Value at risk for a mixture of normal distributions: The use of quasi-Bayesian estimation techniques

Subu Venkataraman

Rapid globalization, innovations in the design of derivative securities, and examples of spectacular losses associated with derivatives over the past decade have made firms recognize the growing importance of risk management. This increased focus on risk management has led to the development of various methods and tools to measure the risks firms face.

One popular risk-measurement tool is value at risk (VaR), which is defined as the minimum loss expected on a portfolio of assets over a certain holding period at a given confidence level (probability). For example, consider a trader who is concerned about the risk, over the next ten days, associated with holding a specific portfolio of assets. A statement that, at the 95 percent confidence level, the VaR of this portfolio is \$100,000 implies that 95 percent of the time, losses over the 10-day holding period should not exceed \$100,000 (or losses should exceed \$100,000 only 5 percent of the time).

The use of value at risk techniques in risk management has exploded over the last few years. Financial institutions now routinely use VaR techniques in managing their trading risk and nonfinancial firms have started adopting the technology for their risk-management purposes as well. In addition, regulators are beginning to design new regulations around it. Examples of these regulations include the determination of bank capital standards for market risk and the reporting requirements for the risks associated with derivatives used by corporations.

Proponents of VaR argue that the ability to quantify risk exposure into a single number represents the single most powerful advantage of the technique.[1] Despite its simplicity, however, the technique is only as good as the inputs into the VaR model.[2] Many implementations of VaR assume that asset returns are normally distributed. This assumption simplifies the computation of VaR considerably. However, it is inconsistent with the empirical evidence of asset returns, which finds that asset returns are *fat tailed*. This implies that extreme events are much more likely to occur in practice than would be predicted based on the assumption of normality. Take, for example, the stock market crash of October 1987. The assumption of normality would imply that such an extreme market movement should occur only once in approximately 5,900 years. As we know, however, there have been worse stock crashes than that of October 1987 even in this century. This suggests that the normality assumption can produce VaR numbers that are inappropriate measures of the true risk faced by the firm.

While alternative return distributions have been proposed that better reflect the empirical evidence, any replacement for the normality assumption must confront the issue of the simplicity of computations, which is one of the

Subu Venkataraman is a senior economist in the research department at the Federal Reserve Bank of Chicago. The author would like to thank Mike Sterling for exceptional research assistance.

primary benefits of VaR. In this article, I examine one such alternative assumption that simultaneously allows for asset returns that are fat tailed and for tractable estimations of VaR. This distribution, based on a mixture of normal densities, has also been proposed by Zangari (1996). First, I relate the *mixture of distributions* approach to alternatives that have been presented in the academic literature on the stochastic processes governing asset returns. Second, I use an estimation technique for the parameters of the mixture of distributions that is computationally simpler than the techniques suggested by Zangari—the quasi-Bayesian maximum likelihood estimation (QB-MLE) approach (first suggested by Hamilton, 1991).[3] Third, using simulated data, I show that the QB-MLE combined with the mixture of normals assumption provides better measures of value at risk for fat-tailed distributions (like the Student's *t*) than the traditional normality assumption. I then establish that the technique does not suffer from the problems associated with the traditional maximum likelihood approach and that it is effective in recovering the parameters from simulated data.

Finally, this methodology is applied to foreign exchange data for eight currencies from 1978 to 1996. It is well known that returns in the foreign exchange market show dramatic violations of the assumption of normality by exhibiting fat tails (Jorion, 1995). I compute VaR estimates under both the assumption of normality and the mixture of normals approach for each of the eight currencies. I show that the mixture of normals assumption combined with QB-MLE outperforms the traditional normality assumption. First, the traditional normality assumption leads to a significantly larger number of violations of VaR than the mixture of normals. Moreover, the number of violations of VaR observed over the sample period under the QB-MLE is consistent with the stated goals of VaR.

To evaluate the performance of portfolio VaRs (as in Hendricks, 1996), I examine how information on the parameters governing individual currencies can be aggregated in the context of portfolios of these currencies. In contrast to the normality assumption, however, the use of the mixture of normals complicates this aggregation considerably. I propose a specific algorithm for computing portfolio statistics from the individual components that keeps the analysis computationally simple. The effectiveness of the approximations underlying this algorithm is judged by examining the magnitude of violations from simulated portfolios of currencies. Again, I compare the results under the QB-MLE approach with the normality-based results and the expected outcomes. I find that, despite the simplifying aggregation assumptions, the QB-MLE technique again outperforms the normality-based approach and provides VaR estimates consistent with what one would expect. These results suggest that combining the mixture of normals approach and the QB-MLE estimation technique allows us to capture fat-tailed distributions, while maintaining a computationally tractable approach to VaR computations.

VaR estimation under normality

Below, I review the concept of VaR under the assumption of normality and how this assumption simplifies the computation of VaR considerably. Suppose that the return for any asset, i, (i=1 to N) at a given point in time, t, is normally distributed, that is, $R_{it} \sim N(\mu_i, \sigma_i^2)$. Moreover, assume that asset returns are uncorrelated over time, that is, $cov(R_{i,t}, R_{i,t-j}) = 0, j = 1, 2...$, but could be contemporaneously correlated across assets, that is, $cov(R_{i,t}, R_{\hat{i},t}) = \sigma_{ij}\ \forall i \neq \hat{i}, t$, with the covariance matrix being denoted by Σ. For any portfolio of these assets, with portfolio weights given by $\omega = [\omega_1\ \omega_2\ ...\ \omega_N]$, with $\sum_i \omega_i = 1$, the returns can be written as the weighted average of the returns on the individual assets, that is, $R_p = \sum_i \omega_i R_i$. The returns on this portfolio are also normally distributed, with mean $\mu_p = \sum_i \omega_i \mu_i$ and variance $\sigma_p^2 = \omega \Sigma \omega^T$. This represents the first major advantage of assuming normality. If individual asset returns are normally distributed, then the returns on any portfolio of these assets has a normal distribution as well. At a critical probability of α, the VaR is the solution to

$$1)\quad \int_{-\infty}^{VaR_\alpha} f(R_p \mid \mu_p, \sigma_p)\, dR_p = \alpha,$$

where $f(.)$ is the normal density for portfolio returns. Typical values of α range from 1 percent to 10 percent. The second advantage of the normality assumption is that the computation of VaRs at different critical values (that is, solving equation 1) is relatively straightforward.

However, these simplifying assumptions have two drawbacks. First, many derivative securities have payoffs that are nonlinear functions of the underlying assets. The fact that the asset satisfies the normality assumption does not imply that the derivative has a normal distribution. This raises questions about whether this version of VaR analysis can be applied universally. This has been the focus of much concern (Beder, 1995) and several variants have been proposed to alleviate the problem (Fallon, 1996).[4] Second, there is considerable evidence in the academic literature to suggest that security returns are non-normal, typically exhibiting fat tails and volatility clustering (Kim and Kon, 1994, and the references cited therein).

Several alternatives to normality have been proposed in the literature. For example, in their comprehensive survey of alternative definitions for the stochastic process for stock returns, Kim and Kon classify the return processes as time-dependent and time-independent models of conditional heteroscedasticity (that is, of changes in the volatility of asset returns).[5] While the time-dependent models are more successful as models of asset returns, they are also considerably more complicated. Moreover, when firms are attempting to forecast the risk of losses over short holding periods (ranging from one day to two weeks), simpler models might be adequate. This seems to have been the justification behind the RiskMetrics™ framework developed by JP Morgan, as well as the variant proposed by Zangari (1996), which uses a simple version of the mixture of normals approach.[6] Clearly, the trade-off between having a procedure that accurately reflects the risk of the portfolio and one that is not too computationally intensive for the end user needs to be considered. Below, I discuss this mixture of distributions approach, relate it to the existing academic findings, and discuss problems with conventional estimation techniques. I then describe the alternative approach that overcomes these problems—the QB-MLE technique (Hamilton, 1991).

VaR estimation for a discrete mixture of normals

Empirical evidence suggests that the assumption that asset returns are normally distributed is inappropriate and that returns are actually fat tailed. One way to model such a distribution is to assume that returns are generated from a mixture of normal distributions. Specifically, suppose the stochastic process for the returns for security i is defined by

$$2)\quad R_{it} = \lambda_{it} R^{n}_{it} + (1 - \lambda_{it}) R^{ß}_{it},$$

where $R^{n}_{it} \sim N(\mu_n, \sigma^2_n)$, $R^{ß}_{it} \sim N(0, \sigma^2_{ß})$, and λ_{it} takes on a value of 1 with probability p, being equal to zero otherwise. The three random variables $\{R^{n}_{it}, R^{ß}_{it}, \lambda_{it}\}$ are assumed to be uncorrelated with each other and over time.[7]

Intuitively, the return on an asset at any given time can be drawn from one of two normal distributions, with the outcome, λ, determining which distribution is chosen. For example, most of the time (with probability p) the returns might be from the first distribution, that is, $\lambda = 1$. Occasionally (with probability $1 - p$), something unusual might happen (like the stock market crash of October 1987) that significantly increases volatility. This would be reflected in equation 2 by returns generated from the second (potentially higher variance) distribution, that is, $\lambda = 0$. The benefit of such a specification is that it allows for the possibility that occasionally the return is generated from a distribution with a higher variance, while simultaneously maintaining the structure of normal densities, conditional on the realization of λ (a *jump* from one distribution to another).[8]

The first issue of concern, then, is the estimation of the parameters $\{p, \mu_n, \sigma_n, \sigma_{ß}\}$ for individual assets (since the realization of λ is not typically observed by the researcher). I discuss three alternative methodologies for estimating $\{p,\mu_n,\sigma_n,\sigma_{ß}\}$. First, I consider traditional maximum likelihood. It turns out that there are problems associated with this approach in the context of mixtures. These problems motivate the next approach, which is the QB-MLE technique. I discuss alternative interpretations of the approach, and assess its effectiveness estimating parameters in simulated data. For completeness, I briefly compare the QB-MLE approach to the Bayesian (Gibbs-sampling based) approach proposed by Zangari (1996).

Traditional maximum likelihood approach

This approach would require the researcher to select the parameters that maximize the following log-likelihood function (dropping subscript *i* for convenience) for the mixture of normal densities

$$3)\quad \ell\Big((p, \mu_n, \sigma_n, \sigma_\beta)|\{R_t\}\Big) = \sum_t \log\Big[\frac{p}{\sigma_n}\exp\Big(-\frac{1}{2}\frac{(R_t-\mu_n)^2}{\sigma_n^2}\Big) + \frac{1-p}{\sigma_\beta}\exp\Big(-\frac{1}{2}\frac{R_t^2}{\sigma_\beta^2}\Big)\Big].$$

Unfortunately, as pointed out by Hamilton (1991), a global maximum does not exist for this function.[9] Consequently, attempting to use this approach to parameter estimation leads to instability, local solutions, and nonconvergence problems.

Quasi-Bayesian maximum likelihood estimation

Hamilton (1991) points out that the estimation problem would have been simplified considerably if the researcher had observations on the realization of λ available directly. Moreover, even if one had some observations, or some priors, this estimate could be improved. Second, while technical restrictions get around the problem of the failure of the existence of a global maximum, this still leaves the question how to deal with these problems in the small sample case. The method suggested by Hamilton is to maximize the following variant to the likelihood function:

$$4)\quad \ell\Big((p, \mu_n, \sigma_n, \sigma_\beta)|\{R_t\}\Big) - \frac{a_n}{2}\log(\sigma_n^2) - \frac{a_\beta}{2}\log(\sigma_\beta^2) - \frac{b_n}{\sigma_n^2} - \frac{b_\beta}{\sigma_\beta^2} - \frac{c_n(m_n-\mu_n)^2}{2\sigma_n^2},$$

where $\ell(.)$ is the likelihood function defined in equation 3 and $\{a_n, b_n, c_n, m_n, a_\beta, b_\beta,\}$ are (nonnegative) constants that reflect one's prior beliefs about the parameters that are being estimated.[10] Hamilton presents four alternative interpretations for the functional form that he has suggested and the manner in which the constants reflect the researcher's prior beliefs about the parameters.[11] Under three of these four approaches, the estimator can be interpreted as being based on Bayesian updating of the researcher's prior beliefs.

Zangari's methodology is based on the Bayesian updating of the densities for the relevant parameters using the observed return series. Since the computation of this posterior distribution is difficult in practice, Zangari suggests the use of the Gibbs sampler instead.[12] This procedure is time consuming; consequently, Zangari proposes that the mixture-related parameters be reestimated only once a month. Moreover, as pointed out earlier, the QB-MLE technique also has several Bayesian interpretations and the method is relatively straightforward to implement.

A direct comparison of the QB-MLE with the Bayesian estimation technique is beyond the scope of this article. Instead, the analysis focuses on how well the mixture of normals assumption combined with QB-MLE does relative to the traditional normality assumption.

Results based on simulated data

To examine both the effectiveness of the estimation technique and the ability of the mixture of normals to capture fat tails, I provide two sets of results. First, I examine how well the QB-MLE performs in estimating the parameters in simulated data generated from a mixture of normals data-generating process. Then, I compare the implications of assuming normality with those of the mixture of normals when the underlying density has a fat-tailed distribution.

Returns generated from a mixture of normals

To examine the robustness of the QB-MLE technique, I generate a variety of samples and examine the ability of the algorithm to estimate the parameters. Specifically, I consider mixtures drawn from two normals with zero means, variances $\sigma_n = 2$, $\sigma_\beta = 10$, and p ranging from 0.10 to 0.90. For each set of parameter inputs, I generate 100 samples of size 1,000 and estimate the parameters for each subsample. The results of this process, the mean and the standard deviation of the parameter estimates, are presented in table 1. The estimation routines are stable and do a good job in estimating the underlying parameters. The next step is to evaluate the effectiveness of this technique when the return-generating process exhibits fat tails (without necessarily being drawn from a mixture of normal distributions).

TABLE 1

Estimates from simulated data

Probability (p)	Estimates		
	$\hat{p}$	$\hat{\sigma}_n$	$\hat{\sigma}_\beta$
0.10	0.10741 (0.03919))	1.95615 (0.62693)	10.02450 (0.32751)
0.20	0.20857 (0.03249)	1.99397 (0.30232)	10.01270 (0.32215)
0.30	0.29628 (0.02919)	1.98501 (0.21278)	10.02834 (0.31111)
0.40	0.39521 (0.02837)	1.98489 (0.15897)	9.93684 (0.40044)
0.50	0.50197 (0.02800)	1.99356 (0.10294)	10.02182 (0.39484)
0.60	0.60376 (0.02981)	2.00225 (0.11112)	9.98221 (0.39651)
0.70	0.70108 (0.02322)	1.99523 (0.08838)	9.95835 (0.48623)
0.80	0.80018 (0.021877)	1.99578 (0.07755)	10.05155 (0.64645)
0.90	0.89690 (0.01380)	1.99648 (0.05238)	9.87647 (0.90500)

Notes: The maximum likelihood estimates are based on equation 4, with $a_i = b_i = 0.20$, $c_i = 0.10$, $m_i = 0$, for $i = n,\beta$ (as in Hamilton, 1991). Averages for 100 samples (of size 1,000) drawn from a mixture of normals distribution with σ_n=2, σ_β=10, and p varying from 0.10 to 0.90 across the runs. Standard deviations of the estimates are in parentheses.

Returns generated from a Student's t distribution

A distribution that exhibits the typical property of fat tails seen in asset returns is the Student's *t* distribution, which is characterized by its *degrees of freedom*. Fat-tailed behavior is more pronounced at lower degrees of freedom, with the distribution resembling a normal density at higher degrees of freedom. I generate simulated data from Student's *t* distributions with 2, 4, 10, and 100 degrees of freedom. For each of these, I generate a sample of size 10,000. The VaR for each simulation is computed in two ways. The theoretical VaR is computed based on the parameters used in the simulation. The sample VaR is based on the appropriate percentile from the sample itself.

Then, I estimate parameters under the assumption of normality as well as the assumption that the data have been generated from a mixture of normals. Based on these parameters, I compute the VaRs at different probability levels for the normal and mixture of normals approach and compare them to the theoretical and sample VaRs.

Table 2 illustrates that the mixture of normals has smaller errors than the normal approach at higher percentile levels. Moreover, when it has higher absolute errors than the normal approach, it errs toward conservative (high) VaR estimates. This is in contrast to the normal approach, which tends to generate low VaR estimates. Looking first at the columns labeled *error relative to theoretical*, we see that both the normal and the mixture of normals approach do a better job of measuring VaR at higher degrees of freedom. This is not a surprise, since the distribution begins to more closely resemble a normal density. Moreover, the normal approach understates (in absolute terms) the VaR relative to the true value at very high levels of confidence and overstates it at lower levels. In contrast, the mixture of normals approach reflects the opposite behavior, understating VaRs only under very low levels of confidence. While the percentage error under the mixture can be quite high (as much as 36.10 percent), it is generally biased toward being higher than the normals when a high level VaR is required. This represents a desirable characteristic of such a risk measure. Contrary to conventional wisdom, assuming normality when the distribution is fat tailed need not result in VaRs that are consistently understated. Similar patterns are also observed if one compares the computed VaRs to the sample VaR, which is defined as the critical return, in the simulated sample, such that μ percent of the returns lie below this threshold.

Estimation results for foreign exchange data

To assess the ability of the mixture of normals and the QB-MLE technique to estimate parameters and measure VaR more accurately than the normal distribution, I examine how well it does with a sample of daily foreign exchange returns for eight currencies—the Canadian dollar, French franc, German mark, Italian lira, Japanese yen, Swiss franc, British pound, and Dutch guilder. Returns are measured from January 1, 1978, to August 26, 1996.[13]

TABLE 2

Simulation results comparing VaR estimates for Student's *t* distributions

Percentile	*t* dist (theoretical)	*t* dist (actual)	Normal	Mixture of normals	Error relative to theoretical (normal)	Error relative to theoretical (mixture)	Error relative to sample (normal)	Error relative to sample (mixture)
				Student's *t* with 2 degrees of freedom				
0.5	–9.9248	–9.6126	–6.7562	–11.8241	–31.93%	19.14%	–29.72%	23.01%
1.0	–6.9646	–6.7726	–6.1018	–9.4787	–12.39%	36.10%	–9.90%	39.96%
2.5	–4.3027	–4.3153	–5.1408	–5.6513	19.48%	31.34%	19.13%	30.96%
5.0	–2.9200	–2.9249	–4.3143	–2.7758	47.74%	–4.94%	47.50%	–5.10%
				Student's *t* with 4 degrees of freedom				
0.5	–4.6041	–4.4062	–3.5883	–4.9642	–22.06%	7.82%	–18.56%	12.66%
1.0	–3.7469	–3.5614	–3.2407	–4.0344	–13.51%	7.67%	–9.00%	13.28%
2.5	–2.7764	–2.6265	–2.7303	–2.7487	–1.66%	–1.00%	3.95%	4.65%
5.0	–2.1318	–2.0445	–2.2914	–2.0543	7.49%	–3.64%	12.08%	0.48%
				Student's *t* with 10 degrees of freedom				
0.5	–3.1693	–3.1248	–2.8916	–3.2123	–8.76%	1.36%	–7.46%	2.80%
1.0	–2.7638	–2.7663	–2.6115	–2.7647	–5.51%	0.03%	–5.60%	–0.06%
2.5	–2.2281	–2.2517	–2.2002	–2.2112	–1.25%	–0.76%	–2.29%	–1.80%
5.0	–1.8125	–1.8064	–1.8465	–1.8038	1.88%	–0.48%	2.22%	–0.14%
				Student's *t* with 100 degrees of freedom				
0.5	–2.6259	–2.6388	–2.6105	–2.6368	–0.59%	0.42%	–1.07%	–0.08%
1.0	–2.3642	–2.3869	–2.3576	–2.3777	–0.28%	0.57%	–1.23%	–0.39%
2.5	–1.984	–2.0292	–1.9863	–1.9970	0.12%	0.66%	–2.11%	–1.59%
5.0	–1.6602	–1.6896	–1.6670	–1.6700	0.41%	0.59%	–1.34%	–1.16%

Notes: Errors are computed based on the (percent) difference between the VaR based on either the normal or the mixture of normals assumption and a benchmark VaR. This benchmark is computed using the known degrees of freedom for the *t* distribution (*theoretical VaR)* as well as the appropriate percentile in the sample (*sample VaR)*.

Summary statistics for the currency returns are provided in table 3. The hypothesis that these returns are drawn from a normal distribution is strongly rejected.[14]

First, I evaluate the difference between VaR measures based on the normal versus the mixture of normals for each currency. I compute VaRs for each currency on a daily basis

TABLE 3

Sample statistics
Daily foreign exchange returns

Mean	Canadian dollar	French franc	German mark	Italian lira	Japanese yen	Swiss franc	British pound	Dutch guilder
(X 10^5)	3.37	–7.63	.807	10.2	–16.0	–7.03	–11.0	–3.99
Median	0	0	0	0	0	0	0	0.000123
Maximum	0.01728	0.058678	0.058746	0.066893	0.035571	0.063879	0.103479	0.045885
Minimum	–0.01864	–0.04141	–0.03876	–0.03672	–0.05155	–0.03985	–0.09723	–0.03843
Std. deviation	0.002621	0.006929	0.006784	0.006542	0.006621	0.006884	0.008205	0.006643
Skewness	0.11387	0.035704	0.173338	0.531506	–0.39195	0.0313	0.097693	–0.08654
Kurtosis	6.62272	6.29632	7.666992	10.25995	6.501476	6.82522	15.23094	6.239683
Jarque-Bera statistic	2,471.592	2,039.186	4,108.262	10,098.9	2,415.103	2,745.512	28,068.85	1,974.41
Probability	0	0	0	0	0	0	0	0
Number of observations	4,502	4,502	4,502	4,502	4,502	4,502	4,502	4,502

Notes: The sample consists of daily returns from January 1, 1978, to August 26, 1996. A normal distribution should have a skewness (S) of 0 and kurtosis (K) of 3. The Jarque-Bera statistic is

$\frac{T}{6}\left[S_2 + \frac{1}{4}(K-3)^2\right]$, where T is the number of observations. The test statistic has a χ^2 distribution with 2 degrees of freedom.

and examine the frequency and the magnitude of the violations that occur. These are compared to what one would have expected if the VaRs had been correctly computed.

I use a 250-day estimation window and compute VaRs based on a one-day holding period, at the 97.5 percent confidence level, based on a recent survey of typical assumptions underlying VaR models used by firms.[15] The survey found that the confidence interval used by firms ranges from 95 percent to 99 percent, the one-day holding period VaR is typically computed, an observation period of one year (250 trading days) is used, and the historical data are equally weighted. The original sample consists of 4,502 daily (log) return observations. The initial estimation window and the need to compare VaRs with the next day's outcome reduce the VaR comparison to 4,251 observations. On any given day, t, I use the return series $\{R_{t-i}\}_{i=1}^{i=250}$ to compute the parameters and, therefore, VaR_t. This is compared with R_t and a violation is said to occur whenever $|R_t| > |VAR_t|$.[16] If the VaR is computed correctly, the expected number of violations is 0.05 times the number of observations, implying 212.5 violations. I examine the implications of both the assumption of normality and the mixture of normals approach. Figure 1 shows the time variation in the parameter estimates over the sample period for the German mark.[17] There is considerable time variation in the volatility measures (panels A, C, and D) under both approaches. Interestingly, there is also considerable variation over time in the estimate of p, the probability that returns are drawn from one distribution in the mixture.

The results from comparing VaR estimates for the eight currencies with the actual number of violations are summarized in table 4. In a sample size of 4,251, one would expect 212.5 violations. The number of violations that occur under the assumption of normality is significantly higher than one would expect and a likelihood ratio test rejects the hypothesis that the true underlying probability of a violation is 5 percent. In contrast, the number of violations of the VaR estimated under the mixture of normals is much lower than under the normal. In addition, one cannot reject the hypothesis that the model has a probability of violations equal to 5 percent.

TABLE 4

Violations of VaR under alternative methodologies

Currency	Number of violations: Normal	Number of violations: Mixture	Average size of violations (percent): Normal	Average size of violations (percent): Mixture	Expected size of violations (percent): Normal	Expected size of violations (percent): Mixture
Canadian dollar	256 (8.801)**	233 (2.011)	0.7113	0.7421	0.0428	0.0407
French franc	245 (4.981)*	227 (1.012)	1.7722	1.7975	0.1021	0.0960
German mark	263 (11.757)**	224 (0.639)	1.7481	1.7463	0.1082	0.0920
Italian lira	251 (6.938)**	223 (0.533)	1.6692	1.7694	0.0986	0.0928
Japanese yen	251 (6.938)**	215 (0.030)	1.4333	1.4052	0.0846	0.0711
Swiss franc	248 (5.921)*	222 (0.436)	1.8809	1.8906	0.1097	0.0987
British pound	279 (19.995)**	226 (0.879)	1.4411	1.4932	0.0946	0.0794
Dutch guilder	261 (10.873)**	237 (2.859)	1.7545	1.7319	0.1077	0.0966

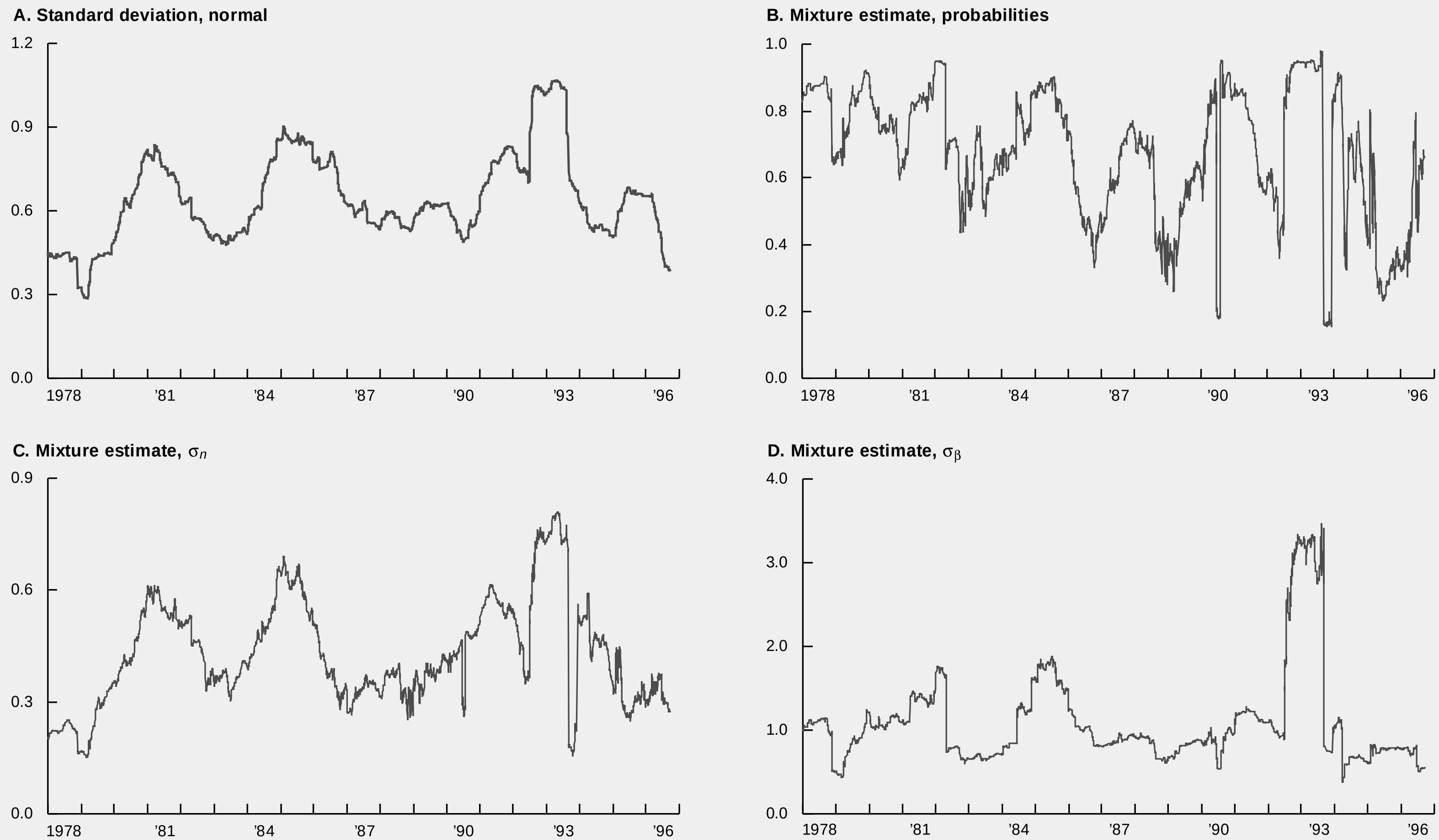
*Significant at the 5 percent level.

**Significant at the 1 percent level.

Notes: The log-likelihood test statistic (reported in parentheses) is $LR = 2[1n[(\alpha^*)^x(1-\alpha^*)^{T-x}] - 1n[\alpha^x(1-\alpha)^{T-x}]]$, T = sample size (4,251), x = number of violations, $\alpha = 0.05$, and $\alpha^* = x/T$ is the sample fraction of violations. The test statistic has an asymptotic χ^2 distribution with 1 degree of freedom. The critical values are 6.6349 and 3.841 at the 99 percent and 95 percent confidence levels, which translates into violations of 250 and 241, respectively.

FIGURE 1

Parameter estimates of the German mark



Notes: The estimates are computed using rolling 250-day windows [t–250,t–1]. The maximum likelihood estimates for panels B–D are based on equation 4, with $a_i = b_i = 0.20$, $c_i = 0.10$, $m_i = 0$, for $i = n, \beta$ (as in Hamilton, 1991).

Table 4 also shows the average size of a violation and the expected size of a violation (defined as the average size times the frequency of a violation). The average size of the violation is larger under the mixture than under the pure normal assumption. While this might seem surprising, recall (from the results of table 2) that the rank ordering of the VaRs under the normal versus mixture of normals depends critically on the level of confidence as well as the shape of the distribution, and this could explain the results in table 4.

The expected size of the violations is uniformly smaller under the mixture (the last two columns of table 4). This suggests that for individual assets, the mixture of normals provides superior VaR estimates than the conventional normality assumption because both the number and expected size of violations are lower under the mixture of normals approach. Next, I examine how well this process works in the context of portfolios.

Portfolio results

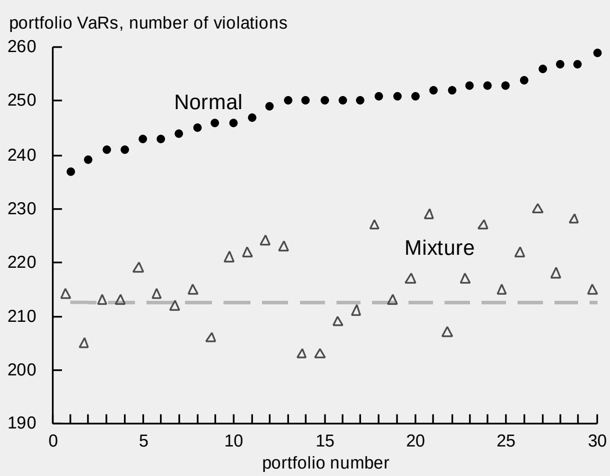
As mentioned earlier, the two benefits of the normality assumption are that it is relatively simple to calculate the VaRs associated with different confidence levels and to aggregate individual parameters to develop the parameters of a portfolio. The mixture of normals shares the first property. But how would one aggregate these parameters in the context of a portfolio? I assume that the covariance across assets is independent of λ_i. This implies that there are only two covariance matrices that could be generating the returns. The off-diagonal terms of these matrices are independent of λ realizations, while the diagonals are either $\sigma^2_{i,n}$ or $\sigma^2_{i,ß}$ depending on the realization of λ_i. The second issue is whether these realizations are independent across assets. The assumption of independence would be consistent with the large literature on jump diffusion models, which typically assumes that the jump risk is fully diversifiable (Merton, 1976). However, it is not immediately clear that this assumption is reasonable in the context of the risk-management activities of a bank, since the prospect of bankruptcy could make the bank worry about risk that might seem diversifiable in an asset pricing context. Moreover, this assumption complicates the mapping between confidence levels and VaRs considerably. For example, with eight assets, one would have to consider $2^8 = 256$ possibilities for the realizations of λ_i, with the process quickly becoming intractable.

The assumption of perfect correlation is not valid either, since one would then expect identical values of p for all eight currencies. Here, I adopt a computationally simpler alternative and test to see whether the approximation works. For a portfolio ω, I use as inputs $p_p = \sum_i \omega_i p_i$ and the two covariance matrices, Σ_n and $\Sigma_ß$, which are identical along the off-diagonals and contain the relevant variances on the diagonals. These assumptions approximate the distribution of portfolio returns by a mixture of normals distribution. To assess how good an approximation this represents, I form 30 random portfolios of the eight currencies and evaluate how well the portfolio VaR estimates do relative to the profits and losses on the portfolio.

In figure 2, I plot the number of violations (the simulation is for 4,251 daily returns for 30 different portfolios) under the mixture of normals approach relative to the conventional normality assumption. The portfolios have been sorted based on their VaR estimates under the assumption of normality. Again, the fraction of violations under the mix is much lower than under

FIGURE 2

Violations for foreign currencies portfolios



Source: Author's calculations from data provided by the Federal Reserve Bank of New York.

TABLE 5

VaR violations for portfolios of currencies

	Normal	Mixture of normals
Number of violations	249	217.4
	(5.5460)	(7.7442)
Proportions of violations	0.0586	0.0511
	(0.013)	(0.0018)
Magnitude of violations	1.0970	1.1293
	(0.1766)	(0.1845)
Average violations	0.0643	0.0580
	(0.0109)	(0.0109)

Notes: The statistics are based on 30 random portfolios of currencies constructed over the entire sample period. Standard deviations are in parentheses.

the normality assumption. Moreover, the critical number of violations to reject the hypothesis that the VaR is consistent with a 2.5 percent confidence level is 250 (at the 1 percent level) or 241 (at the 5 percent level). For the normal density, the hypothesis can be rejected in 26 of the 30 portfolios at the 5 percent level and in 13 of the portfolios at the 1 percent level. One cannot reject the mixture-based VaRs in any of the portfolios.

Table 5 indicates that, similar to the findings for individual currencies, the magnitude of the violations under the mixture tends to be larger than the magnitude under pure normality. However, the expected size of the violations is smaller under the mixture of normals approach.[18] Thus, in the context of portfolios as well as individual assets, the mixture of normals provides superior VaR estimates relative to the conventional normality assumptions, because the number of violations is lower and consistent with expectations and the *ex ante* expected size of violations is smaller.

Conclusion

The analysis in this article highlights the critical nature of the existing assumptions underlying VaR computations and the complications that result when the methodology is used for assets that exhibit fat-tailed return distributions. The mixture of normals approach combined with QB-MLE is shown to perform significantly better, in the context of both individual assets and portfolios. Further research is needed on the number of components to include in the mixture, more complicated intertemporal dependencies, and the development of computationally feasible aggregation algorithms.

NOTES

[1]In fact, the concept of VaR was motivated by this ability to capture risk by one number. Dennis Weatherstone, the chief executive officer of JP Morgan at the time (and also the chairman of the influential Group of Thirty study on derivatives), insisted that such a single measure of the firm's exposure be made available to him every morning, resulting in the development of the underlying quantitative techniques (Financial Engineering, Ltd., *Risk*, special supplement, 1996).

[2]The inputs into the model include 1) assumptions (and estimation techniques) for the stochastic processes that determine the returns on individual assets; 2) a methodology for mapping the return distributions for individual assets into the aggregate return distribution for the portfolio (and hence to profits and losses [P&L]); and 3) a computationally simple process for evaluating VaR at different probability levels for this aggregate P&L distribution. All these steps obviously also depend on the relevant holding period over which the analysis is conducted. In this article, my primary focus is not on the second aspect (refer to JP Morgan's RiskMetrics™ document on position mappings for greater detail). I focus instead on the estimation of the underlying stochastic processes and the difficult trade-off between the need for an approach that is accurate and the need for one that is easy to implement. I discuss some aggregation issues later in this article.

[3]I also briefly review the problems associated with standard estimation techniques (such as maximum likelihood), and the Bayesian approach using Gibbs sampling proposed by JP Morgan.

[4]Most techniques try to approximate these nonlinearities based on Taylor series expansion, leading to methods based on the delta and gamma of security-type approaches, for example.

[5]Under the former, they consider ARMA, GARCH, and EGARCH models (Bollerslev, Engle, and Nelson [1994]) and the Glosten, Jagannathan, and Runkle (1993) specifications for asset returns. Under the latter, they consider Student's *t* models, generalized mixtures of normals, Poisson jump models, and stationary normal models.

[6]See Kon (1984) for a comparison of a general version of the mixture of normals with the Student's *t* density, for example.

[7]The mixture of normals also allows for more complex structures where λ_t could follow a Markov process. See Engel and Hamilton (1990) for an application.

[8]In fact, the traditional jump diffusion model can be interpreted as allowing for the possibility of a jump between an infinite number of normal distributions (Kon [1984]).

[9]Specifically, a singularity arises whenever one of the observations is attributed entirely to a single distribution, since the mean is then imputed to be the value of the observation and the variance approaches zero. However, focusing attention on just the largest local maximum with positive variances leads to consistent estimates. The problem at that stage is one of ensuring that the numerical algorithms are bounded away from zero, which is difficult to do in practice. Moreover, as pointed out by Lehmann (1983) and Robert (1994), in any finite sample, the probability that none of the observations was generated from one of the mixture components is strictly positive. They argue that this also contributes to the instability of the maximum likelihood estimation process. Consequently, estimation under this approach is not recommended.

[10]Notice that in the special case where these constants are all set to zero, the functional form suggested by Hamilton (1991) collapses to the traditional maximum likelihood function. The constants are identical in our estimation to those used by Hamilton. Moreover, perturbing the parameters had no effect on the estimates. Refer to table 1 for additional details.

[11]Specifically, he suggests that one could interpret the estimator as 1) representing prior information which is equivalent to observed data, 2) the mode of a Bayesian posterior distribution, 3) an analogy with a Bayes estimator, or 4) a penalized maximum likelihood function.

[12] For full details of this methodology, see Zangari (1996).

[13]The currencies selected here are identical to those in Hendricks (1996). The start of the time period is also identical, while the ending point reflects when this research project was started.

[14]All the currencies reflect significant skewness and kurtosis relative to what one would expect if the samples had been drawn from normal distributions. In addition, the Jarque-Bera statistic rejects normality for all currencies.

[15]"Amendment to the Capital Market Accord to incorporate market risks: The use of internal models for supervisory purposes," a study conducted by a joint ISDA/LIBA task force which surveys its members to assess the assumptions underlying their use of VaR.

[16]I focus attention on both excessive losses and gains throughout this analysis. Disaggregating these two does not change the nature of the results.

[17]The patterns for the other currencies are substantially similar and are therefore not included.

[18]These simulation runs are computationally intensive, but increasing the number of portfolios to 50 did not change the nature of the results.

REFERENCES

Beder, Tanya Styblo, "VAR: Seductive but dangerous," *Financial Analysts Journal*, September-October 1995, pp. 12–24.

Bollerslev, T., R. Engle, and D. Nelson, "ARCH models," in *Handbook of Econometrics*, Robert F. Engle and D. McFadden (eds.), Vol. 4, Amsterdam: North Holland, 1994.

Casella, G., and E. George, "Explaining the Gibbs Sampler," *The American Statistician*, Vol. 46, No. 3, August 1992, pp. 167–174.

Engel, C., and J. Hamilton, "Long swings in the dollar: Are they in the data and do the markets know it?," *American Economic Review*, Vol. 80, 1990, pp. 689–713.

Fallon, William, "Calculating value-at-risk," Wharton Financial Institutions, working paper, No. 96-49, 1996.

Financial Engineering, Ltd., Risk Publications, "Value at Risk: Special supplement," *Risk*, June 1996.

Glosten, L., R. Jagannathan, and D. Runkle, "On the relationship between the expected value and the volatility of the normal excess return on stocks," *Journal of Finance*, Vol. 48, No. 5, December 1993, pp. 1779–1801.

Hamilton, James, "A quasi-Bayesian approach to estimating parameters for mixtures of normal distributions," *Journal of Business & Economic Statistics*, Vol. 9, No. 1, 1991, pp. 27–39.

Hendricks, Darryll, "Evaluation of value-at-risk models using historical data," *Economic Policy Review*, Federal Reserve Bank of New York, Vol. 2, April 1996, pp. 39–69.

Hopper, Gregory, "Value at risk: A new methodology for measuring portfolio risk," *Business Review*, Federal Reserve Bank of Philadelphia, July/August 1996, pp. 19–29.

Hsieh, David, "Nonlinear dynamics in financial markets: Evidence and implications," *Financial Analysts Journal,* Vol. 51, July-August 1995, pp. 55–62.

Jorion, P., "Predicting volatility in the foreign exchange markets," *Journal of Finance*, Vol. 50, No. 2, 1995, pp. 507–528.

Kim, Dongcheol, and Stanley Kon, "Alternative models for the conditional heteroscedasticity of stock returns," *Journal of Business*, Vol. 67, No. 4, 1994, pp. 563–598.

Kon, Stanley, "Models of stock returns—A comparison," *Journal of Finance*, Vol. 39, No. 1, March 1984, pp. 147–165.

Lehmann, E.L., *Theory of Point Estimation*, New York: John Wiley and Sons, 1983.

Leong, Kenneth, "The right approach," *Risk*, Special Supplement, June 1996.

Merton, R., "Option pricing when under underlying stock returns are discontinuous," *Journal of Financial Economics*, Vol. 3, No. 1/2, 1976, pp. 125–144.

Morgan Guaranty Trust Company, Global Research, RiskMetrics™, technical document, New York, 1995.

Robert, Christian, "Mixture of distributions: inference and estimation," Université de Rouen and Crest, Départment de Math, Insee, Paris, working paper, 1994.

Titterington, D.M., A. Smith, and U. Makov, *Statistical Analysis of Finite Mixture Distributions*, Wiley Series in Probability and Mathematical Statistics, New York: John Wiley and Sons 1985.

Turner, Chris, "VAR as an industrial tool," *Risk*, Vol. 9, No. 3, March 1996.

Zangari, Peter, "An improved methodology for measuring VAR," *RiskMetrics Monitor*, Reuters/JP Morgan, 1996.