Rupee-Dollar Option Pricing and Risk Measurement: Jump Processes, Changing Volatility and Kurtosis Shifts

Jayanth R. Varma

Abstract

Exchange rate movements in the Indian rupee (and many other emerging market currencies) are characterised by long periods of placidity punctuated by abrupt and sharp changes. Many, but by no means all, of these sharp changes are currency depreciations. This paper shows that econometric models of changing volatility like Generalised AutoRegressive Conditional Heteroscedasticity (GARCH) with non normal residuals which perform quite well in other financial markets fail quite miserably in the case of the INR-USD process because they do not allow for such jumps in the exchange rate. The empirical results very convincingly demonstrate the need to model the exchange rate process as a mixed jump-diffusion (or normal mixture) process. Equally importantly, the empirical results provide strong evidence that the jump probabilities are not constant over time. From a statistical point of view, changes in the jump probabilities induce large shifts in the kurtosis of the process. The failure of GARCH processes arises because they allow for changes in volatility but not for changes in kurtosis. The time varying mixture models are able to accommodate regime shifts by allowing both volatility and kurtosis (not to mention skewness) to change. This also shows that the periods of calm in the exchange rate are extremely deceptive; in these periods, the variance of rate changes is quite low, but the kurtosis is so high (in the triple digit range) that the probability of large rate changes is non trivial. The empirical results also show that the Black-Scholes-Garman-Kohlhagen model for valuation of currency options is quite inappropriate for valuing rupee-dollar options and that the Merton jump-diffusion model is the model of choice for this purpose.

Rupee-Dollar Option Pricing and Risk Measurement: Jump Processes, Changing Volatility and Kurtosis Shifts

Jayanth R. Varma

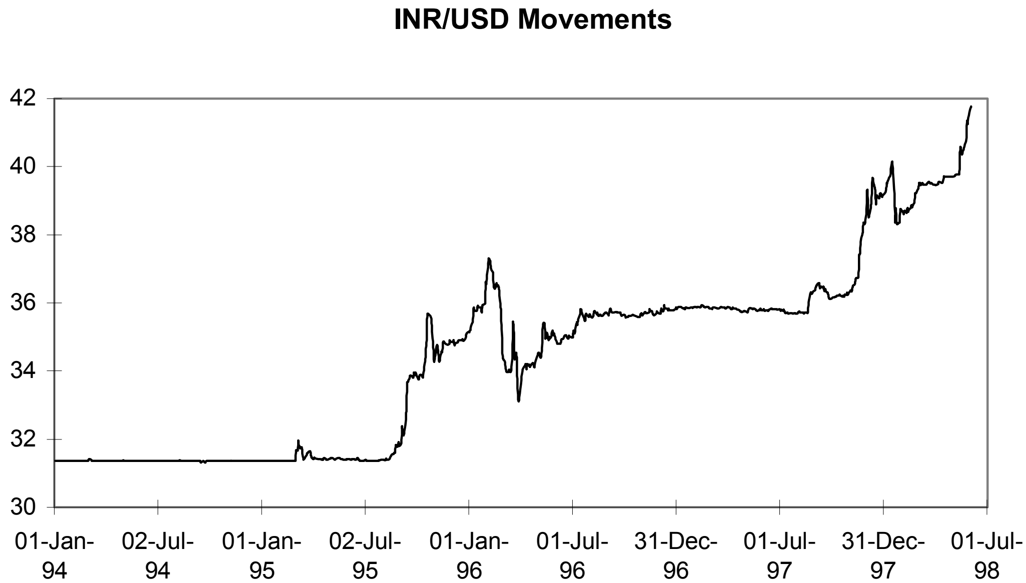
The volatility of exchange rate movements plays an important role in the pricing of foreign exchange options and in the measurement and management of risks arising out of open foreign exchange positions. In the case of the Indian rupee and many other emerging market currencies, the foreign exchange markets behave very differently from other financial markets because of the dominant role of the central bank, the presence of exchange controls particularly on capital account transactions and the structure of the market itself. Exchange rate movements in these currencies are characterised by long periods of placidity punctuated by abrupt and sharp changes. Many, but by no means all, of these sharp changes are currency depreciations. These peculiarities of the exchange rate process make the normal distribution an inappropriate model of exchange rate changes. Consequently, higher order moments of the distribution (particularly the skewness and the kurtosis) are as important as the variance (which is what is captured by the volatility). Since there is a great deal of evidence for regime shifts in the exchange rate process, modelling the time path of the volatility and the kurtosis becomes essential for accurate option pricing and risk measurement.

This paper studies the statistical process underlying changes in the rupee-dollar exchange rate and develops robust volatility and kurtosis prediction models that can be used both for pricing rupee-dollar options and for value at risk calculations on open rupee-dollar positions.

Exploratory overview of the data

Figure 1 shows the movement of the rupee-dollar exchange rate over the period January 1994 to June 1998. A cursory glance at this chart reveals four broad periods (i) the period up to August 1995 during which the rupee remained stable at the level of Rs 31.37 to the dollar; (ii) August 1995 to February 1996 when the rupee displayed high volatility while depreciating by over 10%; (iii) March 1995 to August 1997 during which the rupee remained stable at around Rs 35 to the dollar; and (iv) August 1997 to June 1998 when the rupee was subjected to renewed volatility and depreciation. A closer examination would reveal finer subdivisions within the two periods of volatility mentioned above. Clearly any useable statistical model of the currency movements must accommodate these widely varying exchange rate regimes.

Figure 1



Price Discreteness

In most studies of financial prices, it is harmless to regard prices as continuous variables capable of assuming arbitrary fractional values. In reality however, quoted prices are always discrete because of the minimum tick size conventions that operate in all financial markets. Stock prices in India, for example, typically move in multiples of 25 paise or 50 paise while in the United States, they move in multiples of 1/8 or 1/16 of a dollar. This discreteness is however usually important only when studying intra day price movements. In a study of daily closing prices, discreteness can be ignored because the typical daily fluctuation in share prices is several times the tick size.

In the foreign exchange markets, the reference rate published by the Reserve Bank of India does not contain any fractions of a paise; this implies a "tick size" of one paisa. (Market participants may sometimes quote rates including fractions of half a paisa; in many cases, however, these reflect half-paise spreads around a mid rate which is in integral paise). During periods of relative calm in the foreign exchange markets, a tick size of one paisa would induce extreme discreteness in rate changes. During the first half of 1997 for example, absolute changes in the exchange rate on most days were 0, 1, 2 or 3 paise.

What this means is that an adjustment for price discreteness is necessary while modelling exchange rate changes using a normal distribution or any other continuous distribution. Price discreteness does tend to be smeared out a bit when percentage changes in the exchange rate are computed. A one paise move from 35.85 is slightly different from a one paise move from 35.80 when both are converted into percentage changes. However, price changes of zero paise escape this smearing out because a zero paise change is zero percent whether the base is 35.85 or 35.80 or even 31.37. What this means is that the distribution of percentages changes in the exchange rate would show a large probability mass at 0. This atomic distribution poses serious problems while using maximum likelihood methods of estimation. Maximum likelihood would tend to favour a degenerate distribution with all the mass concentrated at zero. For example, if the distribution were modelled as a GED (generalised error distribution), maximum likelihood estimation drives the parameter ν (which controls the thickness of the tails) to its limiting value of zero. Similarly, if a mixture of normals were to be estimated by maximum likelihood, one of the distributions would be a degenerate normal distribution with mean and standard deviation both equal to zero which concentrates all its mass at zero. In each of these cases, one can use an *ad hoc* method to prevent degeneracy; for example, one can impose a lower limit on the standard deviation in a normal mixture or on the parameter ν of a GED. It is better however to solve the problem directly by adjusting for price discreteness.

The standard approach to discreteness is to regard the observed discrete price as the rounded value of a true[1] unobserved price which is continuous. The unobserved price can then itself be estimated by maximum likelihood or other means, but this can become quite complex when the statistical model includes time varying parameters. This paper therefore adopts the following simple discreteness adjustment which eliminates degeneracy at the cost of a slight loss of accuracy. The observed prices provide lower and upper bounds on the unobserved return. For example, if the observed prices move from 35.80 to 35.82, the lower bound on the percentage return corresponds to the unobserved prices moving from 35.805 to 35.815 and the upper bound corresponds to the unobserved prices moving from 35.795 to 35.825. In all maximum likelihood estimates in this paper, the likelihood is calculated at both of these bounds and the lower of the two values is taken as the likelihood for maximisation purposes. If we consider a normal (or any other symmetric) distribution, the likelihood defined in this way would be low when the mean of the distribution is far from the observed return or when the standard deviation of the distribution is very small in relation to the difference between the lower and upper bounds on the unobserved returns. The method thus penalises very low standard deviations and eliminates degeneracy while ensuring that means are estimated correctly.

The empirical distribution function of returns is required for estimating and plotting densities as well as for estimates based on matching fractiles. For these purposes, the empirical distribution function is computed by replacing the observed returns by a pair of returns corresponding to the lower and upper bounds on the returns. This doubles the apparent sample size but the actual sample size is used for significance tests, kernel bandwidth and other sample size dependent parameters.

Distribution of Exchange Rate Changes

Figure 2 shows the probability density function of the distribution of exchange rate changes estimated using a gaussian kernel[2] with a bandwidth of 0.25 standard deviations. (in this and all subsequent density plots, the units on the X axis are in terms of the historical standard deviation calculated over the full sample). Exchange rate changes (here and, unless otherwise stated, elsewhere in this paper) are defined to be changes in the logarithm of the exchange

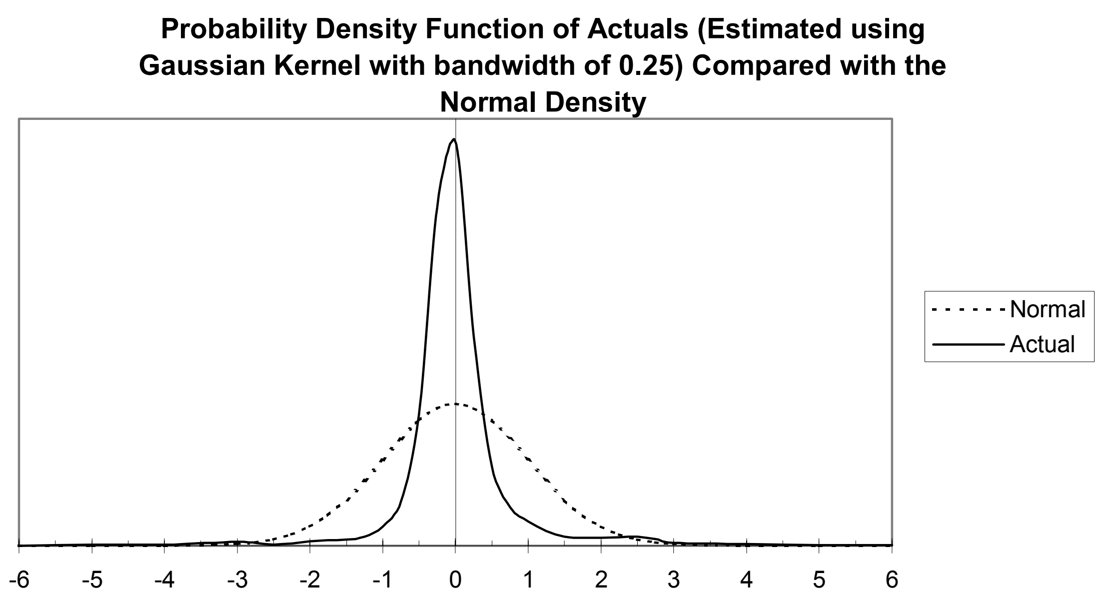
[1] Some authors object to the description of this unobserved price as the “true” price. They contend that the observed price is the true price. (Campbell, et al. 1997).

[2] For a description of kernel and other methods of density estimation, see Silverman (1986).

rate or equivalently as the logarithmic return $\ln(S_t / S_{t-1})$. The logarithmic return is widely used in finance in preference to the proportional return ($[S_t - S_{t-1}]/S_{t-1}$) because its distribution is more symmetric and therefore closer to normality. As can be seen the distribution is characterised by a thin waist, fat tails and a slight asymmetry induced by the hump in the right tail between 2 and 3 standard deviations above the mean. The summary statistics of the distribution are as follows:

Mean	0.03%
Median	0.00%
Standard Deviation (σ)	0.36%
Quartile Deviation x 0.7413 (this should equal the standard deviation for a normal distribution)	0.06%
Skewness	0.52
Excess Kurtosis (Excess of the kurtosis over the normal distribution value of 3).	21.00
Maximum	3.04% (= 8.4 σ)
Minimum	-2.73% (= -7.5 σ)

Figure 2



The extreme non normality of the distribution is evident from the large extreme values, the very high excess kurtosis, and the fact that 0.74 times quartile deviation is only one-sixth of the standard deviation while for an normal distribution it should be equal to the standard deviation.

GARCH Approaches to Modelling Exchange Rates

This section reports results obtained using the GARCH (Generalised Autoregressive Conditional Heteroscedasticity) approach (Bollerslev, 1986) to modelling the exchange rate.

This includes as a special case, the Exponentially Weighted Moving Average (EWMA) approach adopted in the RiskMetrics® methodology of J. P. Morgan.

The simple GARCH (1,1) model can be written as follows:

$$\begin{aligned} &\sigma_t^2 = \omega^2 + \beta\sigma_{t-1}^2 + \alpha r_{t-1}^2 \\ &r_t / \sigma_t \quad \sim \quad N(0,1) \text{ or more generally iid with zero mean \& unit variance} \end{aligned} \tag{1}$$

where r_t is the logarithmic return on day t (defined as $\ln(S_t/S_{t-1})$ where S_t is the exchange rate on day t), σ_t is the standard deviation of r_t, α and β are parameters satisfying $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$, $\alpha + \beta \leq 1$ and $\omega^2/(1 - \alpha - \beta)$ is the long run variance. This is the simplest GARCH model in that it contains only one lagged term each in σ and r and uses the normal distribution. More general models can be obtained by considering longer lag polynomials in σ and r and using non normal distributions.

Essentially, the GARCH model accommodates different exchange rate regimes by allowing the volatility of exchange rates to vary over time. It also postulates that a large change in the exchange rate (whether positive or negative) is likely to be followed by other large changes in subsequent days. This effect is captured by using the squared return to update the estimated variance for the next day. In fact the posterior variance σ_t^2 is a weighted average of three quantities: (i) the long run variance $\omega^2/(1 - \alpha - \beta)$ with weight $(1 - \alpha - \beta)$, (ii) the prior variance σ_{t-1}^2 with weight β and (iii) the squared return r_{t-1}^2 with weight α. The restrictions on the parameters α and β ensure that the weights are positive and sum to unity.

A special case of the GARCH model arises when $\alpha + \beta = 1$ and $\omega = 0$. In this case, it is common to use the symbol λ for α and Eq 1 takes the simpler form

$$\begin{aligned} &\sigma_t^2 = (1-\lambda)\,\sigma_{t-1}^2 + \lambda\; r_{t-1}^2 \\ &r_t / \sigma_t \quad \sim \quad N(0,1) \text{ or more generally iid with zero mean \& unit variance} \end{aligned} \tag{2}$$

The variance estimate can in this case be also interpreted as a weighted average of all past squared returns with the weights declining exponentially (λ, λ^2, λ^3 ...) as we go further and further back. For this reason, Eq. 2 is known as the Exponentially Weighted Moving Average (EWMA) model.

Initial estimation of both Eq 1 and Eq 2 using the normal distribution indicated significant non normality. They were therefore estimated using the Generalised Error Distribution (GED) which was popularised in financial econometrics by Nelson (1991). The GED with zero mean and unit variance and tail parameter ν ($0 < \nu$) is defined by the density:

$$g(x) = \frac{\nu \exp\left(-\frac{1}{2}\left|\frac{x}{\kappa}\right|^{\nu}\right)}{\kappa\, 2^{\frac{\nu+1}{\nu}}\, \Gamma\left(\frac{1}{\nu}\right)}$$

$$\kappa = \sqrt{\frac{2^{\frac{-2}{\nu}}\Gamma\left(\frac{1}{\nu}\right)}{\Gamma\left(\frac{3}{\nu}\right)}} \qquad (3)$$

The normal distribution is the special case where $\nu = 2$. Low values of ν imply fatter tails than the normal while higher values imply thinner tails. In most of the GARCH estimates, the tail parameter ν of the GED was close to 1 (as against the value of 2 for the normal distribution) implying significantly fatter tails than the normal distribution.

Estimation of the GARCH model (Eq 1) with GED residuals revealed several interesting features:

1. Unrestricted estimation led to positive estimates of both α (0.67) and β (0.58), but $\alpha + \beta$ was significantly above unity (1.25). However, this model is unacceptable since it has the absurd implication that σ_t grows without bound over time since $E(\sigma^2_{t+1}) = (\alpha + \beta)\sigma^2_t + \omega^2$. The log likelihood for this model was 5095.97 with $\nu = 1.05$.

2. When the constraint $\alpha + \beta \leq 1$ was imposed, the estimates of α and β were 0.789994 and 0.209835 respectively implying $\alpha + \beta$ equal to 0.999829 while ν was estimated to be 1.05. The estimate of ω^2 was 6.83E-08 implying a long run *standard deviation* of daily changes in exchange rates of about 2%[3]. The log likelihood for this model was 5080.85. Though the log likelihood ratio test rejected this model in favour of model 1 (χ^2 with 1 df = 29.4, P < 0.001%), model 1 is, as already stated, an unacceptable model from a conceptual point of view.

3. Imposition of the further restriction that $\omega = 0$ collapses the GARCH model to the EWMA model (Eq 2). The estimate of α (or λ) was 0.886649 and ν was estimated to be 1.04. The log likelihood for this model was 5013.86. Therefore, the log likelihood ratio test rejected this model in favour of model 2 above (χ^2 with 1 df = 134.0, P < 0.001%).

---

[3] This is about five times the sample standard deviation over the whole period and comparable to the maximum standard deviation observed at any point of time using the EWMA model (model 3 below). Model 2 was therefore re-estimated with the constraint that the long run standard deviation should equal the historical average of 0.361%. The estimates of α and β were now 0.200238 and 0.794701 respectively implying $\alpha + \beta$ equal to 0.994939 while ν was estimated to be 1.05. The log likelihood ratio test was unable to reject this model as against model 2 (χ^2 with 1 df = 2.54, P = 11%). This implies that the long run variance is very poorly estimated in the GARCH model. The large value of the long run variance is not therefore considered a sufficient reason to reject model 2.

Goodness of fit

Out of the above models, the first is patently unacceptable, and, therefore, the second and third were taken up for detailed study. These are referred to below as the GARCH-GED and the EWMA-GED models respectively. The performance of these models can be measured by examining the distribution of the standardised residuals r_t/σ_t.

	Expected Value based on		EWMA - GED	GARCH - GED
	Normal	GED ($\nu = 1$)		
Standard Deviation	1.00	1.00	1.48	1.14
Quartile Deviation	1.35	1.00	0.35	0.37
Skewness	0.00	0.00	10.22	3.83
Excess Kurtosis	0.00	2.71	192.79	48.76
Maximum			30.85	15.05
Minimum			-7.71	-9.66
Number beyond ± 5	≈ 0	1.00	10	9
Number beyond ± 6.5	≈ 0	0.10	7	5

The standardised residuals from both the EWMA-GED and the GARCH-GED models have markedly thinner waists and fatter tails than even the GED distribution with $\nu = 1$. They also have a very high degree of positive skewness. This indicates that both the EWMA-GED and the GARCH-GED models are failing to capture the true process driving changes in the exchange rate. This would suggest that the GARCH process would be unsuitable for valuation of options and other derivatives on the foreign exchange rate.

Value at risk

It follows from the above results that these models would be unsuitable for managing risks arising out of options and other derivative positions on the foreign exchange rate as these models would fail to even value them correctly. It is however possible that a GARCH and EWMA models may still be adequate for risk management of spot and forward open positions in certain situations. Risk management emphasises the calculation of Value at Risk (VaR) which is defined in terms of the percentiles of the distribution of asset values. Where the asset is an open position in foreign exchange spot or forward contracts, this reduces to the percentiles of the distribution of exchange rate changes. What the above numbers say is that the GARCH model would be grossly wrong in estimating extreme percentiles like the 0.1 or 0.01 percentile and also of percentiles close to the quartiles. However, most value at risk computations are based on the 1, 5 and 10 percentiles. At the 5% level, the GARCH and EWMA models fare reasonably level, but at the 1% and 10% levels, the performance is not satisfactory as seen from the following table.

	10% level two sided	5% level two sided	1% level two sided
Percentile as number of standard deviations[4] (using GED with $\nu = 1$)	1.65	2.15	3.25
Expected number of violations of VaR limit	101	50	10
EWMA-GED: Actual number of violations	71	43	22
EWMA-GED: Actual percentile	7.04%	4.26%	2.18%
EWMA-GED: Significance test of actual versus expected	Significant ($P \approx 0.2\%$)	Not significant	Significant ($P \approx 0.02\%$)
GARCH-GED: Actual number of violations	63	43	18
GARCH-GED: Actual percentile	6.24%	4.26%	1.78%
GARCH-GED: Significance test of actual versus expected	Significant ($P \approx 0.01\%$)	Not significant	Significant ($P \approx 1.25\%$)

At even lower risk levels like 0.50% and 0.25%, the models fare disastrously as shown below:

	0.50% level two sided	0.25% level two sided
Percentile as number of standard deviations (using GED with $\nu = 1$)	3.75	4.25
Expected number of violations of VaR limit	5	3
EWMA-GED: Actual number of violations	16	11
EWMA-GED: Actual percentile	1.59%	1.09%
EWMA-GED: Significance test of actual versus expected	Significant ($P < 0.01\%$)	Significant ($P < 0.01\%$)
GARCH-GED: Actual number of violations	11	10
GARCH-GED: Actual percentile	1.09%	0.99%
GARCH-GED: Significance test of actual versus expected	Significant ($P \approx 0.8\%$)	Significant ($P < 0.01\%$)

We are forced to conclude that the GARCH based approach is satisfactory neither for risk management nor for option pricing.

Models Based on Jump Processes and Normal Mixtures

We therefore turn to other models for the exchange rate process. Visual inspection of the movements of the exchange rate (Figure 1) suggests that there are occasional jumps in the exchange rate where the rupee/dollar rate moves by almost one rupee, while, on most days, the movement is only a few paise. This immediately suggests Merton's (1976) jump-diffusion model in which the exchange rate process has a diffusion component and a jump

[4] The cdf of the GED has to be calculated by numerical integration of the GED density. The percentile can then be obtained by a simple one dimensional search procedure or by Newton-Raphson iterations.

component. The diffusion component would produce changes in the exchange rate which are normally distributed with a fairly small standard deviation. Superimposed on this are jumps in the exchange rate driven by a Poisson process. Whenever the Poisson event occurs, there is a jump in the exchange rate, and the jump size is itself a random variable. The most tractable version of the Merton model is where the jump size is normally distributed. The diffusion process and the Poisson process are assumed to be probabilistically independent of each other.

For estimation purposes using daily data, it is convenient to approximate the Poisson jump process by a binomial jump process. In a Poisson jump process, there can theoretically be any number of jumps even in a short time period. However, if we consider a short time period Δt, the most likely event (with probability approximately $1-\lambda\Delta t$) is that there are no jumps and there is a small probability (approximately $\lambda\Delta t$) that there is exactly one jump where λ is the intensity of the Poisson process. The probability of more than one jump is of order $(\Delta t)^2$ and can be ignored if the time period Δt is short. This leads to the binomial approximation in which there is either no jump or exactly one jump.

The binomial jump model with a normally distributed jump size leads directly to a normal mixture for daily rate changes. If there is no jump, then the exchange rate change is derived from the diffusion component and is $\sim N(\mu_0, \sigma^2)$, i.e., it is normally distributed with a mean of μ and variance σ^2. (The mean μ_0 would be very close to 0 for daily returns - even a 10% annual depreciation would amount to less than 0.05% per day). If there is a jump, the exchange rate change is the sum of two normal variables - the first is $\sim N(\mu_0, \sigma^2)$ as above and the second is $\sim N(\mu_1,\delta^2)$ where μ_1 is the mean jump size and δ^2 is its variance. Therefore conditional on a jump, the rate change is normally distributed $\sim N(\mu_0+\mu_1, \sigma^2 + \delta^2)$. The unconditional distribution of rate changes is a mixture of two normal distributions $N(\mu_0, \sigma^2)$ and $N(\mu_0+\mu_1, \sigma^2 + \delta^2)$. The mixture weights (or probabilities) are the probability of no jump and the probability of one jump.

The normal mixture model can be extended to more than two normal distributions. In general, one can consider n normal distributions $N(\mu_i, \sigma_i^2)$ with weights p_i summing to 1. This can also be interpreted as a binomial approximation to a jump diffusion process with several Poisson jump processes which have different jump size distributions.

Static Normal Mixtures

Estimation of the normal mixture model requires estimation of the means and variances of each component distribution and of the mixture weights. These parameters can be estimated by maximum likelihood, but there are alternative estimation methods available. Hull and White (1998) argue that it is preferable to estimate a normal mixture by matching fractiles. Specifically, they recommend dividing the observations into several groups and computing the fraction f_i of observations falling in each group. Then they suggest forming the likelihood function $\Sigma f_i \log p_i$ where p_i is the probability that an observation falls in group i under a given vector of mixture parameters. The mixture parameters are chosen to maximise this likelihood.

The normal mixture model with three component distributions was estimated for the rupee-dollar exchange rate changes. Estimation was carried out using maximum likelihood as well as by matching fractiles and the results were qualitatively similar. The results reported here are based on maximum likelihood.

The three component distributions are as follows:

	Component A	Component B	Component C
Standard Deviation	1.20%	0.13%	0.39%
Mean	0.03%	0.00%	0.18%
Mixing Probability	11.27%	67.26%	21.47%

Component B is the base distribution (representing the diffusion process) and accounts for the behaviour of the exchange rate change on more than two-thirds of all days. However, the two components arising out of the jump process are very important: component A accounts for more than 75% of the variance of the mixture and component C for more than 90% of its mean. A and C thus represent two very different kinds of jumps in the exchange rate. A involves large jumps (a mean jump size[5] of 40 paise at an exchange rate of Rs 42 to the dollar) but these jumps are almost equally likely to be upward jumps (probability of 51%) or downward jumps (probability of 49%). C on the other hand involves more modest jumps (a mean jump size of 17 paise at an exchange rate of Rs 42 to the dollar) but these jumps are about twice as likely to be upward jumps (depreciations of the rupee) as downward jumps.

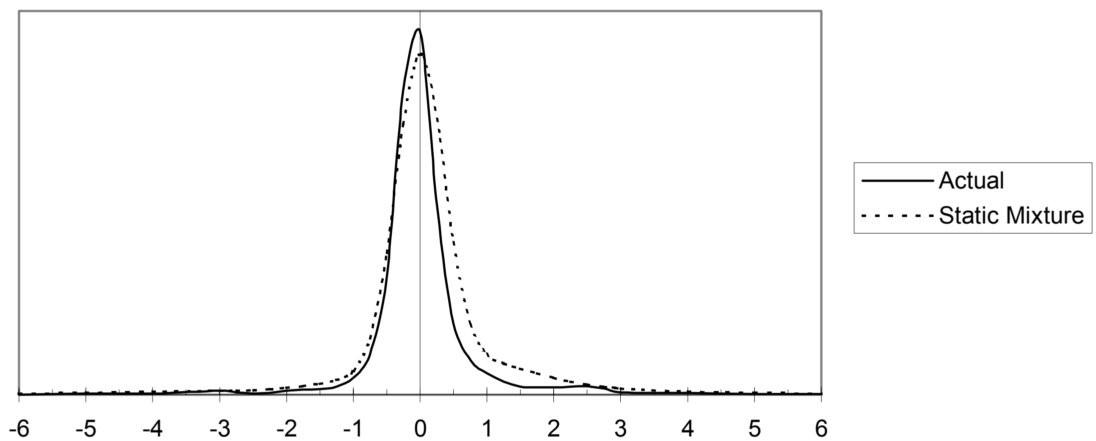
So the general picture that emerges is as follows: most of the time, the rupee remains steady against the dollar with rate changes of a few paise a day. Superimposed on these are days when the rupee undergoes depreciation in modest jumps with occasional slightly smaller corrections[6]. More rarely, the rupee enters periods of high volatility when the rupee moves upwards or downward by nearly half a rupee a day. The rupee depreciations of this order often happen when there is panic in the market, while rupee appreciations of this order are often the result of decisive intervention by the Reserve Bank.

[5] The mean jump size is the expectation of the absolute value of the jump. This can be easily computed using the truncated means of a normal distribution.

[6] The expected depreciation is the conditional expectation of the absolute jump size given that the jump is positive (about 18 paise) while the expected correction is the conditional expectation of the absolute jump size given that the jump is negative (about 15 paise). Again, these values are computed using the truncated means of a normal distribution.

Figure 3

Probability Density Function of Static Normal Mixture Distribution Compared with Density of Actuals (Estimated using Gaussian Kernel with bandwidth of 0.25)



The summary statistics of the mixture are compared with that of the actual data in the following table, while Figure 3 shows a plot of the mixture density compared with the estimated density of the actual data. It is evident that though the normal mixture captures the important qualitative characteristics of the empirical distribution it still has lower skewness, fatter waists and thinner tails than the actual data.

	Actual data	Normal Mixture	Pure Normal
Mean	0.03%	0.04%	0.03%
Median	0.00%	0.02%	0.03%
Standard Deviation (σ)	0.36%	0.46%	0.36%
Quartile Deviation x 0.7413 (this should equal the standard deviation for a normal distribution)	0.06%	0.18%	0.36%
Skewness	0.52	0.09	0.00
Excess Kurtosis (Excess of the kurtosis over the normal distribution value of 3).	21.00	13.01	0.00
Number beyond ±2% (approximately 6 σ) in a sample of 1000	7	11	0.00003

Dynamic Normal Mixtures

For many purposes, the normal mixture estimated above may be a reasonable approximation to the exchange rate process and one could try to improve the fit even better by using more than three normal distributions. However, the static normal mixture would differ sharply from the actual return generating process in one very important manner. It is visually apparent from Figure 1 that the exchange rate process is not constant over time. As already pointed out, one can readily identify periods of turbulence and periods of placidity in the movement of the exchange rate. The GARCH and EWMA models discussed above were motivated by this fact and attempted to model these shifts in the exchange rate regime. The natural analogue of these models in the context of a jump-diffusion model would be to allow the jump probability to vary over time. Periods of placidity arise when the jump probability is very low and turbulent periods arise when the jump probability is higher. In the context of normal mixtures, we should allow the mixing probabilities (or weights) to vary over time.

One could think of several different ways in which to model time varying mixing probabilities. This paper uses a method which is inspired by Bayesian ideas. After observing the rate change on any day, we can use Bayes theorem to estimate the probability that a jump occurred on that day. Bayes theorem says that

$$P(J_t \mid X_t) = \frac{P(X_t \mid J_t)\, P(J_t)}{P(X_t)}$$

If J_t denotes the occurrence of a jump today and X_t denotes the observed rate change today, Bayes theorem says that the posterior probability that a jump occurred today ($P(J_t|X_t)$) is obtained by multiplying the prior probability of the jump ($P(J_t)$) by the ratio of the likelihood of the observation conditional on a jump ($P(X_t|J_t)$) to the unconditional likelihood of the observation ($P(X_t)$).

What can we conclude about tomorrow if the posterior probability turns out to be high? Under the static model discussed earlier, the answer is nothing. The Poisson process is memoryless, and the fact that a jump occurred today tells us nothing about whether a jump will or will not occur tomorrow. However, the time varying model that we are discussing in this section takes a different view. It says that there are different exchange rate regimes in some of which jumps are more common that in others. The fact that a jump occurred today suggests that we are in a regime of frequent jumps and that the probability of a jump tomorrow is higher than the prior probability.

This means that we must have a mechanism for changing the prior probability of a jump tomorrow on the basis of the posterior probability that a jump occurred today. To model this, we take our cue from the GARCH model where the prior variance for tomorrow is modelled as a weighted average of three quantities: (i) the long run variance, (ii) the prior variance for today and (iii) the squared return observed today. Analogously, we postulate that the prior probability for a jump tomorrow is equal to a weighted average of three quantities: the long run probability of a jump (taken from the static mixture), the prior probability of a jump today and the posterior probability that a jump occurred today:

$$P_t(J_{t+1}) = (1-\alpha' - \beta')\, P^*(J) + \beta'\, P_{t-1}(J_t) + \alpha'\, P_t(J_t) \qquad 0 \le \alpha', \beta' \le 1 \qquad (4)$$

where $P^*(J)$ is the long run probability of a jump as estimated in the static mixture, and the subscript t in P_t means that the probability is conditional on all the information available at the end of day t including the exchange rate change on day t. The parameters α' and β' have to be estimated by maximum likelihood.

GARCH-like Taper

There is one additional complexity to be taken care of in order to complete the model. The normal mixture model with time varying mixing probabilities is constrained to produce a standard deviation which lies within the range of standard deviations of the three mixing distributions. In our example, the lowest possible standard deviation is 0.13% which results in the limiting case when the prior probability of the jumps becomes zero and the mixture coincides with component B. Similarly, the highest possible standard deviation is 1.29% which is achieved when the probability of component A goes to unity. These are not acceptable restrictions from a risk management point of view. There are extended periods of time when the standard deviation of changes in the exchange rate are very much smaller than 0.13%. The upper bound of 1.29% is not too strongly violated in the period under study, but it may well be breached in future. It is worth recalling that during the East Asian crisis in 1997, the daily standard deviation of changes in currencies like the Indonesian rupiah might have reached 10%.

The GARCH model is more flexible in this respect as it is capable of driving standard deviations as high or as low as necessary unless the weight on the long run variance, $(1 - \alpha - \beta)$, is very high. We would therefore like the dynamic mixture model to taper off into a GARCH-like process when the variances become too small or too large. In other words, in addition to adjusting the mixing probabilities, we must also be willing to adjust the variances of the mixing distributions. However, we should do this only at the two ends of the spectrum when the adjustment of probabilities has gone far enough. We achieve this goal without introducing too many free parameters as follows. We postulate a GARCH like model for the variance of Component A:

$$\sigma_{A,t}^2 = (1-\alpha''-\beta'')\,\sigma_A^2 + \beta''\,\sigma_{A,t-1}^2 + \alpha''\,r_t^2 \qquad 0 \le \alpha'', \beta'' \le 1$$

Instead of letting α'' and β'' be free parameters, we constrain them as follows. First we enforce the condition that α''/β'' should always equal α'/β'. Second, we note that if (1- $- \alpha'' - \beta''$) is kept close to unity, the GARCH like model for Component A variance is virtually shut off and the variance remains close to the static estimate σ^2_A. On the other hand the influence of σ^2_A is virtually eliminated when we drive $(1 - \alpha'' - \beta'')$ close to zero. What we therefore need is a mechanism for (i) making $(1 - \alpha'' - \beta'')$ close to zero when the mixing probability $p_{A,t}$ is very high and (ii) driving it close to unity when $p_{A,t}$ is close to or below the static probability p_A. The logistic function $\frac{1}{1+ae^{bt}}$ ($a > 0$) is an ideal candidate for providing a smooth bridge between these two values of zero and unity. This function attains the values of 0 and 1 only at $\pm\infty$, but the parameters a and b can be so chosen that the function attains the values of say 0.999 at $p_{A,t} = 1$ and say 0.001 at $p_{A,t} = 2p_A$. Since α'' and β'' are completely determined if we fix α''/β'' and $(1 - \alpha'' - \beta'')$, this completes the model for σ^2_A. An entirely analogous model is formulated for σ^2_B.

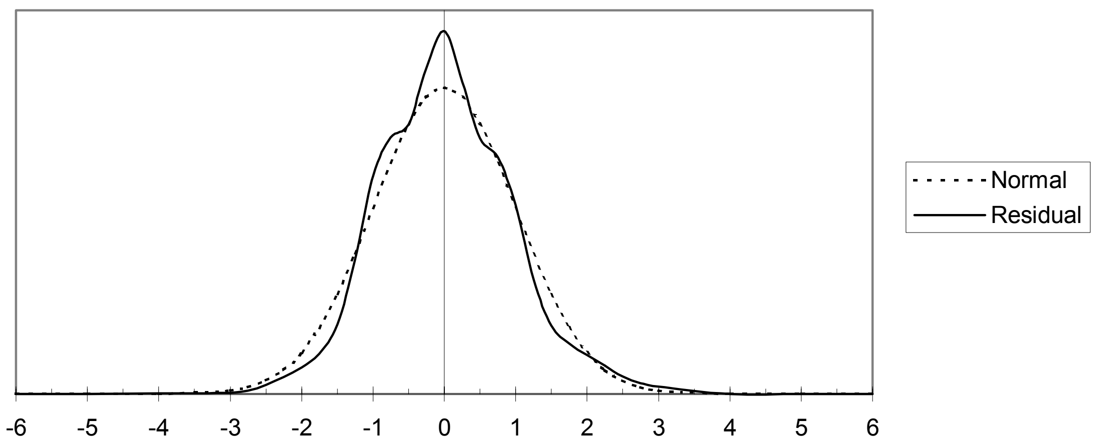
Goodness of Fit

The maximum likelihood estimates of α', β' and $1 - \alpha' - \beta'$ were 0.5807, 0.3823 and 0.0370 respectively. The fairly low value of $1 - \alpha' - \beta'$ allows the dynamic mixing probabilities to deviate quite substantially from the static probabilities. Nevertheless, the value is large enough to produce a time path of mixing probabilities which is qualitatively different from what would result if it were arbitrarily set equal to zero. In particular, the non zero value of $1 - \alpha' - \beta'$ keeps the mixing probabilities bounded away from zero and unity, while if it were zero, prolonged periods of calm would drive the jump probabilities to near zero levels.

In case of the GARCH and EWMA models, one could measure how well the model describes the data by comparing the standardised residuals to a reference distribution (normal or GED). In the time varying mixture, we can do something similar but the transformation that is to be done to the rate change is more complex than dividing by a time varying standard deviation. We can compute $N^{-1}(G_t(X_t))$ where N denotes the cumulative distribution function (cdf) of the normal distribution and G_t denotes the cdf of the normal mixture for day t using data up to and including day t-1. Standardisation of residuals which was done in case of GARCH and EWMA is similar to this process because in these cases, G_t differs from the reference distribution only by a scale factor (standard deviation).

Figure 4

Probability Density of Residuals from Dynamic Normal Mixtures
(Estimated using Gaussian Kernel with bandwidth of 0.30)
Compared with Normal Density



On doing this transformation, we find the following picture which is far better than the corresponding picture for the GARCH and EWMA models. In fact, on most counts, the distribution is extremely close to the normal distribution. This is confirmed by the plots of the estimated probability density function of the normalised residuals $N^{-1}(G_t(X_t))$ in Figure 4.

	Expected Value (Normal distribution)	Dynamic Normal Mixture Residuals $N^{-1}(G_t(X_t))$
Standard Deviation	1.00	0.91
Quartile Deviation	1.35	1.28
Skewness	0.00	0.23
Excess Kurtosis	0.00	0.53
Maximum		3.17
Minimum		-3.65
Number beyond ± 3.29	1	1
Number beyond ± 3.89	0.1	0

Value at risk analysis

The table below shows the results of a value at risk analysis at the 10%, 5% and 1% levels using the dynamic normal mixture model[7]. The model does exceedingly well at 5% and 1% levels. Even at the 10% level where the model is over-conservative, the performance is significantly better than that of the GARCH-GED model.

	10% level two sided	5% level two sided	1% level two sided
Expected number of violations of VaR limit	101	50	10
Actual number of violations	74	41	8
Actual percentile	7.33%	4.06%	0.79%
Significance test of actual versus expected	Significant ($P \approx 0.5\%$)	Not significant	Not significant

What is more interesting is that the model does exceedingly well at risk levels of 0.5% and 0.25% while the GARCH-GED models fared disastrously at these levels:

	0. 50% level two sided	0.25% level two sided
Expected number of violations of VaR limit	5	3
Actual number of violations	8	4
Actual percentile	0.79%	0.40%
Significance test of actual versus expected	Not significant	Not significant

[7] The VaR limits have to be computed by numerical methods each day. The mixture cdf is a weighted average of the component normal cdfs and can therefore be computed easily given any algorithm for computing the normal cdf. Since the mixture density is available in closed form (weighted average of the component normal densities), we can compute the mixture percentiles by Newton-Raphson iterations. The computational cost of doing this on a personal computer using a spreadsheet macro is quite modest.

Figure 5

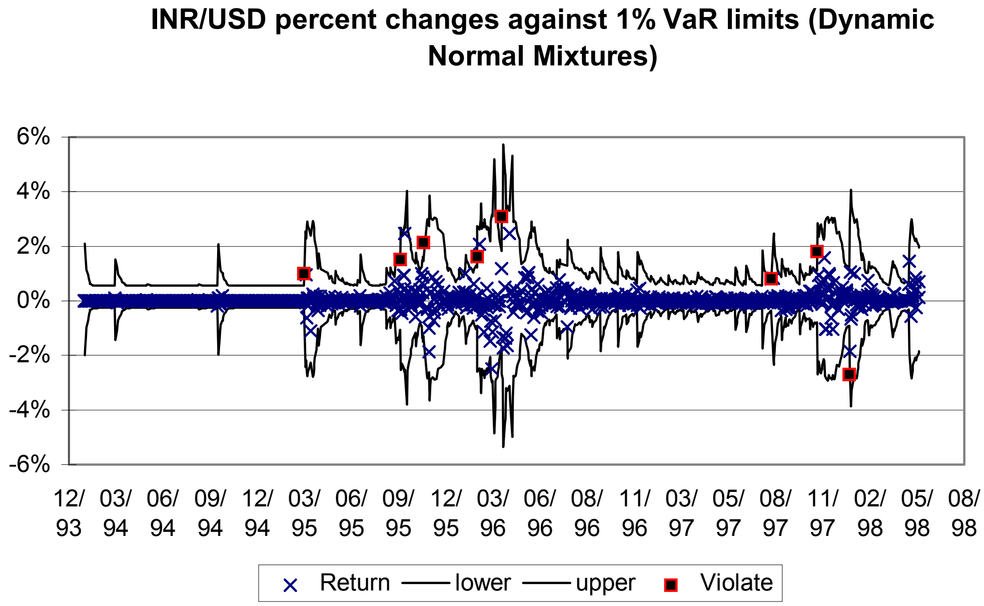


Figure 5 shows the performance of the 1% Value at Risk limits based on this model. During the periods of calm in the markets, the VaR limits are quite narrow, but they widen very dramatically during periods of turbulence. As already stated, the number of violations of the VaR limits (which are indicated by solid squares) are within the expected range. Most of the violations appear to be on the positive side, but this is partly due to the positive skewness of the distribution. More importantly, the downward jumps in the INR/USD rate (rupee appreciation) tend to happen only during periods of turbulence when the VaR limits are very wide. The truly unexpected jumps (which happen during periods of relative calm) are almost all rupee depreciations (upward moves in the INR/USD rate)

Kurtosis Shifts

The analysis of the time path of the moments of the dynamic normal mixtures revealed a very interesting phenomenon which explains why this model does much better than the traditional GARCH model. In times of calm in the foreign exchange markets, all that the GARCH model can do is to drive the standard deviation of the price down to very low levels (as low as 0.07% compared to the full-sample historical standard deviation of 0.36%) reflecting the fact that most price moves have been very small. In the same situation, the dynamic normal mixture also pushes the standard deviation down but it not as far down as GARCH (the minimum to which it goes is about 0.12%). However, the dynamic mixture model can and does alter the higher moments of the distribution very sharply and effectively: there are periods in which the kurtosis goes above 200 (see table below).

Time path of moments of dynamic normal mixture

	Mean	Standard Deviation	Skewness	Excess Kurtosis
Minimum	0.00%	0.12%	-0.30	0.18
Maximum	0.16%	2.02%	1.32	218.11
Median	0.01%	0.19%	0.76	44.43
Mean	0.03%	0.31%	0.71	91.87

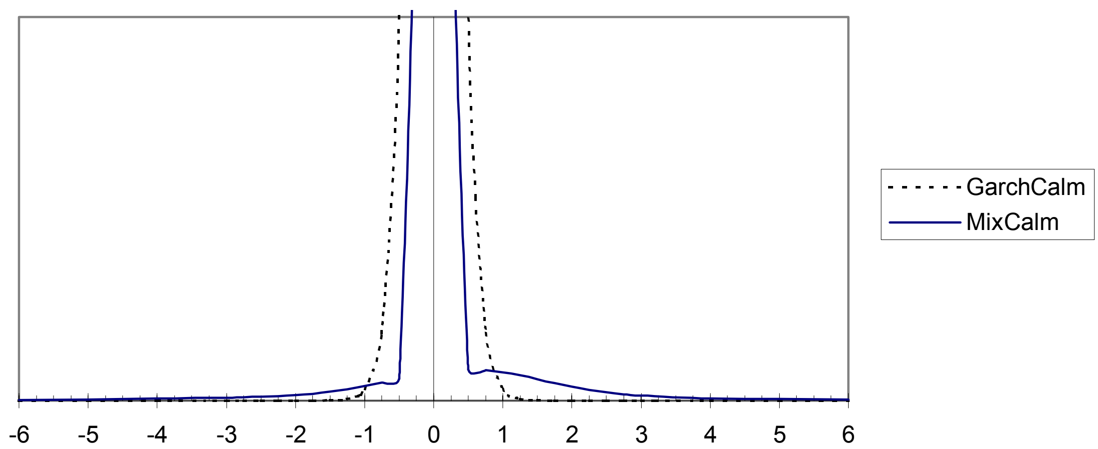
The dynamic mixture model accommodates periods of calm by lowering the standard deviation, raising the kurtosis and increasing the skewness. A plot of the densities of the GARCH model and of the dynamic mixture model in a period of extreme calm (Figure 6) is very revealing. In the band of ± 1 full-sample historical standard deviation, the dynamic mixture model has a narrower (and also taller[8]) peak than the GARCH-GED and is therefore in closer conformity to the recent pattern of constant or nearly constant exchange rates. In this range, the dynamic mixture has a more narrowly concentrated density than the GARCH-GED. The GARCH-GED can achieve this degree of narrow concentration only by reducing the standard deviation even lower than it has already done.

Yet, beyond ± 1 full-sample historical standard deviation, the picture is dramatically different. The mixture has a far thicker tail than the GARCH-GED; furthermore the right tail is noticeably thicker. The dynamic mixture is saying that though there is a very high probability that the exchange rate will remain virtually unchanged the next day, there is a small probability of a sharp downward jump in the value of the rupee. This is reflected in a low standard deviation, a high kurtosis and a high positive skewness of the distribution. The GARCH-GED can change the standard deviation but cannot dynamically adjust skewness and kurtosis. In other words, GARCH-GED leaves the shape of the distribution unchanged, and only changes the scale. It can therefore match the tails of the dynamic mixture only by increasing the standard deviation very substantially, but this would make the peak even shorter and wider than it already is. What GARCH-GED gives us is a compromise value reflecting a trade-off between these two conflicting goals. The mixture is far more flexible in that it can produce both a tall and narrow peak and a thick asymmetric tail by dynamically adjusting variance, skewness and kurtosis.

[8] This is not apparent from the figure since both the peaks has been lopped off to reveal the tails more clearly. The taller peak is found between ± 0.25 full-sample historical standard deviations where the mixture density lies above the GARCH-GED. The narrowness of the peak is evident between 0.25 and 1 standard deviations where the mixture density lies below the GARCH-GED.

Figure 6

Probability Density Function of Dynamic Mixture Distribution Compared with Density of GARCH-GED in Calm Markets (early 1994)



Importance of the GARCH-like tapers

For some purposes, the GARCH-like tapers are not very important. For example, the value at risk analysis at 1% or lower risk levels is quite satisfactory without the GARCH-like tapers, but the model is intolerably over-conservative at the 10% or higher levels. The quartiles are even more badly off and the overall goodness of fit is unsatisfactory. Quite apart from this, the GARCH-like tapers are essential from a theoretical point of view.

Implications for Option Pricing

The empirical results show that a jump-diffusion (or normal mixture) model (at least in its dynamic version) provides an excellent description of the exchange rate process and that this model performs far better than GARCH models that do not allow for jumps. This suggests that the valuation of options on the rupee-dollar exchange rate must be based on the jump-diffusion model of Merton(1976). .

Merton's model is based on the calculation of Black-Scholes option values conditional on there being exactly *n* jumps during the life of the option. These Black-Scholes option values for different values of *n* are then weighted with the probabilities of these many jumps from the Poisson distribution and the resulting weighted average is shown to be the true value of the option. Merton's model is based on one jump process and corresponds to a mixture of two normals, but the model can be readily generalised to the case of two independent jump processes giving rise to a mixture of three normals.

Consider an asset price process in which there is a diffusion process with variance σ^2, and there are two independent jump processes with probabilities λ_1 and λ_2, and the jumps are normally distributed with means μ_1 and μ_2 and variances δ_1^2 and δ_2^2. Suppose that the current asset price is *S*, and that the risk free interest rate is r. Consider a call option with an exercise price of X. The Merton model generalised to two jump process asserts that the value V of this option is:

$$
\begin{aligned}
&V = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P_{\lambda_1' t}(m) P_{\lambda_2' t}(n) f_{m,n} \\
&\lambda_i' = \lambda_i (1+\mu_i) \qquad (5) \\
&f_{m,n} = W\left(S, X, r - \lambda_1\mu_1 - \lambda_2\mu_2 + \frac{m\ln(1+\mu_1)}{t} + \frac{n\ln(1+\mu_2)}{t}, \sigma^2 + \frac{m\delta_1^2}{t} + \frac{n\delta_2^2}{t}, t \right)
\end{aligned}
$$

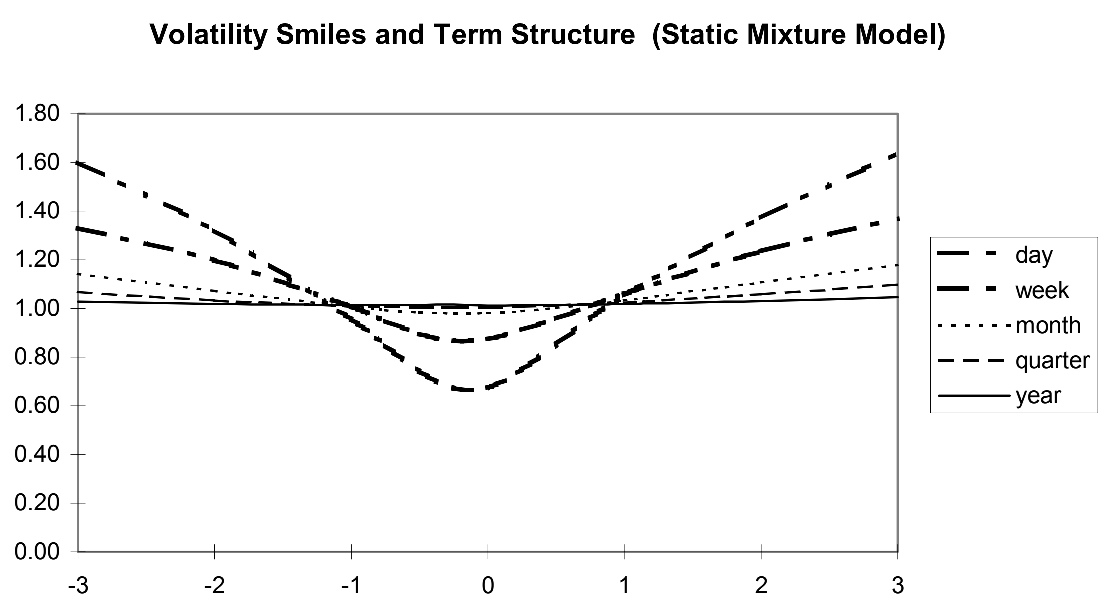
where $W(S, X, r, \sigma^2, t)$ is the Black-Scholes value of an option when the asset price is S, the exercise price is X, the risk free rate is r, the variance is σ^2 and the time to expiry is t. In practice the infinite summation over m and n would be truncated at levels M and N such that the respective cumulative Poisson probabilities are very close to 1.

Merton Model for Static Mixtures

To see what difference the mixture distributions make to option pricing, the option prices from this model were computed numerically for the static mixture distribution for a variety of exercise prices and maturities. To compare these prices with the naive Black-Scholes process, we use the notion of volatility smiles. Practitioners tend to think of option prices in terms of implied volatilities computed from the Black-Scholes price. For any option price V (observed in the market or computed from an alternative option pricing model), one can calculate an implied volatility v such that the Black-Scholes model with σ replaced by v gives a price of V. In other words, v is implicitly defined by

$W(S, X, r, v^2, t) = V$

Figure 7



Plots of v for different exercise prices are known as volatility smiles, and there is a different smile for each maturity. (The dependence of v on maturity is known as the term structure of volatility). If the Black-Scholes model were correct, the volatility smiles would be horizontal straight lines equal to σ throughout. Figure 7 shows the volatility smiles[9] (in the form of plots of v/σ) for the exchange rate process computed using the Merton model. It is apparent that the smiles are quite pronounced for maturities of one month or shorter, but are less significant at longer maturities. This suggests that the Black-Scholes[10] model is quite inappropriate for valuing exchange rate options for maturities of one month or less.

At longer maturities, however, the Black-Scholes model appears to be a reasonable model. However, this result must be treated with caution since the empirical results showed that while the static normal mixture is a very good model of the tails of the distribution, it is less than satisfactory in the waists of the distribution. A more complete and accurate model of option valuation must use the dynamic mixture model.

Merton Model for Dynamic Mixtures

When we use dynamic mixture models to calculate option prices, the relevant comparison is with the Black-Scholes prices using volatility estimates derived from the GARCH model. During periods of calm in the foreign exchange markets, the difference between these two prices is strikingly large. Consider, for example, a call option with an exercise price 1% above the current rate and a maturity of one day. This is about three full-sample historical standard deviations out of the money. In early 1994[11], according to the GARCH model, the volatility of the exchange rate was about 0.07%, the exercise price was about 15 standard deviations out of the money and the option was practically worthless. According to the Merton model based on dynamic mixtures, the option value is small but not negligible. The implied volatility corresponding to the Merton value of the option is about six times the GARCH estimate of the volatility and over three times the volatility estimated by the dynamic mixture. Similarly, in early 1997[12], the same option had a Merton value corresponding to an implied volatility five times the GARCH estimated volatility of 0.09% or about three times the dynamic mixture estimated volatility of 0.143%. On both occasions, it is the high kurtosis that makes the deep out of the money option so valuable. On the same two days, an at the money option is priced at implied volatilities which are somewhat lower than the GARCH estimated volatilities. The implied volatilities for these at the money

---

[9] These smiles are plotted for maturities of a day, week, month, quarter and year. The exercise prices for different maturities are chosen to correspond to the same number of standard deviations from the prevailing exchange rate. In other words, a fixed set of numbers k_i were chosen (-3, -2, -1, -0.75, -0.50, -0.25, 0, 0.25, 0.50, 0.75, 1, 2, 3) and the X_i for each t were chosen so that $\ln(S/X_i)$ was equal to $k_i \sigma\sqrt{t}$. This makes the smiles comparable across all maturities. For simplicity, we interpret S as the forward exchange rate rather than the spot rate and set r equal to 0. This eliminates the need to use the Garman-Kohlhagen (1983) model of exchange rate options and makes option prices practically insensitive to the interest rate.

[10] In this context, the Black-Scholes model is actually the Black-Scholes-Garman-Kohlhagen model. See footnote 9 above.

[11] February 3, 1994 to be precise

[12] February 20, 1997 to be precise.

options are only about half the volatilities estimated from the dynamic mixtures because the high kurtosis (thin waists) makes at the money options less valuable.

These results are in accordance with our intuitive expectations. Given the evidence for sudden jumps in the exchange rate process that we see visually in Figure 1, a deep out of the money option is potentially quite valuable because there is always a non trivial probability that the process will jump sharply enough to bring the option into the money.

If one were to assume that the dynamic mixture observed in say early 1994 would hold good for the next one year, one could use this to value options with a maturity of one year. Consider a one year call option with a strike price 1% above the current rate (approximately 0.2 annualised historical standard deviations). The Merton model values this at an implied of nearly twice the GARCH estimated standard deviation, but approximately the same as the mixture standard deviation. When the exercise price is 20% out of the money (approximately 3.5 annualised historical standard deviations), the implied is nearly thrice the GARCH estimated standard deviation, and about 1.6 times the mixture standard deviation. Of course, the assumption that the mixture would remain the same for the next one year is patently wrong given the reversion to the static mixture probabilities inherent in Eq. 4. A more complex simulation model is required to value this option correctly, but the above numbers show that the departures from Black-Scholes values are likely to be quite pronounced even for as long a maturity as one year.

These results suggest that the Black-Scholes model is even more inappropriate for pricing foreign exchange options than was suggested by the moderate smiles calculated from the static mixtures. It is quite likely that there are periods of calm in which the Black-Scholes model is inappropriate even at the longer maturities at which it appeared satisfactory when we considered the static normal mixture.

Conclusion

The empirical results very convincingly demonstrate the need to model the exchange rate process as a mixed jump-diffusion process. Sophisticated econometric models like the GARCH with GED residuals which perform quite well in other financial markets fail quite miserably in the case of the INR-USD process because they do not allow for jumps in the exchange rate. Equally importantly, the empirical results provide strong evidence that the jump probabilities are not constant over time. From a statistical point of view, changes in the jump process induce large shifts in the kurtosis of the process. The results suggest that the failure of GARCH processes arises because they allow for changes in volatility but not for changes in kurtosis. The dynamic mixture models do precisely that; they accommodate regime shifts by allowing both volatility and kurtosis (not to mention skewness) to change. This also shows that the periods of calm in the exchange rate are extremely deceptive; in these periods, the variance of rate changes is quite low, but the kurtosis is so high (in the triple digit range) that the probability of large rate changes is non trivial. The empirical results also show that the Black-Scholes-Garman-Kohlhagen model for valuation of currency options are quite inappropriate for valuing rupee-dollar options and that the Merton jump-diffusion model is the model of choice for this purpose.

References

Bollerslev, T. (1986), "Generalised Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics,* 31, 307-327

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets,* Princeton, Princeton University Press.

Garman, M. B., and Kohlhagen, S. W. (1983), "Foreign Currency Option Values", *Journal of International Money and Finance,* 2, 231-237.

Hull, J. and White, A. (1998), "Value at Risk When Daily Changes in Market Variables Are Not Normally Distributed", *The Journal of Derivatives,* Spring 1998, 9-19

Merton, R. C. (1976), "Option Pricing When Underlying Stock Returns Are Discontinuous ", *Journal of Financial Economics*, 3, 125-144

Nelson, D. (1991), "Conditional Heteroscedasticity in Asset Returns: A New Approach", *Econometrica*, 59, 347-370

Silverman, B. W. (1986), *Density Estimation for Statistical and Data Analysis,* London, Chapman and Hall.