehat{\pi}_t (u) \, du = \widehat{\Pi}_t (X_t) \qquad t = 1, ..., N \tag{5}$$

Under the null hypothesis of correct forecast densities (i.e. $\Pi_t (X_t) = \widehat{\Pi}_t (X_t)$), the sequence of integral transformed realization $\{z_t\}_{t=1}^N$ is $U[0,1]$ and their theoretical *cdf* $F(pr_n) = pr_n$ is equal to the proportion of z_t's that is less than a number pr_n in the interval $(0,1)$ (see appendix).

The step between our estimates of the diffusion function, $\hat{\sigma}_t(X,\tau,\widehat{\beta})$, and calculation of the cumulative distribution of the observed t day ahead draw, X, is easy. We simulate a large number of draws from the diffusion process defined by $\hat{\sigma}_t(X,\tau,\widehat{\beta})$ (through a discrete Markov approximation) and record the proportion less than or equal to X, to get the estimated cumulative distribution ($Ecdf$) $\widehat{\Pi}_t(X_t) = z_t$.

We then test the null that the observed sequence of z_t's is a sequence of uniformly (though not necessarily independently) distributed random variables. In this paper, we use inference based upon bootstrap samples that preserve the time series properties of our original sample, $z_1, ...z_t, ..., z_N$. From this bootstrap we can construct confidence intervals for a variety of statistics. We report results from a distance statistic, the so-called Cramer-von Mises statistic [von Mises (1931)]. This statistic is defined as

$$\widehat{CvM} \equiv \int_0^1 (F(pr_n) - \widehat{F}(pr_n))^2 d(pr_n). \tag{6}$$

Note that this is a distance in the L_2 topology between the empirical distribution function ($Ecdf$) of z_t, $\widehat{F}(pr_n)$, and its theoretical value, $F(pr_n) = pr_n$, representing the uniform null. A similar statistic that was also computed with the same results lies in the L_∞ topology, the so called Kolmogorov-Smirnov statistic,

$$KS \equiv \sup_{pr_n} |F(pr_n) - \widehat{F}(pr_n)|. \tag{7}$$

This bootstrap procedure is lacking in that rejection of the null does not indicate where the proposed densities apparently fail.

For this, we test whether the $Ecdf$ (where in the following definition, $I^n(z_t \leq pr_n) = I^{n,t}$ is the indicator function),

$$\widehat{F}(pr_n) \equiv \frac{\sum_{t=1}^{N} I^n(z_t \leq pr_n)}{N}, \tag{8}$$

is equal to pr_n for a large number of different pr_n in the interval (0,1). We perform this test separately for each pr_n. These tests have the advantage of showing what quality of the outcome density is missing in the estimated forecasting density. For example, if the option implied densities have thicker tails than the forecasting outcomes then this shows up graphically as $\widehat{F}(pr_n) < pr_n$ for values of pr_n close to 0 or 1 and as $\widehat{F}(pr_n) > pr_n$ for values close to 0.5. However, this is not a powerful test because it fails to account for the departures of $\widehat{F}(pr_n)$ from pr_n jointly for all n.

One possible way to jointly test the departures for each pr_n would be to sum up their squares, as was suggested by Karl Pearson (1905) very early in the history of specification tests. However, this leads to problems of choosing the individual pr_n, and, ultimately, to the theory of inference in the presence of unbounded operators. We pursue that line of research in a separate paper.

First we expand our discussion of the tests based on the stationary bootstrap.

4.1 The stationary bootstrap approach

The stationary bootstrap approach (IFSB) of Politis and Romano (1994) uses a resampling procedure to calculate standard errors of estimators that account for weak data dependence in stationary observations. The procedure requires a sample of random blocks of random lengths out of the original time series, where the length L of each block is drawn from a geometric distribution, so that the probability of drawing a block of length L is $(1-prob)^{L-1}prob$ for $L = 1, 2, ...$ End effects (in case of a block going beyond the last observation) are handled by ordering the observations in a circle, so that the series "restarts" after the last observation. A difficult aspect in applying this procedure is the choice of the parameter governing the stochastic length of the blocks, $prob$. Politis and Romano suggest a data-based choice of $prob$ so that $prob = prob_N \to N^{-1/3}$, with N equal to the number of observations. By this choice the mean squared error of $\widehat{\sigma}^2_{bt,prob_N}$ as an estimator of σ^2_N is minimal. Fortunately, as long as $prob \to 0$ and $Nprob \to \infty$ fundamental consistency properties of the bootstrap are unaffected by choosing $prob$ suboptimaly. As can be directly seen, these requirements are clearly met by the choice of $prob = N^{-1/3}$.

We use the sample sequence $\{z_t\}$ to calculate the Cramer-von Mises statistic $\widehat{CvM}$ directly for our sample $Ecdf$, $\widehat{F}(pr_n)$, and then to calculate whether this is a significant distance from the 45°-line through the bootstrapped samples. Bootstrapped distribution functions, $F_b(pr_b)$ are also formed and the CvM_b statistic,

$$CvM_b \equiv \int_0^1 (F(pr_b) - F_b(pr_b))^2 d(pr_b) \tag{9}$$

is evaluated for each bootstrapped sample. Because the sample distribution function $\widehat{CvM}$ and all bootstrapped sample distribution functions CvM_b are step functions, the integral expression in CvM_b is calculated directly. 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. A value of CvM_b less than some critical value, α_0, rejects the hypothesis of $z_1, ...z_t, ..., z_N$ being drawn from a uniform distribution at the α_0 level.

4.2 Tests based on deviations of the empirical density from individual quantiles, pr_n

Figure 2 depicts the integral transformation. The simulated density $\widehat{\Pi}_t$ of the diffusion function $\hat{\sigma}_t(X, \tau, \widehat{\beta})$ is on the right side at the top and $\widehat{\pi}_t$, the corresponding first empirical derivative of $\widehat{\Pi}_t$ with respect to K, is situated at the bottom. The sequence of actual $\{z_1, ...z_t, ...z_N\}$ are on the left side at the top and the estimated $\widehat{F}(pr_n)$ are plotted below. Note, that the whole sequence of N actual z_t is generated by N diffusion functions, since for each diffusion process, estimated on t, one obtains only one z_t. However, the $\widehat{\Pi}_t$, $\widehat{\pi}_t$ and the sequence $\{z_1, ...z_t, ...z_N\}$ in figure 2 share the same forecast horizon θ (here $\theta = 30$ days).

The null hypothesis of correct forecasts corresponds to the dashed 45°-line that connects the origin of the diagram (on the left side at the bottom) to the upper left corner. The empirical proportion $\widehat{F}(pr_n)$ of the sequence $\{z_t\}$ being less than $F(pr_n)$ is represented by the n^{th} bar. The total number of bars is $\overline{N}$. The basis of each bar equals $1/\ \overline{N}$ and $F(pr_n) = pr_n = \sum_{i=1}^{n} 1/\ \overline{N}$. Under the null, each bar is crossed by the dashed line at its right corner.

To address the question whether violations of the uniformity $(\widehat{F}(pr_n) \neq F(pr_n))$

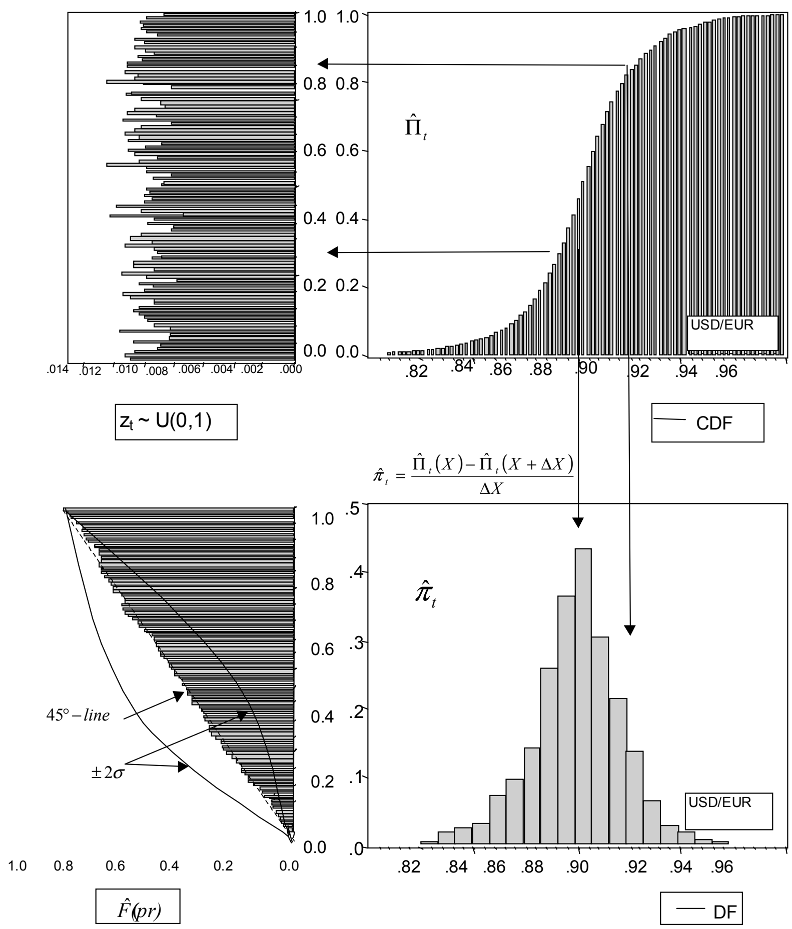


Figure 2: The integral transformation

are significant we need to estimate the standard deviations of $\widehat{F}(pr_n)$. However, since our θ-days ahead forecast densities are calculated daily, the evaluated realizations X_t and consequently the sequence of $\{z_t\}$ are time dependent due to the overlapping data problem. This issue arrises when the forecast horizon is longer than the sample frequency. If i.e. the sample frequency is daily and the forecasts are 1-month ahead (24 business days), the overlap amounts to 24 days. Thus, the forecast errors are no longer iid but follow a moving average process (MA) equal to the length of the forecast horizon θ. In this case inference from standard tests, which are based on the assumption of iid observations, is misleading. If the forecast errors are dependent, different types of standard tests, as Chi-squared tests and *Ecdf* tests i.e., lead the researcher to reject the true null hypothesis too often.

Our test consists of calculating confidence intervals for individual $\widehat{F}(pr_n)$ by using the function values $I^{n,t}(z_t \leq pr_n)$ of the indicator function. The time dependence of the observations is considered up to the order of the theoretical data overlap θ.

$$\widehat{var}(\widehat{F}(pr_n)) = \widehat{\sigma}^2(\widehat{F}(pr_n)) = \frac{1}{N}\left[\widehat{\gamma}^n(0) + 2\sum_{j=1}^{\theta}\left(1 - \frac{j}{N}\right)\widehat{\gamma}^n(j)\right] \tag{10}$$

where the sample autocovariance is defined by $\widehat{\gamma}(j)$.

$$\widehat{\gamma}^n(j) = \frac{1}{N}\sum_{t=j+1}^{N}\left(I^{n,t} - \overline{I^n}\right)\left(I^{n,t-j} - \overline{I^n}\right) \tag{11}$$

Under the null hypothesis, the ratio $t = (F(pr_n) - \widehat{F}(pr_n))/\widehat{\sigma}(\widehat{F}(pr_n))$ has a t distribution with $N-2$ degrees of freedoms.

5 The results

The results of the CvM_b statistics are reported in table 1. These tables present the probabilities that bootstrapped samples differ from original sample in the Cramer-von Mises distance by as much as the original sample differs from the null of the 45°-line. Lower values than .05 imply a rejection of the null at the five percent level. Several things are immediately clear from these tests. First, the data strongly support the options price densities as useful forecasting densities at the one to three month forecast horizon. In no case was the model rejected. Second, all of the models do a poor job of predicting the densities at the one week or shorter time horizon. Third, the simpler models and the more complicated models do about equally well at the thirty to ninety day horizon, and all do extremely poorly at the very short time horizon. Fourth, the complicated cubic polynomial does a better job of fitting the density for the half year horizon than the less complicated models. Indeed, it can not be rejected as a forecasting model of the density for this long horizon.

Horizon θ in days	1	7	14	30	90	180
Specification*						
$\beta_1 X$	.001	.004	.049	**.225**	**.204**	.015
$\beta_0 + \beta_1 X$	.005	.006	**.059**	**.218**	**.160**	.008
$\beta_0 + \beta_1 X + \beta_2 X^2$	.001	.005	**.055**	**.230**	**.133**	.023
$\beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$	.001	.002	.035	**.166**	**.151**	**.058**

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

table 1: Test results of the stationary bootstrap approach

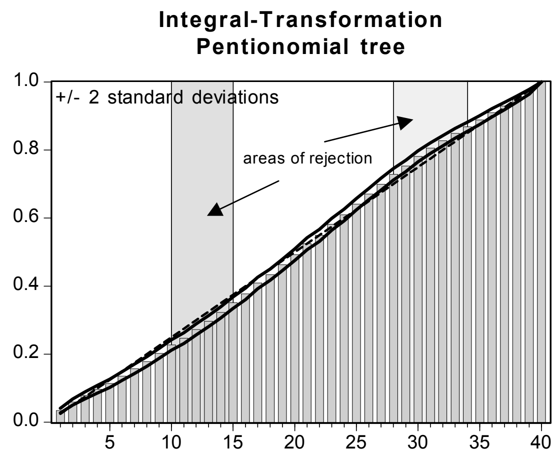
These broad patterns were also supported by other tests based on the bootstrapped variance of the CvM_b. To assess where the forecast densities fail, we plot the actual *Ecdf* of the z_t against the theoretical *cdf* for the extremely long and extremely short forecasted densities of the log Normal model. Results are shown in figures 3a and 3b. The number of bins $\overline{N}$ is 40, so that the basis of each bin corresponds to a probability mass of 0.025. Here, the n^{th} bar shows the function value $\widehat{F}(pr_n)$, while the theoretical *cdf* value $F(pr)$ is given by the dashed 45°-line above bin no. n. The thick bulging out lines surrounding the estimated $\widehat{F}(pr_n)$ indicate approximately the 95% confidence interval $(\pm 2\widehat{\sigma}\widehat{F}(pr_n))$, where $\widehat{\sigma}(\widehat{F}(pr_n))$ are calculated by (10).[2]

[2] In case of independent data $Var(pr_n) = ((1-pr_n)pr_n)/N$, which has its maximum at $2pr_n = 1$. Therefore at $pr_n = 0.5$ the bulging out of the confidence interval is biggest.

Lognormal Densities

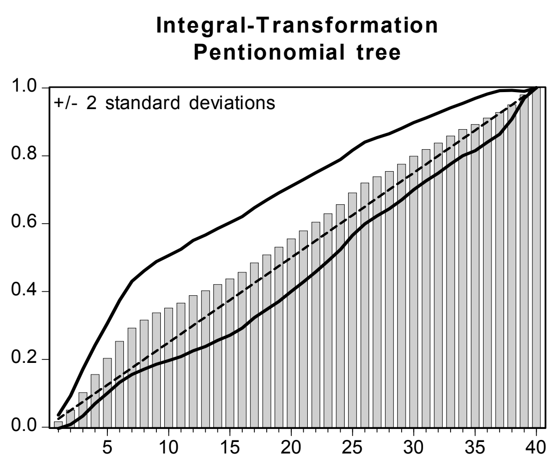
1 day horizon

Figure 3a



180 day horizon

Figure 3b



----- H_0: theoretical cdf of uniform (45-degree line)

bars: empirical cdf of uniform

——— estimated 2σ standard-deviation bounds of empirical cdf

Figure 3: The deviations of the empirical density from individual quantiles

Figure 3a shows that the areas of rejection of the density, represented by the greyed out areas, are somewhat symmetric. The density corresponding to the *Ecdf*'s displayed in figure 3a would be slightly below the uniform line on the extreme ends of the distribution and then slightly above it on the other portions of the density. Thus, the option forecast densities fail at the short horizon because they do not place enough mass at the extreme ends of the densities. The tails are not fat enough. Note also that the confidence bands are fairly tight with the one day horizon. The Cramer-von Mises test has fairly good power at this horizon, due to the large number of independent daily observations.

Figure 3b shows a very different picture for the longer time horizon. Although the power of the graphic test is too small to reject the null (which is rejected overwhelmingly by the Cramer-von Mises test) it is clear that the log Normal model overpredicts very low outcomes of the Dollar to Deutsche Mark ratio. Following a forecasting model based on a Black-Scholes model results in the lowest twenty percent of possible outcomes being too pessimistic in terms of the value of the DM. The confidence bands for these estimates are much wider because of the high correlation of the actual outcomes in dates within a month of one another.

6 Concluding remarks

Our results fall into two groups, one the thirty to ninety day time horizon for which the forecasting densities seem to fit the data fairly well, and the very short and the very long horizons which are poor specifications for forecast densities (except for the cubic diffusion model which is not rejected for the long horizon.)

Where the densities fail as forecasting tools, several points should be noted. First, the polynomial expansion of a single state specification of the variance clearly limits the set of models, that can be fitted to the date. More work can be done to specify a set of models that are sufficiently rich to match the option prices, either by increasing the dimension of the states, controlling the diffusion process or by incorporating time dependence into the process. Second, the time horizons for which we do not fit the data correspond exactly to the expiration dates of the option contracts which we cast out of our data set on the grounds that these are typically low liquidity markets. Thus, we did not use all price data of options which expired within ten days of the trading day, options that are perhaps best designed to forecast the future exchange rate at one and seven days ahead. Clearly there is fruitful work to be done in examining the trade-off in estimating risk neutral densities between the signal and noise provided by thinly traded options.

The second group of conclusions concern the thirty to ninety day horizons where our tests clearly do not reject any of the specifications of the diffusion process as forecasting densities. This is in spite of the fact that we used techniques that allowed the information from all of the daily observations between the years 1983 and 1998. This finding is not solely a result of poor power of our tests. In other research (Craig and Keller (2001)), we resoundingly reject densities on the thirty day horizon implied by other methods, such as a GARCH technique, or based on other options with lower liquidity, even though these tests are based only on less than three years of data.

The implications of the lack of rejection of these state space densities are of some importance. The first one is that the pricing of risk in these very liquid markets is very low. In other words, any risk premium built into the state price densities is small enough that the risk neutral implied density is indistinguishable from the forecasting density. Any theory, such as uncovered interest rate parity, which relies on large shifts in risk premia in order to reconcile it with the data is thus less convincing.

Second, these risk neutral densities are fairly good estimates of the market's forecast of future prices. These densities can be computed daily, and thus form a useful policy tool, as well as providing an important set of data with which to test deeper theories of foreign exchange determination.

Having stated that, we must admit that there is much left to do in testing the densities before we can say more. The tests are not powerful enough to distinguish the fairly simple parameterization offered in this paper from each other, or from more elegant parameterizations of the densities. The diffusion densities offered here seem very crude approximations when compared with the densities often calculated with non-parametric techniques from European options. In contrast to these multi-modal, quickly changing shapes, our densities are often unimodal, and usually are close to the density of the previous day. A more powerful test might have much to say about which of the densities represent the market's true assessment of possibilities.

The tests of the densities that are explored in this paper are of lower power than other more specific tests in part because of their all encompassing character. In other words, the CvM test is designed to cover all possible specifications against all possible alternatives. The CvM test does perform well against other such tests, including the Kolmorgorov-Smirnov test, in terms of power. However, as shown in Csőrgő and Horváth (1993), the CvM test does not exploit much information that may be known about the null, such as behavior of the density in the tails. Further, in the space of probability distribution functions, the CvM is only optimal for deviations in the L_2-direction $cos(\sigma x)$ as shown by Gregory (1980). The theory of statistical distribution specification testing is still fruitful, offering major new advances each year. It is our hope that with these advances tests of sufficient power to distinguish different parameterizations of the diffusion process may be developed.

References

Breeden, D. and R.H. Litzenberger [1978]: "Prices of State-Contingent Claims Implicit in Option Prices," Journal of Business, 51, pp. 621-651.

Clements, M.P. and J. Smith [2001]: "Evaluating Forecast from SETAR Models of Exchange Rates," Journal of International Money and Finance, 20, pp. 133-148.

Craig, B. and J. Keller [2001]: "A Comparison of Densities Implied by Over-the-Counter and Exchange Traded Options in Foreign Exchange," unpublished M.S.

Cramer, H. [1928]: "On the Composition of Elementary Errors," Skand. Aktuartidskr., pp. 141-180.

Csőrgő, M. and L. Horváth [1993]: "Weighted Approximations in Probability and Statistics," Wiley, New York.

Diebold, F.X. and T.A. Gunther and A.S. Tay [1998]: "Evaluating Density Forecasts with Applications to Financial Risk Management," International Economic Review, 39, 4, pp. 863-883.

Dumas, B., J. Fleming and R.E. Whaley [1998]: "Implied Volatility Functions: Empirical Tests," Journal of Finance, 56, pp. 2059-2106.

Figlewski, S. and B. Gao [1999]: "The Adaptive Mesh Model: A New Approach to Efficient Option Pricing," Journal of Financial Economics, 53, pp. 313-351.

Fischer, R.A. [1930]:"Inverse Probability," Proceedings of the Cambridge Philosophical Society, 36, 4, pp. 528 - 535.

Gregory, G.G. [1980]: "On the Efficiency and Optimality of Quadratic Tests," Annals of Statistics, pp. 116-131.

JACKWERTH, C.J. AND M. RUBINSTEIN [1996]: "Recovering Probability Distributions," Journal of Finance, 51, pp. 1611-1631.

MALZ, A.M. [1997]: "Estimating the Probability Distribution of the Future Exchange Rate from Option Prices," The Journal of Derivatives, 4, pp. 18-36.

NEUHAUS, H. [1995]: "The Information Content of Derivatives for Monetary Policy", Discussion Paper 3/95, Economic Research Group of the Deutsche Bundesbank.

PEARSON, K. [1905]: "On the General Theory of Skew Correlation and Nonlinear Regression," Draper's Non-linear Regression, Draper's Company Memoirs, Biometric Series II.

POLITIS, D.N. AND J.R. ROMANO [1994]: "The Stationary Bootstrap," American Statistical Association, 89, 428, pp. 1303-1313.

QUESENBERRY, C.P. [1986]: "Some Transformation Methods in Goodness-of-Fit", in: Goodness-of-Fit Techniques, eds. Ralph B. D'Agostino and Micheal A. Stephens, pp. 235-275.

ROSENBLATT [1952]: "Remarks on a Multivariate Transformation," An. Math. Stat. 23, pp. 470-472.

SHIMKO, D.[1993]: "Bounds of Probability," Risk, 6, 4, pp. 33-37.

STUTZER, M. [1996]: "Maximum Entropy and Options Prices," Journal of Finance, 51, pp. 1588-1610.

VON MISES, R. [1931]: "Wahrscheinlichkeitsrechnung", Wien, Leipzig.

Appendix

We are interested in the density of the integral transforms $f_t(z_t)$ at time $t+\theta$ From (5) we know that the integral transforms $z_t = \widehat{\Pi}_t(X_t)$, where $\widehat{\Pi}_t(X_t)$ is the estimated *cdf* (and therefore monotonic function) of X_t based on information at time t. Moreover, assume that the true density of X_t is $\pi_t(X_t)$ at t.

- Then, since $\widehat{\Pi}_t(X_t)$ is monotonic, the inverse transformation $X_t = \widehat{\Pi}^{-1}(z_t)$ exists.
- The Jacobian of the transformation is the absolute value of the determinant of the partial derivative $J = \left|\frac{\varrho X_t}{\varrho z_t}\right| = \left|\frac{\varrho \widehat{\Pi}_t^{-1}(z_t)}{\varrho z_t}\right|$.
- Then the density $f_t(z_t) = \pi_t(X_t)\left|\frac{\varrho \widehat{\Pi}_t^{-1}(z_t)}{\varrho z_t}\right| = \pi_t(\widehat{\Pi}^{-1}(z_t))\left|\frac{\varrho \widehat{\Pi}_t^{-1}(z_t)}{\varrho z_t}\right|$.
- Inserting values for z_t and $\widehat{\Pi}_t^{-1}(z_t)$ in $\left|\frac{\varrho \widehat{\Pi}_t^{-1}(z_t)}{\varrho z_t}\right|$ yields $\left|\frac{\varrho X_t}{\varrho \widehat{\Pi}_t(X_t)}\right| = \frac{1}{\widehat{\pi}_t(X_t)}$
- Therefore $f_t(z_t) = \frac{\pi(\widehat{\Pi}^{-1}(z_t))}{\widehat{\pi}_t(X_t)} = \frac{\pi_t(X_t)}{\widehat{\pi}_t(X_t)}$.

  Since $\widehat{\pi}_t(X_t)$ is the estimated density and $\pi_t(X_t)$ is the true density of X_t, $f_t(z_t) \sim U(0,1)$ if $\widehat{\pi}_t(X_t) = \pi_t(X_t)$.

List of Variables

α : restriction parameter on probabilities p_i within tree

α_0 : critical value

$\widehat{\beta}$: parameter vector of diffusion process

$\beta_0, \beta_1, \beta_2, \beta_3$: coefficients of parameter vector $\widehat{\beta}$

$c_t\,(K, X, \tau)$: option price (call option)

cdf : cumulative density function

$\widehat{CvM}$: sample Cramer-von Mises statistics

CvM_b : bootstrapped Cramer-von Mises statistics

Δt : time step within tree

Δh : state space step within tree

Ecdf : empirical or estimated cumulative density function

$f(z_t)$: density of the integral transform z_t, with $f(z_t) \sim U[0, 1]$

$F(pr_n)$: theroretical *cdf* of f, in fig. 2 (45°-line).

$\widehat{F}(pr_n)$: *Ecdf*, in fig. 2 (bars deviating from 45°-line).

$Fp(X, t - \Delta t)$: value of the underlying future price

$I^{n,t}$: indicator function

K : strike price

L : length of the bootstrapping block

L_2 : distance in the L_2 topology

$\phi(PV, X, Fp, \rho, \omega, \Delta t)$: exponential kernel smoother

n : no. of bins of indicator function $I^n()$.

N : no. of observations

$\overline{N}$: total no. of bins

p_i : probabilties within tree at node $i \in \{1, 2, 3, 4, 5\}$

pr_n : proportion number $pr_n \in (0, 1)$

$prob_N$: data based choice of the bootstrap probability.

$\pi_T(X)$ theoretical risk neutral density over the state space of X

$\widehat{\pi}_t$: forecast density

$\widehat{\Pi}_t$: cumulative forecast density

ρ : discount factor

$\widehat{\sigma}_t(X, \tau, \widehat{\beta})$: diffusion process of the state space (spot rate)

t : actual date

T : expiration date

τ : time to maturity

θ : forecast horizon

$U[0,1]$: uniform density

ω : bandwidth of kernel smoother

$V(X, t-\Delta t)$: value function of American option

X : state space (spot rate)

z_t: integral transformed realizations of X_t, with $z_t = \widehat{\Pi}_t(X_t)$