Version of October 26, 2004

"Estimation Error in the Assessment of Financial Risk Exposure"

by

Stephen Figlewski*

* Professor of Finance
NYU Stern School of Business
44 West 4th Street, Suite 9-160
New York, NY 10012-1126

212-998-0712
sfiglews@stern.nyu.edu

I wish to thank James Park and Antonios Sangvinatsos for research assistance, and Luca Benzoni, Rob Engle, Eric Jacquier, Lasse Pedersen, Jim Ramsey and participants in seminars and presentations at NYU, Columbia, Hong Kong University of Science and Technology, City University of Hong Kong, University of Toronto, Vanderbilt, CIFEr 2003, the European Finance Association, Citigroup, and RiskMetrics for many valuable comments and suggestions.

"Estimation Error in the Assessment of Financial Risk Exposure"

Abstract

Measures of financial risk exposure like VaR require predicting the tail of a returns distribution. Due to estimation error, however, a "tail event" may simply be a not-so-rare occurrence at a time when predicted volatility underestimates the true volatility. This problem gets worse the further in the tail one is trying to predict.

In a simulation of 10,000 years of daily returns, we first examine the best possible case, when volatility is an unknown, but constant, parameter. In the more realistic case, volatility changes stochastically over time, which makes estimation error much worse, although strong mean reversion in the variance tends to dampen the effect. Non-normal fat-tailed return shocks make overall risk assessment much worse, especially in the extreme tails, but we find that the effect of tail fatness is much bigger than that of estimation error *per se*. Using an exponentially weighted moving average to downweight older data hurts accuracy if volatility is constant or only slowly changing. But with more volatile variance, downweighting improves performance for the most extreme tails.

We first use non-overlapping independent samples, but in practice, risk exposure is reestimated on a rolling basis, which causes strong autocorrelation in the estimation errors, and can increase the probabilities of multi-day events, like three 1% tail events in a row, by several orders of magnitude. To examine the impact of jumps, we simulate several jump-diffusion models of stock returns from the recent finance literature and find, again, that standard VaR seriously underestimates tail probabilities. Finally, we look at VaR estimates from 40 years of daily S&P 500 returns, which confirm that the issues examined in our simulations are also present in the real world.

1. Introduction

Use of formal statistical procedures for assessing exposure to financial risk, such as Value at Risk (VaR), has now become standard practice in the financial industry.[1] Regulatory authorities are using them for setting capital requirements for banks and much research has been done to examine the performance of different methods.[2,3] VaR and related risk measures involve predicting the size and frequency of large losses that fall in the tails of returns distributions.

Value at Risk addresses the following question: For a specified holding period (typically 1 day) and probability level α (typically 5% or 1%), what is the return such that the probability of experiencing a worse return over the holding period is no more than α? Even though it suffers from a variety of shortcomings, use of VaR as a practical tool for risk measurement has grown rapidly, largely because it is intuitive and relatively easy to calculate.[4] Our investigation of estimation error is not specifically tied to Value at Risk, but VaR will be useful in illustrating the general problem of assessing the risk exposure associated with the occurrence of low probability events.

Plain vanilla VaR uses standard statistical techniques to estimate the relevant parameters of the underlying returns distribution. In a typical calculation, returns are assumed to be normally distributed (asset prices are lognormal). Recent historical data are used to estimate the volatility, and the mean is either estimated or, more commonly, set to 0.[5] The estimated parameters are then plugged into the equation $\text{VaR} = \alpha_c\hat{\sigma} + \hat{\mu}$,

---

[1] Among many references, see: Hull [2002] ch. 16, for a textbook discussion; Jorion [1997]; or Risk Publications [1996]. Schachter's website gives an exceptionally comprehensive bibliography on the subject.

[2] See Basle Committee on Bank Supervision [1996], for example.

[3] See, for example, Duffie and Pan [1997] or Kupiec [1995]. The Journal of Banking and Finance devoted its entire July 2002 issue to VaR-related research.

[4] VaR specifies where the α-tail of the returns distribution begins, but it says nothing about the distribution of outcomes that fall in the tail. In the terminology of Artzner, et al [1999], VaR is not a "coherent" risk measure. A better (and coherent) measure of tail risk is the expected value of the return conditional on being in the α-tail. This is called by various names, including "Conditional Value at Risk" (C-VaR), or "expected shortfall."

[5] Constraining the mean to 0 amounts to a kind of Bayesian procedure, with a strong prior that the true mean daily return for financial assets is much closer to 0 than to a value randomly drawn from the sampling distribution for the mean. The classical estimator for the mean return is the sample average, but the sampling error on this estimate is surprisingly large. For a lognormal diffusion with mean μ and volatility σ, the sample average from a sample spanning T years has expected value μ and standard deviation $\sigma T^{-1/2}$, regardless of observation frequency. With a sample size like those typically used in VaR-type calculations, it is common for the sample average to make no sense economically: It may be negative, which is inconsistent with market equilibrium for normal assets, or outlandishly large. For example, with annual volatility of 20% and a 3-month returns sample, the standard deviation of the sample average around the true mean μ is $0.20 / (1/4)^{1/2} = 40\%$. A 2 standard deviation confidence interval would therefore cover $\mu \pm 80\%$.

where α_c denotes the 5% or 1% critical value of the normal distribution (-1.645 or -2.326) and $\hat{\sigma}$ and $\hat{\mu}$ are the sample volatility and mean ($\hat{\mu} = 0$, if the mean is not calculated).[6]

But the sample parameter values are only statistical estimates of the true parameters and the standard procedure takes no account of the estimation error in these figures. The 5% cutoff for the tail of a normal distribution is at -1.645 standard deviations, but when the sample volatility is an underestimate of the true volatility, there is more than 5% probability that the next observation will fall in the predicted 5% tail. As we will see below, the effect of sampling error increases the further into the tail one is trying to forecast.

Actually, if returns are normally distributed with constant mean and variance, the distribution of the next period return using estimated parameters is not normal at all, but rather, a Student-t distribution, which has fatter tails than the normal.[7] With sampling error, even the fact that the volatility parameter is obtained by taking the square root of the sample variance introduces a bias. The sample variance is an unbiased estimate of the true variance, but because the square root is a concave function, the sample standard deviation is biased low as an estimate of the true volatility, due to Jensen's Inequality:

$$E[\,\hat{\sigma}\,] \;=\; E[\sqrt{\hat{\sigma}^2}\,] \;<\; \sqrt{E[\,\hat{\sigma}^2\,]} \;=\; \sqrt{\sigma^2} \;=\; \sigma$$

Estimation risk is an important factor in using statistical measures to evaluate risk exposure, that seems to be largely ignored in practical risk management.[8]

We first examine this basic sampling problem and find that estimation risk is more serious than might have been recognized previously. But the "Baseline case" we have just described, with constant parameters and normally distributed return shocks, actually represents the best situation one could reasonably hope for. With constant parameters, the classical estimators for mean and volatility are consistent, so the solution to estimation risk is simply to use more data in the calculations. Estimation error would be negligible in a sample of, say, 5 years of returns with constant mean and volatility. Research has shown that in using the classical constant volatility estimator for financial returns, more accurate volatility forecasts can often be obtained with considerably longer historical returns samples than those that are normally used in practice.[9] But risk managers and traders prefer short samples because they do not expect volatility to remain constant over time and they feel their estimates should adapt quickly to changing market conditions.

---

6 Kupiec [1995] discusses the pros and cons of measuring VaR relative to zero, or relative to the expected value.

7 A proof is provided in the Appendix.

8 The impact of estimation risk on optimal investment and portfolio choice was explored many years ago by Bawa, Brown and Klein [1979], but they did not address risk management per se. Jorion [1996] raises the issue in a VaR context, but his focus is primarily on offering an alternative estimation technique that can improve accuracy in the case in which the true form of the returns distribution is known (e.g., normal) and the unknown parameters are assumed to be constant.

9 See Figlewski [1997] or Green and Figlewski [1999].

It is now well-accepted that volatility is time-varying.[10] Figure 1 plots the sample volatility for the S&P 500 stock index in a moving 63-trading day window from 1992 - 2002. During this time, the sample volatility ranged between around 6.0 percent to over 33.0 percent. If the true volatility changes over time, extending a sample backward to bring in more, but older, data may simply contaminate the forecast of the immediate future. Common practice is to use short sample periods (generally well under a year), often with a weighting technique, such as an exponentially weighted moving average (EWMA), to downweight older observations relative to recent ones.[11]

The major problem I focus on in this paper is how estimation risk affects probability predictions for rare events when the parameters of the underlying returns process are nonstationary. If we know the form of the returns distribution *a priori*, normal perhaps, then information accumulates quickly, since every observation drawn from that distribution will yield information about its tails. On the other hand, if we do not know the data generating process and simply tabulate an empirical distribution, we have to wait for many observations before we can learn much about its tails, because the world does not generate information about rare events very rapidly. Now if the data generating process itself is changing over time, estimating the tails accurately, either parametrically or nonparametrically, becomes much harder. It also becomes much less tenable to assume that we know the form of the distribution *a priori*.

I use simulation to explore the problem of evaluating risk exposure when the underlying returns distribution varies stochastically over time. The next section sets out a framework for the problem and introduces the measures of prediction accuracy that we will use throughout the paper. Simulation results show the overall performance of the standard approach to VaR estimation in a repeated sample with nonstochastic volatility. Section 3 extends the simulation to the case of volatility that evolves stochastically over time, according to a mean-reverting square root diffusion process.

In Section 4, we consider whether using an exponentially weighted moving average to downweight observations as they age can reduce the effect of estimation error due to time-varying volatility. We also examine the impact of non-Gaussian return shocks, drawn from a fat-tailed Student-t distribution. Recent models of security returns have introduced important deviations from the classical lognormal diffusion model in the form of Poisson jumps in both returns and volatilities. Jumps greatly exacerbate estimation error for plain vanilla VaR, but we defer explicit consideration of this more complicated departure from lognormality until Section 6.

In the first four sections we explore the sampling properties of the standard estimator using independent samples. But in real world risk management, risk models are refitted

[10] Schwert [1989] provides a broad look at volatility behavior in the historical record of returns and, indeed, much of the research on security price behavior over the last 20 years has focused on modeling time-varying volatility.
[11] For example, RiskMetrics employs an exponentially weighted moving average in its volatilities, which are widely used in the finance industry. See RiskMetrics [1996].

and updated daily as new data arrives. Section 5 looks at the induced autocorrelation in the errors when risk exposure is estimated from a rolling sample of returns. This exacerbates the problem of estimation risk, and can greatly increase the probabilities of multi-day events, such as three tail occurrences in a row. In this section, we also compare our simulation results to estimation risk in the real world, using a rolling volatility estimate on 40 years of Standard and Poor's 500 Index returns.

The stochastic volatility model used in our simulations models returns as a lognormal diffusion and volatility as an independent square root diffusion, but recent theoretical and empirical research provides strong evidence that asset returns follow more complicated processes, with doubly stochastic jumps and complex patterns of correlation among the stochastic factors. In Section 6, we consider the estimation problem when returns are generated by this type of process, using model specifications and parameter values drawn from the finance literature.

Section 7 concludes.

2. Estimating Risk Exposure when Volatility is Constant

To introduce the type of analysis we will employ in the paper, consider first the estimation error in the standard VaR approach to predicting the tails of a returns distribution when the underlying process is well-behaved. Throughout the paper, we will concentrate on predicting the lower tail of the distribution, which corresponds to losses on a long position in the underlying risk. The same kind of analysis clearly applies to the upper tail, which determines losses on a short position.

Assume the asset value follows a standard lognormal diffusion:

$$\frac{dS}{S} = \mu\, dt + \sigma\, dz \tag{1}$$

S is the value of the security (or asset, liability, portfolio, etc.) that we are interested in, constants μ and σ are the instantaneous mean and volatility, at annualized rates, and dz represents standard Brownian motion.

For the simulations, (1) is discretized as

$$r_{t+1} = \ln(S_{t+1}/S_t) = \mu\,\Delta t + \sigma\,\tilde{z}_t\sqrt{\Delta t}\,; \qquad \tilde{z} \sim N(0,1) \tag{2}$$

The time interval Δt represents one trading day, and is set to the value $\Delta t = 1/250$. We will refer to a period of 21 days as a "month," 63 days as "3 months" and 250 days as a "year."

The predicted volatility of the return that will be realized on day t+1 will be estimated from the returns over the previous K days, using equation (3).

(3) $$\sigma_{predicted,t+1} = \sqrt{\frac{250}{K}\sum_{\tau=t-K+1}^{t} r_\tau^2}$$

Consider calculating the 1-day VaR from a single 3-month (K = 63) sample of returns. As discussed in footnote 5, the sampling error of $\bar{r}$ as an estimate of μ is excessively large given what we know about real world asset pricing, so we follow standard practice and suppress calculation of the sample mean.

5% and 1% cutoff values are by far the most frequent choices in practical VaR applications. But these actually represent relatively common occurrences. An event with a 5% probability will occur on average one time out of 20. For daily returns, therefore, 5%-tail events should average more than one per month. A 1% event should occur every few months, about 2-3 times a year. While it is obviously important to be prepared for these events, a cautious risk manager needs to be concerned about losses that are less frequent but much more serious.

The first column in Table 1 shows the α-tail probability values that we will focus on, from the 5% level, $\alpha = 0.05$, out to $\alpha = 0.0001$, a one in 10,000 event. Another way to understand these low probabilities is in terms of how frequent an event with a given probability is. The second column in Table 1 gives the frequency of occurrence for the tail values. For example, an event with a probability of 0.001 can be expected to occur once in 4 years on average (where, as mentioned above, we take a "year" to be 250 days). Even a 1 in 10,000 event happens on average every 40 years: not a common occurrence by any means, but such a "once in a lifetime" event will not be considered out of the range of concern for a prudent risk manager with a long term view.

We expect one one-in-a-thousand $\alpha = 0.001$ event every four years on average. But in a given four year period, the one beginning tomorrow, for example, there might be no events at all, or there could be several events. A different way to express event frequency that takes into account the fact that the occurrence of an event on any given day is independent of the outcomes on other days, arises in the analysis of failure times. From the 1-day probability α, we can calculate the likelihood that it will be more than K days before the first event occurs. The probability of no event in the next K days is $(1 - \alpha)^K$, so the probability that there will be at least one event in K days is

(4) $$q = 1 - (1 - \alpha)^K$$

For a specified probability q, equation (4) can be solved for K_q, the shortest time period such that the probability an event will be observed during that interval is greater than q.

(5) $$K_q = \log(1 - q) / \log(1 - \alpha)$$

K_q is of obvious relevance to a risk manager when the event in question is of a very large "life-threatening" nature. The value q is the probability of "death" within the next K_q days.

Column 3 of Table 1 presents values for $K_{0.50}$, the shortest period going forward such that there is more than 50% probability that an event will be observed. For example, a 1 in 1000 event will occur on average every 4 years, but there is more than 50% chance of one happening within the next 2.8 years. For small values of α and q, K_q is shorter than the average event frequency, and lower q reduces K_q. K_q values for q=5% are shown in Column 4. For example, there is a q = 5% chance that a once in a lifetime $\alpha = 0.0001$ event will happen within just over two years. While standard VaR is a measure of risk exposure over a single period, K_q provides a measure of the risk exposure in a risk management <u>policy</u> that is meant to be used every period over an indefinite horizon.[12] As we report the results of different simulation exercises below, it may be useful to refer back to Table 1 to get an idea of the practical impact of estimation error on these measures of risk exposure.

Now let us consider how an incorrect volatility estimate affects perceived tail probabilities. The location of the α-tail is given by $N^{-1}[\alpha]\,\sigma_{predicted}$, where $N^{-1}[.]$ denotes the inverse normal distribution. Suppose that the true volatility is 20.0% but the estimated volatility calculated from a single 63-day sample is 17.68%. (This substantial underestimate of the true parameter is the 10th percentile of the sampling distribution from our simulation of about 40,000 such 3-month periods.) Because the sample volatility is too low, the predicted VaR values also underestimate the true values. For example, the user believes that the 5% tail is at $-1.645\,\sigma_{predicted}$ $= -1.645 \times 17.68 \times \sqrt{250} = -1.84\%$, but the true 5% tail is actually at $-1.645\,\sigma_{true} = -1.645 \times 20.0 \times \sqrt{250} = -2.08\%$. In terms of the predicted volatility, this is $-1.645 \times \dfrac{\sigma_{true}}{\sigma_{predicted}} = -1.86$ standard deviations.

Columns 2 and 3 in Table 2 show the predicted and actual tail cutoff values with these volatilities for the α values in Column 1, expressed in terms of the predicted standard deviation. The values in Column 3 are given by $N^{-1}[\,\alpha\,]\,\dfrac{\sigma_{true}}{\sigma_{predicted}}$.

A more intuitive way to think about the estimation error is in terms of the underprediction of the probability that next period's return will be worse than the target VaR probability α. The true probability of what someone using the incorrect volatility will perceive to be a tail event, is given by

[12] Note that this calculation is based on the assumption of independent draws from the true returns distribution. Estimates of K_q from a predicted distribution will be subject to estimation error in the same way as the Value at Risk is.

$$\alpha_{true} = N^{-1}\left[\; N[\,\alpha\,]\frac{\sigma_{predicted}}{\sigma_{true}}\;\right]$$

These values are shown in the fourth column of Table 2 and the ratios of actual to predicted tail probabilities, α_{true} / α, are given in Column 5. We will focus on these probability ratios for different α values in reporting the results of our simulations below.

Table 2 shows that in extending the VaR calculations into the more extreme tails of the returns distribution, the farther one looks into the tail, the greater is the effect of the incorrect volatility estimate. For example, while the probability of a 5% tail event is 1.46 times too large, a 0.0001 event is more than 5 times more likely than expected.

These tail estimates and probabilities have all been computed from a single volatility that was substantially lower than the true value. They do not represent the average forecasting performance for a standard VaR calculation. In order to examine that much more relevant issue, we simulate 2,500,000 consecutive returns from equation (2) with $\mu = 0$ and $\sigma = 0.20$, and repeat the process of estimating volatility every 63 days using equation (3). The 2 1/2 million simulated days yield 39,682 non-overlapping 63-day periods. To maximize comparability across simulations based on different assumptions and a variety of estimation strategies, we use the same seed value in the random number generator for every simulation run, until Section 6.

The estimated volatility is used to compute 1-day VaR values for the first day following each sample period. The realized return for that day (simulated using the true volatility) is then converted into a number of standard deviations by dividing it by the estimated volatility. That is, the return shocks are standardized by expressing them in terms of the predicted standard deviations. When the full set of returns data has been processed, for each value of α, we determine the location of the cutoff for the true α-tail (i.e., the number of predicted standard deviations that would have captured exactly α percent of the returns) and the actual percentage of returns that fell within the predicted α-tails (the true probabilities).

Table 3 presents the results. Overall, the root mean squared error (RMSE) in the volatility forecast is 0.0178. This is approximately 9 percent of the true volatility 0.20. How large this volatility forecast error is judged to be depends on what use is to be made of the estimate: For pricing options it may be considered quite large, but for assessing overall risk exposure, it seems at first glance to be relatively small. Table 3 shows that for the 5% tail, estimation error does not seriously affect the standard VaR calculation. For example, the 5% VaR level is estimated to be -1.645 times the sample volatility, while in the simulation, the true 5% VaR was at -1.649 times the sample volatility. The true probability of a return in the predicted 5% tail was 5.05%, only 1.01 times the predicted probability.

Estimation error is a little greater for the 1% VaR calculation. The actual 0.01-tail was at -2.384 standard deviations, and the next period return fell in the predicted 1% tail 1.17% of the time. However, as we saw above, the effect of estimation error is greater for more extreme values of α. For $\alpha = .0005$ (a once-in-8 years event) the true probability is more than twice as large as predicted.

The simulation results reported in Table 3, using a 63-day sample period and a constant value of 0.20 for the true volatility, will be our standard of comparison in later experiments, so we refer to this as the Baseline simulation.

We have chosen a period of 3 months (63 trading days), which is a common sample size for estimating volatility. RiskMetrics, for example, uses an exponentially weighted moving average for volatility on a historical sample of 74 days, with a decay rate of 0.94. That methodology puts about 99% of the weight on observations within the first 63 days. (We will examine the effect of downweighting old data in this calculation below.) The reason to limit the data sample is to reduce the effect of time variation in the volatility. However, in the Baseline case, volatility is a constant parameter, so the way to reduce sampling error is simply to increase sample size.

Table 4 shows the effect of varying the sample size from 21 days to 250 days on the probability ratios, which are plotted in Figure 2. With a fixed 2 1/2 million returns, the number of non-overlapping samples decreases with the sample length, increasing the sampling error in our simulated "true" tail values. A 250-day sample would produce only 10,000 observations, making locating the 0.0001 tail problematical, so we increased the simulation sample size to 10 million for the K=125 and K=250 runs. Even so, in the latter case, the expected number of 0.0001 tail events was only 4, leading one to suspect that Table 4 may understate the true probability ratio for the $\alpha = 0.0001$, $K = 250$ combination (there were no events in the predicted 0.0001 tail in the first 2 1/2 million returns runs, for example). Nevertheless, the results indicate that using more data in the sample substantially reduces the problem of estimation error, making it relatively unimportant when volatility is estimated over the past year, until one gets out beyond the 0.001 tail.

Unfortunately, the Baseline case probably represents the best possible conditions for predicting the extreme tails of a returns distribution using the standard methodology. The true distribution is normal, the true volatility is constant, and the true mean is 0, so computing sample volatility as if the sample mean were zero actually saves a degree of freedom by imposing a true constraint. In the real world, it is safe to say that none of these conditions holds. Predicting the standard deviation of next period's return requires aiming at a moving target, and it is no longer automatically true that a longer sample period produces a more accurate estimate. Procedures that limit the amount of past data used, either simply cutting off the sample by using a fixed window, or downweighting older observations, may improve accuracy. We now turn to the estimation problem when volatility is allowed to change over time.

3. Estimating Risk Exposure when Volatility is Time-Varying

The returns model in equation (1) is the workhorse of continuous-time asset pricing, but the empirical evidence suggests that real-world asset price processes are more complex than this. In this section we will consider the estimation problem when volatility varies stochastically over time.

We will continue to leave aside the behavior of the drift term μ. Even though it should vary stochastically along with the riskless interest rate and, as we have argued, even a constant μ would be hard to predict accurately, the drift does not have a very large effect on volatility estimation when returns are sampled over short intervals. For Δt close to zero, it is of smaller order than the volatility (Δt versus $\sqrt{\Delta t}$). Only when the sample mean is very different from the true mean does the error have any appreciable effect on the sample volatility. This is why substituting 0 for the sample mean is an adequate fix for the estimation error.

A variety of alternatives to (1) that allow volatility to change over time have been explored in the literature. One of the simplest and most widely studied approaches is to model it as a mean-reverting square root process, as in Heston [1993]. We regard this as a reasonable assumption, that will allow us to explore the estimation risk problem when the parameter of interest drifts over time, but it is only one of many alternatives. More elaborate volatility models drawn from the recent empirical literature will be examined in Section 6.

We continue to assume volatility is forecasted using equation (3), or, later in the section, using a modified EWMA version of (3). The natural objection to this procedure is that if volatility is following a time-varying process, one should model that process and use it, rather than (3), to forecast the volatility. Predicting time-varying volatility with an estimator based on an incorrect constant volatility model is bound to lead to errors.

There are two strong reasons not to accept this seemingly logical point. The first is that what we are examining is what the great majority of VaR users in the real world actually do. The results we obtain will give us insight into how volatility estimation error actually affects the risk management process as it is widely implemented. This, of course, leads to the follow-on objection that real world risk managers who irrationally use wrong models would be better served by telling them to use right ones, than by carefully studying the properties of their bad ones.

But, the second, much stronger, argument against using a more sophisticated estimator than (3) is that I believe trying to solve the estimation error problem by more elaborate modeling is ultimately futile. The model of the real world that I have in mind is one in which virtually nothing is ever truly stationary. For whatever structural model of the volatility process we might decide is best, both theoretically and empirically--maybe some kind of multifactor jump-diffusion with long memory and non-Gaussian shocks--we should expect to find that its parameters change over time. And if we try to deal with that problem by modeling the parameter drift, the parameters of the parameter-drift process will themselves turn out to be nonconstant. In short, I believe that parameter drift is endemic to a financial system that is itself constantly changing and evolving. In trying to predict the future value of a variable of interest, one is always using data drawn from previous periods when the system was different.

Since we can not expect that important parameters will stand still while we measure them, forecast error can be limited only if information arrives rapidly relative to the rate at which the parameters are changing. It is this need for rapid information arrival that fails when we try to assess the probabilities of rare events. With instability in the data generating process, it becomes quite possible for less elaborate, but more robust, model specifications to outperform more theoretically correct ones in forecasting.

With a basic square root diffusion model. Equation (1) is replaced by

$$\frac{dS}{S} = \mu\, dt + \sqrt{V_t}\, dz \tag{6}$$

$$dV = \kappa(\overline{V} - V_t)\, dt + \theta \sqrt{V_t}\, dw \tag{7}$$

$\overline{V}$ is the long run variance, κ is the rate of reversion of the current variance V_t toward that long run value, $\theta\sqrt{V_t}$ is the volatility of the variance process and dw is a second Brownian motion, independent of dz.[13]

Equations (6) and (7) are discretized for the simulation study as follows.

$$r_{t+1} = \ln S_{t+1} / S_t = \mu\, \Delta t + \sigma_t\, \tilde{z}_t \sqrt{\Delta t}\,; \qquad \tilde{z} \sim N(0,1) \tag{8}$$

$$V_{t+1} = \kappa(\overline{V} - V_t)\, \Delta t + \theta\, \tilde{w}_t \sqrt{\Delta t}\,; \qquad \tilde{w} \sim N(0,1) \tag{9}$$

Values for mean reversion, κ, and the volatility of variance parameter, θ, in the simulations should be chosen to obey the Feller condition for overall stability of the variance process.[14] For the variance process to remain positive over the long run, we must have

[13] For equity returns, it is common to allow negative correlation between dz and dw. See, for example Bakshi, et al [1997]. We did explore allowing correlation between return and variance shocks in this model, but did not find striking differences from the results presented here. In the interest of limiting the amount of results to be presented, we consider only the case with independent shocks at this point. The models examined in Section 6 have negative correlation between dz and dw.

[14] See Feller [1951].

$$(10) \qquad \kappa \overline{V} > \frac{\theta^2}{2}$$

Otherwise, in finite time, variance converges to zero. Although we choose parameter values that satisfy (10), in a discrete simulation of equation (9), we still get occasional random draws for w_t that would produce negative variances. When that happens, we set V_{t+1} to 10^{-8}, essentially imposing the constraint that annual volatility can not be less than 0.01%.

Table 5 presents simulation results for the standard estimation technique based on equation (3), but with volatility that evolves over time according to equation (9). The user is assumed to calculate sample volatility from the last 63 days of simulated returns as if it were a constant parameter. He then estimates the location of the α-tail of the period t+1 returns distribution for a range of αs, treating the sample volatility as if it were the true volatility.

The first column shows the desired α values and the second column reproduces the Baseline results for a 63-day sample. Runs 1-9 assume four different values for the variance mean-reversion parameter κ: 0.20, 0.40, 1.0, and 2.5. For each κ, we show two or three values for θ, with the largest one in each case being within 0.05 of a value that would violate the Feller condition.

Not surprisingly, allowing variance to change over time increases the RMSE of the forecasted volatility, to about 2.3% when $\theta = 0.10$ and to up to 5.45% with $\theta = 0.20$. The results show the substantial impact of stochastic volatility on predicted probabilities of tail events. And, as we suggested above, the Baseline simulation represents the best case for the standard approach to risk assessment.

We do not know what values of κ and θ would hold for real world asset returns. Bakshi, Cao, and Chen [1997] present estimation results for a variety of stochastic volatility models for the S&P 500 stock index, including one that is close to (9). Their values are $\kappa = 1.15$ and $\theta = 0.39$, but there are qualifications that make it not completely appropriate to take these values as good real world estimates for the parameters in our problem. First, the parameter values are obtained by implying them out from S&P 500 index option prices, not by fitting them to actual returns data. Second, the Bakshi, et al specification of the returns process includes a strong negative correlation between dz and dw, while we have modeled them as independent. Finally, with our value for long term variance, the combination of parameters in Bakshi, et al would violate the Feller condition and cause the variance process to be unstable. In Section 6, we simulate returns from the Bakshi, et al model along with several others from the literature and examine its properties more closely.

Figure 3 plots the effect of volatility of variance on the true / predicted probability ratios. We set $\kappa = 1.0$ and plot the results for a range of θ values, from the Baseline case ($\theta = 0$)

to $\theta = 0.25$. (It may be useful to refer back to Table 1 to get a feel for what the errors in the estimated tail probabilities mean in practical terms.)

In Figure 4, we examine the effect of changing the mean reversion of variance while holding θ fixed at 0.10. It is not surprising that larger volatility of variance makes forecasting harder, as we see in Figure 3. It is less clear what one should expect for the rate of reversion toward long run variance. On the one hand, a rapid rate of reversion tends to keep the process closer to its long term (constant) value, which should make forecasting easier. On the other hand, a larger κ also means that when instantaneous variance differs from $\overline{V}$, it will drift more rapidly over the sample period under the force of mean reversion, which could make post-sample forecasting harder. As Figure 4 shows, the former effect appears to win out, at least with these parameter values: higher κ reduces the impact of estimation error in calculating the location of the α-tails. Even so, in Table 5 it is clear that, overall, time varying volatility makes the problem of assessing risk exposure worse.

4. Further Departures from the Standard Model

The standard assumption that returns come from a normal distribution is made for mathematical convenience, but a large amount of statistical evidence indicates that the true distribution is more fat-tailed than the normal. Time-variation in the variance is one reason for apparent non-normality, but frequently the returns shocks appear fat-tailed even in models with explicitly stochastic volatility. A convenient alternative to the normal is the Student-t distribution, which has one additional parameter, the degrees of freedom(d.f.), that governs the tail behavior. The distribution converges to the standard normal as d.f. goes to infinity, but for small values, the t-distribution is distinctly fatter-tailed than the normal. Indeed, all moments greater than d.f. are infinite. For example, a t(3) has finite mean, variance and skewness, but infinite kurtosis and higher moments.

Table 6 examines the effect of drawing the disturbances in the returns equation from a t-distribution with either 7 or 4 degrees of freedom, compared with normal (0,1) shocks. The shocks to the variance equation are still drawn from a standard normal distribution. We consider three cases: constant parameters, stochastic volatility with a relatively low volatility of variance $\theta = 0.05$ and slow mean reversion $\kappa = 0.40$, and more strongly stochastic volatility with $\theta = 0.20$ and $\kappa = 2.5$. The results show that fat-tailed disturbances significantly worsen the underestimation of exposure to extreme returns. Even with the low volatility of variance process, the true probability of experiencing a 0.0001-tail event is more than 20 times greater than is predicted under the assumption that returns are normal. However, comparing across the different runs it is clear that this result is due much more to the tail-fatness of the t-distribution than to the problem of sampling error that we have been examining. Given the degrees of freedom in the distribution, allowing time variation in the volatility makes little difference to the results. For example, with t(7) shocks, under constant variance the true probability of a return in the 0.0001 tail is about 23 times the probability predicted from a normal distribution. When $\theta = 0.05$, the same multiple of about 23 applies, and increasing the volatility of variance to 0.20 only moves the ratio to 25.5. Assuming the returns distribution is normal when return shocks actually come from a t-distribution leads to huge underestimates of the exposure to large low probability events, but the additional risk exposure that can be attributed to estimation risk is not very important.

The fact that variance is known to be time-varying has led to the use of alternative estimation techniques to reduce the problem. Use of a fixed window, such as 63 days, makes sense if it is felt that restricting the returns sample to recent data give a better estimate of a moving parameter. A fixed window imposes a rather artificial weighting of past data in the estimation, either 1 for an observation in the window, or 0 for one outside the window. A common alternative that downweights data more smoothly as it ages is to use an exponentially weighted moving average.

Under EWMA, each observation is downweighted at a fixed rate of decay as it ages. The volatility estimate is given by

$$(11) \qquad \sigma_{\text{predicted}} = \sqrt{\sum_{t=1}^{K\max} \lambda^{t-1} (r_t)^2 \Big/ \sum_{t=1}^{K\max} \lambda^{t-1}} \ ,$$

where λ is the decay parameter and Kmax is the maximum lag included in the calculation. RiskMetrics, which has made a profitable and influential business out of estimating volatilities and correlations for use in Value at Risk calculations, uses $\lambda = 0.94$ and Kmax = 74 for all of its daily volatility estimates, the latter chosen because with $\lambda = 0.94$, if an infinite number of past observations were available, the total weight applied to those more than 74 days old would be less than 1%. According to RiskMetrics [1996], tests on volatilities from a large number of return series indicate that $\lambda = 0.94$ seems to give the best average forecast performance. In order to compare with the earlier results, we use Kmax = 63 here, which captures most of the weight of an infinite sample, with the fraction ranging from about 72% for $\lambda = 0.98$ to more than 99.8% for $\lambda = 0.90$.

An EWMA offers the possibility of extracting some volatility information from comparatively old data while recognizing that more recent data probably contain more information that is relevant for predicting next period's volatility. The optimal decay factor should be a function of the rate of change of volatility and the size of the stochastic component. A relatively large value for λ, close to 1.0, would be appropriate for a stable and slow moving variance process, while if volatility changes rapidly over time, one would like to reduce the weighting of older data by using a smaller λ.

In Table 7, we present simulation results for three volatility regimes, comparing four decay factors. In results not shown here, we found that varying the rate of mean reversion κ had very little effect on the estimation error in the tail estimates for this case. For example, with $\lambda = 0.94$ and $\theta = 0.10$, κ values of 0.20, 1.0, and 2.5 produced forecast RMSEs of 0.0270, 0.0269, and 0.0267, respectively, and the 0.0001-tails fell at 4.269, 4.277, and 4.273. Given the minuscule effect of changing the rate of variance mean reversion over a broad range, we only report results with $\kappa = 1.0$. θ values were set to 0 (constant volatility), 0.05 (relatively stable variance) or 0.25 (volatile variance), and we considered λs of 1.0 (no downweighting in a fixed 63-day window), 0.97, 0.94 and 0.90.

Not surprisingly, if volatility is constant, downweighting older data points simply throws away useful information, since every observation contains an equal amount of information. The RMSE results show that with $\theta = 0$ forecast accuracy diminishes monotonically as λ is reduced from 1.0 to 0.90, and the probability ratios are consistent with this. The same result holds for the low volatility of variance $\theta = 0.05$ regime. However, in the high θ regime, the pattern is different. Forecast accuracy is better for $\lambda = 0.97$ and $\lambda = 0.94$ than it is for either $\lambda = 1.0$ or $\lambda = 0.90$. Evidently, in this case no downweighting ($\lambda = 1.0$) allows too much noise from obsolete data points into the calculation, while too much downweighting ($\lambda = 0.90$) excludes too much useful information that could have been extracted from observations that are not old enough to have lost their value. This suggests that we would find a similar result for the $\theta = 0.05$ case, for some λ values between 0.97 and 1.0.

Figures 5 and 6 do just that, reporting the true / predicted probability ratios for the low and high θ regimes, respectively. Figure 5 shows that while the best performance for the nearby 0.05 to 0.0005 tails is achieved with no downweighting (or at least, with a decay factor over 0.99), for the further 0.0002 and 0.0001 tails, it is better to use a decay factor of 0.99 than to weight each observation equally. Figure 6 shows that this general pattern is similar and more pronounced when θ is relatively high. The most extreme tails are estimated more accurately using EWMA, with a decay parameter of 0.92 or 0.94. The nearer tails also are more accurate, but with less downweighting. The best λ values in Figure 6 are $\lambda = 0.96$ for the 0.01 tail and $\lambda = 0.97$ for the 0.05 tail. Thus, EWMA appears to give some improvement in tail estimation under conditions of time-varying variance. The overall impact of estimation error on predictions of risk exposure, however, is still very large.

5. Autocorrelation in the Volatility Forecast Errors from a Rolling Sample

One feature of our research design that affects the results substantially is the fact that, so far, we have used only non-overlapping samples. This allowed us to compute the effect of estimation error on the predicted probabilities without the problem of serial dependence that overlapping samples would produce. This is both good and bad. With non-overlapping samples we get a better estimate of the properties of the probability distribution of the one-day VaR calculation in the presence of estimation error, but the forecasting problem that we are modeling is different from what risk managers in the real world actually do. A firm that uses VaR as a risk management tool will reestimate volatility regularly, probably every day, to estimate the exposure for the immediate future. Each time, the most recent day's observation is added to the sample and the oldest day is dropped. This means that the prediction error on date t will be highly correlated with the error on date t-1, perhaps generating a string of volatility underestimates, and multiple tail events. This section explores that issue.

We simulate 250,000 daily returns (1000 years) using the same procedures as above, but then consider estimating volatility from a rolling 63-day sample that is updated each day. This produces (250,000 - 63) = 249,937 1-day VaR forecasts, whose prediction errors will be serially correlated. Table 8 presents the probability ratios for the same set of parameter values as in Table 5. For ease of comparison, the results from Table 5 are repeated here, in the rows labeled "No overlap."

Average prediction accuracy, as measured by the root mean squared error of the volatility forecast, is very similar for the overlapping and non-overlapping samples. Although there are periods when volatility is underestimated, which increases the likelihood of observing what appear to be multiple tail events within a short time interval, these are balanced by periods of overestimated volatility, with a lower than expected chance of an event. Overall, the RMSE of the volatility estimate is not affected very much. In other words, using a rolling sample does not increase the bias of the volatility estimate. Nor does it seem to exacerbate the problem in the Baseline constant volatility case, where the probability ratios for the remote tails are smaller in the overlapping sample.

But once volatility is allowed to vary over time, the offsetting of under- and overestimates in the rolling sample does not produce offsetting errors in estimating the tails of the distribution. Under stochastic volatility, a rolling sample produces many more tail events than were shown in Table 5. Even a low value of θ leads to a substantial increase in the probability ratio, with the difference increasing as one looks further into the tail. For example, with $\theta = 0.10$ and $\kappa = 0.4$ (Run 4), a rolling 63-day sample would experience 58% more 5% tails events than expected (versus only 3% more with no overlap). But at the one-in-a-thousand 0.1% level, the rolling sample would produce more than 11 times as many events as expected, while the non-overlapping sample only experiences twice as many. A faster rate of volatility mean reversion κ mitigates the effect considerably, but even with $\kappa = 2.5$, a rolling sample still produces much larger tail probabilities than the non-overlapping sample.

So far we have been examining results only from simulations. This raises the question of whether these experiments really reflect what happens in the real world. To provide some evidence on this issue, I fitted rolling volatility forecasts on about 40 years of S&P 500 stock index returns, and examined the tail predictions using the same kind of analysis we have been considering. The sample period is July 2, 1962 through August 30, 2002, which yields 10113 observations. Rolling estimations were done using returns from the previous 21, 63 and 250 days. Table 9 presents the results.

Table 9 shows clearly that the standard procedure of estimating volatility over a relatively short historical period and rolling the sample forward each day leads to serious underestimates of the tail probabilities in the real world, just as it does in our simulations. The more remote tails are underestimated to a larger degree, but even the 1% tail had more than 70% more events than were predicted using any of the three sample sizes. One noteworthy feature here is that adding more data by going from 21 to 63 to 250 day estimation periods only improves the tail predictions very slightly. This suggests that the problem is not just sampling error, which can be made to go away by using more data points, as was shown in Table 4. Time variation in the volatility, which does not disappear with a longer estimation sample, is likely to be playing an important role, as well (and, quite likely, fat tails in the returns distribution, too).

Risk exposure is not limited to getting one really bad day. When statistically independent volatility forecasts are produced from non-overlapping data, the probability of getting two tail events in a row is just the square of the probability of one event. But with a rolling sample, the volatility forecast errors will be highly positively autocorrelated. This will produce a much greater chance of getting multiple events over a short period than with independent forecasts. Table 10 examines this phenomenon.

Panel A presents results on the occurrence of two tail events in a row and Panel B does the same for three events in a row. The Probability under Independence shown in column 2 is just the tail probability raised to the power 2 or 3. The remaining columns show the probability ratios when volatility is estimated with overlapping samples for some of the asset price processes examined above. The first is the constant volatility Baseline run. These results show that sampling error alone leads to a substantially higher multi-day risk than expected. For example, three 1% tail events in 3 days should be a one in a million event, but because of sampling error it is 12 times more probable than that, even when volatility is constant and returns are normal.

The next four columns give the probability ratios for different values of θ and κ, ranging from a relatively low θ of 0.10 with κ of 0.2 or 1.0, to high values of $\theta = 0.40$ and $\kappa = 2.50$. The effect is striking. If either θ is high, or θ is moderate but mean reversion is slow, the probability of two or three events in a row grows sharply. For both Run 1 and Run 4, the "one in a million" occurrence of three 1% tail events in a row is actually much closer to 1 in 1000--something that might happen about every 4 years on average.

Finally, the last column gives statistics on actual multi-day tail events observed for the S&P 500 stock index over the 40 year sample period. The results are not as extreme as some of the simulations, but are more extreme than others. Three 5% tail events in a row, for example, should happen only once in 8000 days, or about 32 years. But three in a row was actually almost 10 times more frequent than that, averaging 1 in about 3 1/2 years.

Panels C and D show the same kind of results for two relatively more likely multi-day tail events: 3 events in 5 days and 3 events in 10 days. The theoretical probability of K events in N days under independence can be computed directly from the binomial distribution, with the probability of a single event set equal to the tail cutoff.

If α is the tail cutoff probability, the probability of observing K (or more) events in N days $P_{K,N}$ is given by

$$(12) \qquad P_{K,N} = \sum_{k=K}^{N} \binom{k}{N} \alpha^{k} (1-\alpha)^{N-k}$$

One thing that complicates the interpretation of these results a little is the fact that with a rolling estimation, the same events can be counted more than once. For example, if there are three events in a row, there will be 7 days in the sample in which those 3 events will fall within a 10 day window. However, these "multiple counting" cases will tend to be offset by multiple-under counting periods that will occur, as well. Asymptotically, this should not bias the probability ratios as reported in Table 10. Again, these results indicate that the use of a rolling sample to estimate sequential volatilities and risk exposures increases the problem of estimation risk.

6. Estimation Error from Jump Risk

Equation (1), the returns equation assumed by Black and Scholes for stock price dynamics, has become the new classical model for asset returns. However, empirical research has established that it does not fully capture the behavior of actual stock price movements. Time-varying volatility was accommodated by models like (6) - (7), in which returns still follow a diffusion, but with stochastic volatility. But asset prices also make occasional large jumps, which is inconsistent with a diffusion. A number of recent papers have proposed and estimated jump-diffusion models in which the return and in some cases, the variance, can experience occasional large Poisson jumps. In this section we examine the performance of the standard VaR risk calculation when returns are driven by a jump-diffusion process.

There are many different ways to build jumps into a model of returns, each involving several new parameters. Rather than trying to pick a single model, or presenting results from a range of models with a range of parameter values for each, we will take model specifications and the estimated parameter values from several papers in the literature, specifically Eraker, Johannes, and Polson [2003] (EJP), Andersen, Benzoni, and Lund [2002] (ABL) and Bakshi, Cao and Chen [1997] (BCC).[15] Each of these papers fits several alternative specifications for S&P 500 stock index returns. Adapting the models to our notation, they can all be expressed as special cases of the following general jump-diffusion model. The returns equation is

$$(13) \qquad \frac{dS}{S} = \mu_t \, dt + \sqrt{V_t} \, dz + J_t \, dq$$

The variance equation is either (14) or (15). All three papers fit versions of (14); ABL also estimate (15), which they find fits their data better, so that is the model we examine from their paper.

[15] Duffie, Pan and Singleton [2000] provide the technology for relatively straightforward construction of models with jumps in returns and volatilities, driven by an arbitrary number of underlying factors, so long as they are members of the affine class. Other papers fitting jump-diffusion models of S&P 500 index returns include Bates [2000] and Pan [2002].

$$(14) \quad dV_t = \kappa(\overline{V} - V_t)\,dt + \theta\sqrt{V_t}\,(\rho\, dz + \sqrt{1-\rho^2}\, dw) + J_t^V\, dq_V$$

$$(15) \quad d\ln V_t = \kappa(\ln \overline{V} - \ln V_t)\,dt + \theta(\rho\, dz + \sqrt{1-\rho^2}\, dw)$$

The parameters that appear in (6) - (7) have the same meaning here. The diffusive components of the returns and variance equations are now correlated, with instantaneous correlation coefficient ρ. The asset price can also make discrete jumps, governed by the Poisson process dq, which takes the value 1 with intensity λdt. Conditional on a jump occurring, the stochastic jump size is J_t , drawn from a probability distribution that differs across models. ABL and BCC assume return jumps are lognormal. EJP model them as normal, but in discretizing (13) for estimation, they replace dS / S by $\ln(S_{t+1}) - \ln(S_t)$, so the result is similar to assuming lognormality. The probability distributions for jump size are as follows:

(16a) [ABL, BCC] $\quad \ln(1 + J_t) \sim N(\ln(1 + \mu_J) - \sigma_J^2 / 2, \sigma_J^2)$

(16b) [EJP] $\quad J_t \sim N(\mu_J, \sigma_J^2)$

EJP estimate models in which the variance may also jump, according to a Poisson process dq_V, with intensity λ_V. Conditional on a volatility jump occurring, the jump size is J_t^V, a random draw from an exponential distribution with mean ν. Two versions of the model with variance jumps are considered, one in which jumps in returns and jumps in variance are independent, and one with correlated jumps. In the latter case, $dq_V \equiv dq$ and a jump, if it occurs, strikes both return and variance at the same time.

Each of the papers fits a variety of specifications, from which we have selected a subset chosen to explore different assumptions. Table 11 shows the models and parameter values used in each. EJP use Markov Chain Monte Carlo estimation to fit their models to the daily change in the log of the S&P 500 index, over the period 1980 - 1999.

We consider four model variants from EJP. SV (stochastic volatility with no jumps) is similar to the pure diffusion model (6) - (7), with the exception that the shocks in the returns equation and the variance equation are negatively correlated, with $\rho = -0.397$. SVJ allows jumps in the returns equation, but the variance continues to follow a pure square root diffusion. In this model, there can be occasional large price shocks, but they don't affect the diffusive variance. But real world markets seem to experience periods in which variance itself rises sharply. EJP model this phenomenon by adding a jump component to the variance equation. SVIJ allows jumps in both returns and variance, with independent jump arrivals. If a variance jump occurs, the variance of the diffusive component of returns rises sharply, and then gradually decays at a rate determined by κ. But it is hard to understand how a variance jump might occur in an actual securities market without there also being a large shock to returns. EJP's SVCJ model allows jumps in returns and variance but constrains them to occur at the same time.

ABL also model daily S&P 500 returns, from 1953-1996. They consider models with and without jumps, with two different variance equations, as explained above. We simulate their SV and SVJ models with (15) as the variance equation. One advantage of modeling the change in the log of variance is that it is not necessary to constrain the right hand side of equation (15) in estimation or simulation to prevent variance from going negative. ABL use the Efficient Method of Moments estimation technique.

BCC imply out their model parameter values from prices of S&P 500 index options. This makes their models potentially quite different from the others. The sample period is only three years, June 1989 - May 1991. Unlike EJP and ABL, who estimate a single parameter for the drift in the returns equation, BCC use the current market riskless interest rate, modified to adjust for the impact of jumps on the total return. We prefer to simulate using a single value for μ, so we set the drift equal to 5% in BCC's SV model, and to 5% - $\lambda\mu_J$ = 8.54% in their SVJ model.

The occurrence of a (large negative) jump in returns will produce a Value at Risk tail event. A return jump of either sign will also enter the data sample used to estimate future volatilities, until it drops out of the estimation window. This will lead to overestimates of the diffusive volatility and fewer perceived tail events than expected, so long as no further jump occurs. If a second large jump does occur, however, it will also be a tail event, since estimating volatility from a large number of returns generated by a diffusion along with a single jump return will still understate the jump variance. Note, however, that in all of these models, jumps can come in all sizes, including quite small ones, that might be hard to distinguish from a large 1-day realization of the diffusive shock. This pattern of estimation error can lead to fewer not very rare (e.g., 5%) tail events than expected, but more, and larger, events from the more remote tails.

Table 12 gives the results of simulating returns from each of the models, with the parameter values shown in Table 11, over 250,000 consecutive days, and estimating VaR values from a rolling sample under the assumption of constant volatility and normal returns. In Panel A, the sample period is 63 days. In all cases, the root mean squared error of the estimated volatility as a forecast of realized volatility is quite large, more than 4%. However, the probability ratios for 5% Value at Risk do not look bad. In fact, for four of the seven cases, fewer than 5 percent of the realized returns end up in the predicted 5% tail. Thus, an outside monitor, such as a bank regulator, might well conclude that a risk management system with these properties was functioning properly.

But stochastically time-varying volatility, especially coupled with occasional large jumps in returns, causes tail risk to be much greater than the 5% VaR figures suggest. Even 1% events are more than 50 percent more prevalent than expected, in most cases, and the more disastrous events, that should only occur once in many years, actually can be anticipated to happen on average every 1-2 years.

Comparing across models, the one showing the smallest problems is the EJP version of the SV model. It does not allow jumps, and it also has quite a rapid rate of mean reversion in the variance, which allows it to give more accurate estimates than the BCC

SV equation. Unfortunately, all three articles report that their SV specifications are strongly rejected by the data, when compared to models with jumps.

Models that allow jumps show a large amount of tail risk, for the reasons explained above. ABL's SVJ model has the largest estimated jump frequency, leading to the worst underestimation of large tail events. Allowing jumps in the variance equation in EJP's SVIJ and SVCJ models permits them to fit the returns data better than the more restrictive specifications, but if anything, the tail risk problem becomes worse. There is little difference between the independent and correlated variance jumps models in this dimension.

Panel B shows the results of lengthening the sample horizon to one year (250 days). The results are largely unchanged from Panel A. The time variation in risk characteristics in the real world, as captured by these models of the US stock market, appears to have eliminated the possibility of improving accuracy by simply adding more data. Panel C examines using an exponentially weighted moving average to estimate volatility, with the 74 day window and 0.94 decay factor favored by RiskMetrics. In most cases, this appears to improve performance somewhat.

7. Conclusion

Statistical procedures to quantify exposure to financial risk have been widely adopted by real-world risk managers, with Value at Risk probably representing the single most common technique at present. Alternatives to VaR also involve trying to estimate the tails of a probability distribution of asset values or returns. This inevitably entails estimation error, whose effect is seldom considered explicitly. But we have seen that even with constant volatility and normal distributions, the standard VaR estimation technique, based on daily updating of a volatility estimate drawn from a few months of recent historical data, leads to substantial misestimation of tail probabilities. The problem grows worse the farther into the extreme tails one looks.

Empirical evidence shows that security returns in the real world depart in important ways from the assumptions plain vanilla VaR is based upon. The major focus of this paper has been on stochastically time-varying volatility. It is not plausible to expect volatility, nor any other property of the returns generating process, to remain constant over long periods while major economic, technological, and political changes take place. When the financial environment evolves randomly, the error in estimating the probabilities of rare events gets much larger, and the possibility of increasing accuracy simply by using longer data samples disappears.

Departures from the assumed distribution of return (and volatility) innovations also increase estimation error. Fat-tailed return shocks-- one symptom of time variation in parameters--and, especially, jumps produce large tail events that are more frequent than predicted. Events that are expected to occur "once in a lifetime" may actually be more likely to happen every year or two. Unmodeled return jumps can be particularly insidious, because they can lead to fewer than expected medium-sized 5% tail events but, at the same time, much larger numbers of very serious ones. On the other hand, while we can, and should, build more sophisticated models that incorporate time-variation and non-Gaussian return shocks, the world continues to evolve as it is generating the data we will need to fit those models. The models from the recent finance literature examined in Section 6 all fit the data significantly better than the constant parameter models they are designed to replace, but they are also nearly all statistically rejected themselves--in in-sample tests, which assume stationarity in the data generating process.

How should one conduct practical risk management given these results? Several principles are suggested. First, we need increased awareness that estimation risk exists and that it is real risk. It is not the case that overestimates and underestimates of VaR due to sampling error will wash out over time, and that tail estimates will be "right, on average." Second, stress testing of parameter estimates, not just of economic scenarios, is probably useful to see how estimation error affects tail risk assessments. Third, alternative estimation procedures for tail risk may help.

A common alternative to estimating VaR parametrically is historical simulation, using past returns from a much longer sample period, typically several years, and simply tabulating the empirical returns distribution. This may ameliorate some of the estimation risk problems that we have seen with volatility estimation on a short data sample. For example, if return shocks are non-normal, using a long sample period will allow a better empirical fit to the actual tail behavior. Also, if volatility varies stochastically but the rate of mean reversion is fairly rapid, a sample of several years of returns may produce a reasonably good estimate of the ergodic distribution. On the other hand, it is not possible to say anything about the remote tails from a limited sample. For example, in 2 years of data, one would not expect to see even one 0.1% event. Thus, historical simulation needs to be combined with some way to model the remote tails, such as extreme value theory. Also, the problem of serial correlation of the errors should be considerably worse with longer historical samples. Estimation risk in hybrid models of risk exposure combining theory and historical simulation will be explored in subsequent research.

References

Andersen, T., L. Benzoni, and J. Lund. "Towards and Empirical Foundation for Continuous-Time Equity Return Models." Journal of Finance 57, 2002, 1239-1284.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. "Coherent Measures of Risk." Mathematical Finance 9, 1999, 203-28.

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen. "Empirical Performance of Alternative Option Pricing Models." Journal of Finance 52, 1997, 2003-2049.

Basle Committee on Bank Supervision. "Amendment to the Capital Accord to Incorporate Market Risks." Basle: Bank for International Settlements, January 1996. (http://www.bis.org/publ/bcbs24a.htm)

Bates, D. "Post-'87 Crash Fears in S&P 500 Futures Options." Journal of Econometrics 94, 2000, 181-238.

Bawa, Vijay, Stephen Brown and Roger Klein. Estimation Risk and Optimal Portfolio Choice. Amsterdam: North Holland Publishing Co., 1979.

Bollerslev, Tim, R. Chou and Ken Kroner. "ARCH Modeling in Finance: A Selective Review of the Theory and Empirical Evidence." Journal of Econometrics 52, 1992, 5-59.

Boudoukh, Kobi, Matthew Richardson, and Robert Whitelaw. "The Best of Both Worlds." RISK, May 1998, 64-67.

Duffie, Darrell and Jun Pan. "An Overview of Value at Risk." Journal of Derivatives 4, Spring 1997, 7-49.

Duffie, D., J. Pan, and K. Singleton. "Transform Analysis and Asset Pricing for Affine Jump-Diffusions." Econometrica 68, 2000, 1343-1376.

Eraker, B., M. Johannes, and N. Polson. "The Impact of Jumps in Volatility and Returns." Journal of Finance 58, 2003, 1269-1300.

Feller, W. "Two Singular Diffusion Problems." Annals of Mathematics 54, 1951, 173-82.

Heston, Steven. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options." Review of Financial Studies 6, 1993, 327-343.

Jorion, Philippe. "Risk^2: Measuring the Risk in Value at Risk." Financial Analysts Journal, Nov/Dec 1996, 47-56.

Jorion, Philippe. Value at Risk: The New Benchmark for Controlling Market Risk. Chicago: Irwin, 1997.

Karatzas, Ioannis and Steven Shreve. Brownian Motion and Stochastic Calculus 2nd ed. New York: Springer-Verlag, 1991.

Kupiec, Paul. "Techniques for Verifying the Accuracy of Risk Measurement Models." Journal of Derivatives 2, Winter 1995, 73-84.

Pan, J. "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study." Journal of Financial Economics 63, 3-50.

RISK Publications, "Value at Risk," RISK Supplement, June 1996.

RiskMetrics. RiskMetrics Technical Document, Fourth Edition. New York. 1996. (http://www.riskmetrics.com/techdoc.html).

Schachter, Barry. Website: <http://www.gloriamundi.org/>

Schwert, G. William. "Why Does Stock Market Volatility Change over Time?" Journal of Finance 44, 1989, 1115-1154.

Schwert, G. William. "Stock Market Volatility." Financial Analysts Journal 46, 1990, 23-34.

Theil, Henri. Principles of Econometrics. New York: John Wiley and Sons, 1971.

Appendix

Proof that when r is drawn from a normal distribution with constant mean and variance, the conditional distribution of r_{t+1}, given $\{r_{t-K+1}, \ldots , r_t\}$ is Student-t.

The returns $\{r_\tau\}$ are independent draws from a normal distribution with mean μ and standard deviation σ.

We will use the following result from Theil [1971, p. 82]. Let X be a standardized normal variate, let Y^2 have a χ^2 distribution with K degrees of freedom, and let X and $Y = +\sqrt{Y^2}$ be independent. Then $X\sqrt{K}/Y$ is distributed as Student-t with K degrees of freedom. Here, X will be the standardized forecast error for r_{t+1} and Y will be the estimated standard deviation computed from the most recent K returns $\{r_{t-K+1}, \ldots, r_t\}$.

<u>Case 1</u>: Both mean and standard deviation are sample estimates.

$$\hat{\mu} = \frac{1}{K}\sum_{\tau=1}^{K} r_{t-\tau+1} \qquad \hat{\sigma} = \sqrt{\frac{1}{K-1}\sum_{\tau=1}^{K}(r_{t-\tau+1}-\hat{\mu})^2}$$

The first post-sample return, r_{t+1}, is independent of the returns used in computing the sample mean and variance. The expected value of $\hat{\mu}$ is μ and variance of $\hat{\mu}$ is σ^2/K. This means

$$(r_{t+1} - \hat{\mu}) \sim N(0, \sigma^2 + \sigma^2/K) \quad \text{and} \quad \sqrt{\frac{K}{K+1}}\,\frac{r_{t+1}-\hat{\mu}}{\sigma} \sim N(0,1).$$

From basic statistics, (e.g., Theil [1971], p. 91), we have $\dfrac{\hat{\sigma}^2}{\sigma^2(K-1)} \sim \chi^2(K-1)$.

Applying the result stated above and simplifying gives $\sqrt{\dfrac{K}{K+1}}\,\dfrac{r_{t+1}-\hat{\mu}}{\hat{\sigma}} \sim t(K-1)$. The return r_{t+1} is distributed like a Student-t with K-1 degrees of freedom, but scaled up by the factor $\sqrt{\dfrac{K+1}{K}}$. It has the same zero mean as a t(K-1) variate but its standard deviation is larger, making the distribution a mean-preserving spread on a standard Student-t(K-1). The distribution of r_{t+1} has fatter tails than the normal, and because of the scaling factor each quantile in the tail (5%, 1%, etc.) is more negative than the corresponding quantile for a standard Student-t(K-1).

Case 2: The sample mean is set to zero; only the standard deviation is estimated: $\hat{\sigma} = \sqrt{\frac{1}{K} \sum_{t-K+1}^{t} r_{\tau}^{2}}$. If the true mean $\mu = 0$, the constraint is true and a similar computation as in Case 1 yields $\frac{r_{t+1}}{\hat{\sigma}} \sim t(K)$

Case 3: If the true mean is nonzero, suppressing calculation of the sample mean in the estimation procedure introduces a specification error and the proof does not go through. The sample variance is a biased estimate of the true variance and $\frac{r_{t+1}}{\hat{\sigma}}$ will not satisfy the conditions of the theorem.

Table 1: Tail Probabilities, Frequency of Occurrence, and Failure Times

The table shows the risk characteristics for different tail probabilities α.

K_q is a measure of failure time. An event with a one-period probability α has a probability q of occurring within the next K_q periods.

K_q is the solution to: $q = 1 - (1 - \alpha)^{K_q}$

Prob-ability α	Estimated Frequency $1 / \alpha$	$K_{0.50}$	$K_{0.05}$
.05	20 days	14 days	1 day
.01	100 days	69 days	5 days
.005	200 days	139 days	10 days
.002	2 years	1.4 years	26 days
.001	4 years	2.8 years	51 days
.0005	8 years	5.5 years	103 days
.0002	20 years	13.9 years	1.0 years
.0001	40 years	27.7 years	2.1 years

Table 2: Predicted and Actual Tail Risk when Volatility is Mis-Estimated

The table shows the predicted and actual tail characteristics for different desired probabilities α, when volatility is underestimated.

True volatility σ_{true} = 20.0%. Predicted volatility $\sigma_{predicted}$ = 17.68%.

The actual α-tail cutoff, expressed as a multiple of the predicted volatility, is given by $N^{-1}[\alpha] \dfrac{\sigma_{true}}{\sigma_{predicted}}$, where $N^{-1}[.]$ denotes the inverse normal distribution.

Probability α	Predicted α-tail cutoff	Actual α-tail cutoff	Fraction of actual returns in predicted α-tail	True Prob / Predicted Prob
.05	-1.645	-1.86	0.0730	1.46
.01	-2.326	-2.63	0.0199	1.99
.005	-2.576	-2.91	0.0114	2.28
.002	-2.878	-3.26	0.0055	2.74
.001	-3.090	-3.50	0.0032	3.15
.0005	-3.290	-3.72	0.0018	3.63
.0002	-3.540	-4.01	0.00088	4.38
.0001	-3.719	-4.21	0.00051	5.05

Table 3: Constant Volatility Baseline Simulation

Simulation: Sequential returns are simulated for a period of 2,500,000 days (10,000 years).
Estimation sample: K = 63 days; 39,682 non-overlapping intervals.
True Volatility: σ = 0.20; True mean: μ = 0; Sample mean is not estimated.
Tail cutoffs are expressed as multiples of the predicted standard deviation.

RMSE of volatility estimate = 0.0178

Probability α	Predicted α-tail cutoff	Actual α-tail cutoff	Fraction of actual returns in predicted α-tail	True Prob / Predicted Prob
0.05	-1.645	-1.649	.05048	1.01
0.01	-2.326	-2.384	.01173	1.17
0.005	-2.576	-2.660	.00615	1.23
0.002	-2.878	-3.015	.00292	1.46
0.001	-3.090	-3.297	.00176	1.76
0.0005	-3.290	-3.514	.00102	2.03
0.0002	-3.540	-3.777	.00044	2.21
0.0001	-3.719	-4.144	.00024	2.37

Table 4: Probability Ratios for Different Estimation Samples Sizes

Simulation: Sequential returns are simulated for a period of 2,500,000 days. (10,000,000 days for K = 125 or 250)
Estimation sample: Non-overlapping samples of K days, K = 21, 42, 63, 125, 250
True Volatility: $\sigma = 0.20$; True mean: $\mu = 0$; Sample mean is not estimated.

	Probability ratio K = 21	Probability ratio K = 42	Probability ratio K = 63	Probability ratio K = 125	Probability ratio K = 250
Runs	119047	59523	39682	80000	40000
Forecast RMSE	0.0307	0.0219	0.0178	0.0127	0.0089
Prob= 0.05	1.15	1.05	1.01	1.04	1.03
0.01	1.51	1.24	1.17	1.06	1.07
0.001	3.03	1.71	1.76	1.15	1.14
0.0002	5.43	1.99	2.21	2.21	1.66
0.0001	6.94	2.54	2.37	2.44	1.64

Table 5: Probability Ratios for Different θ and κ Values

Simulation: Sequential returns for 2,500,000 days; Estimation sample: 63 days; 39,682 non-overlapping periods.
True Volatility: $\sigma = 0.20$; True mean: $\mu = 0$; Sample mean is not estimated.

	Baseline	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9
θ (Volatility of variance)	0	0.05	0.10	0.05	0.10	0.10	0.20	0.10	0.20	0.40
κ (mean reversion)	0	0.2	0.2	0.4	0.4	1.0	1.0	2.5	2.5	2.5
RMSE	0.0178	0.0193	0.0230	0.0192	0.0229	0.0227	0.0331	0.0222	0.0318	0.0545
Prob= 0.05	1.01	1.02	1.03	1.02	1.03	1.03	1.07	1.02	1.05	1.15
0.01	1.17	1.14	1.26	1.15	1.21	1.19	1.43	1.19	1.32	1.86
0.002	1.46	1.55	1.82	1.56	1.72	1.69	2.43	1.66	2.07	3.76
0.001	1.76	1.70	2.12	1.70	2.03	1.88	2.95	1.84	2.34	5.56
0.0005	2.03	2.02	2.69	2.05	2.36	2.18	3.59	2.17	3.12	7.57
0.0002	2.21	2.39	3.50	2.39	3.27	3.06	5.36	2.70	4.05	12.65
0.0001	2.37	2.45	3.35	2.40	2.59	2.95	7.07	2.70	4.94	19.06

Table 6: Probability Ratios under Student-t Return Shocks

Simulation: Sequential returns for 2,500,000 days; Estimation sample: 63 days; 39,682 observations.
Return shocks are drawn from a Normal (0,1) and Student-t distributions with 7 and 4 degrees of freedom.
True Volatility: σ = 0.20; True mean: μ = 0; Sample mean is not estimated.

	Baseline	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8
θ	0	0	0	0.05	0.05	0.05	0.20	0.20	0.20
κ	0	0	0	0.40	0.40	0.40	2.5	2.5	2.5
Shocks	N(0,1)	t(7)	t(4)	N(0,1)	t(7)	t(4)	N(0,1)	t(7)	t(4)
RMSE	0.0178	0.0243	0.0389	0.0192	0.0254	0.0396	0.0318	0.0360	0.0473
Prob= 0.05	1.01	0.98	0.91	1.02	0.98	0.92	1.05	0.99	0.93
0.01	1.17	1.61	1.83	1.15	1.61	1.80	1.32	1.71	1.88
0.002	1.46	3.43	4.64	1.56	3.54	4.74	2.07	4.08	5.13
0.001	1.76	5.39	7.68	1.70	5.56	7.63	2.34	5.93	8.55
0.0005	2.03	8.49	12.75	2.05	8.61	12.94	3.12	9.07	14.02
0.0002	2.21	14.72	26.84	2.39	14.52	26.43	4.05	15.78	27.20
0.0001	2.37	23.21	45.79	2.40	22.67	46.21	4.94	25.52	48.02

Table 7: Probability Ratios for Exponentially Weighted Moving Average Volatility

Simulation: Sequential returns for 2,500,000 days; Estimation sample: 63 days; 39,682 observations.
Volatility is calculated using an exponentially weighted moving average with decay factors D = 0.97, 0.94, 0.90.
True Volatility: $\sigma = 0.20$; True mean: $\mu = 0$; Sample mean is not estimated.

	Run 1 Baseline	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10	Run 11	Run 12
θ	0	0	0	0	0.05	0.05	0.05	0.05	0.25	0.25	0.25	0.25
κ	0	0	0	0	1	1	1	1	1	1	1	1
Decay	1.0	0.97	0.94	0.90	1.0	0.97	0.94	0.90	1.0	0.97	0.94	0.90
RMSE	0.0178	0.0202	0.0251	0.0319	0.0191	0.0210	0.0256	0.0322	0.0390	0.0345	0.0340	0.0372
Prob= 0.05	1.01	1.03	1.06	1.14	1.01	1.02	1.06	1.14	1.11	1.06	1.09	1.14
0.01	1.11	1.14	1.22	1.32	1.10	1.13	1.21	1.33	1.31	1.23	1.25	1.37
0.002	1.17	1.20	1.33	1.52	1.15	1.21	1.32	1.54	1.56	1.45	1.51	1.64
0.001	1.23	1.25	1.46	1.82	1.26	1.27	1.45	1.81	2.11	1.83	1.76	2.05
0.0005	1.46	1.64	1.75	2.18	1.57	1.57	1.76	2.12	3.03	2.52	2.42	2.56
0.0002	1.76	1.81	2.04	2.65	1.73	1.81	2.04	2.38	3.92	3.19	2.74	2.97
0.0001	2.03	2.16	2.18	2.81	2.02	2.13	2.13	2.75	5.45	4.08	3.83	4.25

Table 8: 63-Day Rolling Sample Probability Ratios with Different θ and κ Values

Simulation: Sequential simulated returns for 250,000 days; Estimation sample: 63-day rolling sample.
True Volatility: $\sigma = 0.20$; True mean: $\mu = 0$; Sample mean is not estimated.
"No overlap" lines duplicate results from Table 5, for comparison.

		Baseline	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9
θ (Vol'y of variance)		0	0.05	0.10	0.05	0.10	0.10	0.20	0.10	0.20	0.40
κ (mean reversion)		0	0.2	0.2	0.4	0.4	1.0	1.0	2.5	2.5	2.5
RMSE	No overlap	0.0178	0.0193	0.0230	0.0192	0.0229	0.0227	0.0331	0.0222	0.0318	0.0545
	overlap	0.0177	0.0188	0.0233	0.0192	0.0223	0.0224	0.0328	0.0219	0.0313	0.0540
Prob= 0.05	No overlap	1.01	1.02	1.03	1.02	1.03	1.03	1.07	1.02	1.05	1.15
	overlap	1.03	1.30	1.80	1.12	1.58	1.15	1.64	1.09	1.30	1.88
0.01	No overlap	1.17	1.14	1.26	1.15	1.21	1.19	1.43	1.19	1.32	1.86
	overlap	1.16	2.03	4.25	1.49	3.17	1.58	3.52	1.40	2.00	4.52
0.002	No overlap	1.46	1.55	1.82	1.56	1.72	1.69	2.43	1.66	2.07	3.76
	overlap	1.41	3.62	12.09	2.28	7.56	2.56	8.98	1.84	3.86	12.95
0.001	No overlap	1.76	1.70	2.12	1.70	2.03	1.88	2.95	1.84	2.34	5.56
	overlap	1.50	4.74	19.99	2.66	11.41	3.20	14.01	2.26	5.08	20.96
0.0005	No overlap	2.03	2.02	2.69	2.05	2.36	2.18	3.59	2.17	3.12	7.57
	overlap	1.54	6.57	33.17	3.22	17.84	4.12	22.51	2.59	7.13	34.46
0.0002	No overlap	2.21	2.39	3.50	2.39	3.27	3.06	5.36	2.70	4.05	12.65
	overlap	1.90	9.74	66.48	4.48	32.19	6.06	42.61	3.04	11.54	68.40
0.0001	No overlap	2.37	2.45	3.35	2.40	2.59	2.95	7.07	2.70	4.94	19.06
	overlap	1.32	10.64	101.87	3.68	44.01	6.04	60.02	2.60	13.44	101.95

Table 9: Realized Tail Events for the Standard and Poor's 500 Index

Simulation: Historical sample of returns on the S&P 500 Index July 2, 1962 - August 30, 2002 (10,113 days).
Estimation sample: 63 day rolling sample; Sample mean is not estimated.

Sample		Tail Probability								
		.05	.02	.01	.005	.002	.001	.0005	.0002	.0001
21-day	Events predicted	505	202	101	50	20	10	5	2	1
	Actual events	587	304	195	128	85	65	50	42	30
	Probability ratio	1.16	1.51	1.93	2.54	4.21	6.44	9.91	20.81	29.73
63-day	Events predicted	502	201	100	50	20	10	5	2	1
	Actual events	548	283	184	130	75	59	43	30	22
	Probability ratio	1.09	1.41	1.83	2.59	3.73	5.87	8.56	14.93	21.89
250-day	Events predicted	493	197	99	49	20	10	5	2	1
	Actual events	507	256	171	114	77	56	43	29	22
	Probability ratio	1.03	1.30	1.73	2.31	3.90	5.68	8.72	14.70	22.31

Table 10: 63-Day Rolling Sample Probability Ratios for Multi-Day Events

Simulation: Sequential simulated returns for 250,000 days; Historical sample 1962 - 2002 for S&P 500 Index (10,113 days); Estimation sample: 63 day rolling sample.
True Volatility: $\sigma = 0.20$; True mean: $\mu = 0$; Sample mean is not estimated.
The symbol = in a cell indicates that theoretically expected number of events if they were independent was 0 and none occurred in the sample.

Panel A: Two Events in Two Days

	Probability under Independence	Ratio of Realized Probability to Probability under Independence					
		Baseline	Run 1	Run 2	Run 3	Run 4	S&P500
θ (Vol'y of variance)	-	0	0.10	0.10	0.20	0.40	-
κ (mean reversion)	-	0	0.2	1.0	1.0	2.50	-
Prob= 0.10	0.01	1.04	2.61	1.32	2.28	2.79	1.65
0.05	0.0025	1.12	5.66	1.82	4.68	6.03	3.03
0.02	0.00040	1.39	19.54	3.10	14.10	20.11	6.72
0.01	0.00010	1.68	54.65	4.56	35.21	55.13	16.92
0.005	0.000025	2.56	160.36	8.48	90.90	152.20	47.77
0.002	0.000004	2.00	690.18	22.01	360.09	631.16	74.64

Panel B: Three Events in Three Days

	Probability under Independence	Ratio of Realized Probability to Probability under Independence					
		Baseline	Run 1	Run 2	Run 3	Run 4	S&P500
θ (Vol'y of variance)	-	0	0.10	0.10	0.20	0.40	-
κ (mean reversion)	-	0	0.2	1.0	1.0	2.50	-
Prob= 0.10	0.001	1.08	5.75	1.81	4.67	6.44	3.19
0.05	0.0001	1.25	22.89	3.52	17.09	24.39	9.56
0.02	0.000008	3.00	184.55	11.00	120.03	173.55	24.88
0.01	0.000001	12.00	980.25	36.01	576.15	908.24	=
0.005	1.25E-7	32.01	5665.5	160.0	2944.8	4641.2	=
0.002	8.0E-9	=	59515.5	1000.3	26506.9	42010.9	=

Table 10: 63-Day Rolling Sample Probability Ratios for Multi-Day Events, continued

Panel C: Three Events in Five Days

	Probability under Independence	Ratio of Realized Probability to Probability under Independence					
		Baseline	Run 1	Run 2	Run 3	Run 4	S&P500
θ (Vol'y of variance)	-	0	0.10	0.10	0.20	0.40	-
κ (mean reversion)	-	0	0.2	1.0	1.0	2.50	-
Prob= 0.10	0.009	1.06	4.61	1.68	3.79	4.93	2.34
0.05	0.0012	1.22	17.24	3.26	12.93	17.83	7.13
0.02	0.00008	1.75	132.53	9.23	86.86	130.93	20.52
0.01	0.000010	6.90	716.1	27.21	388.7	655.2	80.85
0.005	1.24E-06	16.12	3873.2	77.4	1947.9	3354.0	481.5
0.002	7.98E-08	=	41886.8	351.1	18159.3	34111.4	2496.3

Panel D: Three Events in Ten Days

	Probability under Independence	Ratio of Realized Probability to Probability under Independence					
		Baseline	Run 1	Run 2	Run 3	Run 4	S&P500
θ (Vol'y of variance)	-	0	0.10	0.10	0.20	0.40	-
κ (mean reversion)	-	0	0.2	1.0	1.0	2.50	-
Prob= 0.10	0.070	1.03	2.70	1.39	2.39	2.88	1.39
0.05	0.0115	1.10	8.97	2.50	7.33	9.67	3.58
0.02	0.00086	1.37	64.72	7.34	46.02	69.30	14.07
0.01	0.000114	2.60	341.4	19.40	214.7	339.6	52.49
0.005	1.46E-05	5.20	1895.0	62.7	1080.9	1840.2	231.8
0.002	9.50E-07	=	20082.2	290.6	10171.7	18384.9	838.8

Table 11: Stochastic Volatility Models with Jumps: Specifications and Parameter Values

Model: SV = stochastic volatility, no jumps
SVJ = stochastic volatility, jumps in returns equation
SVIJ = stochastic volatility, independent jumps in returns and variance
SVCJ = stochastic volatility, simultaneous jumps in returns and variance

Source:

ABL = Andersen, Benzoni, and Lund [2002]. Sample: daily S&P 500 index returns, 1953-1996; lognormal jumps in returns; variance follows equation (15); parameter estimation by Efficient Method of Moments.

BCC = Bakshi, Cao, and Chen [1997]. Sample: S&P 500 index options, June 1989 - May 1991; μ set to 5% less the expected value of the jump component; lognormal jumps in returns; variance follows equation (14); parameters implied from option prices by minimizing squared price discrepancies.

EJP = Eraker, Johannes, Polson [2003]. Sample: daily change in log of S&P 500 index, 1980-1999; normal jumps in returns; variance follows equation (14); parameter estimation by Markov Chain Monte Carlo.

All parameters are expressed as annualized decimal values, converted as necessary from the values reported in the original articles.

	SV	**SV**	**SVJ**	**SVJ**	**SVJ**	**SVIJ**	**SVCJ**
	EJP	**BCC**	**EJP**	**ABL**	**BCC**	**EJP**	**EJP**
μ	0.112	0.050	0.125	0.077	0.085	0.128	0.141
κ	5.82	1.15	3.23	3.65	2.03	6.30	6.55
$\sqrt{\overline{V}}$	0.151	0.187	0.143	0.105	0.140	0.119	0.116
θ	0.361	0.390	0.240	0.291	0.380	0.226	0.199
ρ	-0.397	-0.640	-0.467	-0.613	-0.570	-0.504	-0.484
λ	-	-	1.51	3.45	0.59	1.16	1.66
$E[J_t \mid dq=1]$	-	-	??	0	??	??	??
σ_J	-	-	0.0407	0.0700	0.0700	0.0299	0.0289
λ_V	-	-	-	-	-	1.386	1.66 (=λ)
ν	-	-	-	-	-	.000180	.000148

Table 12. Rolling Sample Probability Ratios for Jump Diffusion Models, Using 63-day, 250-day, and EWMA Windows

Simulation: Sequential simulated returns for 250,000 days from the models and parameter values specified in Table 11; Estimation samples: Rolling sample of 63 days (Panel A), 250 days (Panel B), 74 days, with EWMA with decay factor 0.94 (Panel C). Sample mean is not estimated.

Panel A: 63-Day Sample

	SV	**SV**	**SVJ**	**SVJ**	**SVJ**	**SVIJ**	**SVCJ**
	EJP	**BCC**	**EJP**	**ABL**	**BCC**	**EJP**	**EJP**
RMSE	4.626	5.379	4.309	8.931	6.240	4.932	4.253
α = 0.05	1.03	1.16	0.92	0.72	1.04	0.95	0.92
0.01	1.65	2.40	1.51	1.31	1.79	1.66	1.53
0.005	2.14	3.46	2.07	1.94	2.43	2.32	2.15
0.002	3.14	5.90	3.44	3.63	3.91	3.94	3.69
0.001	4.34	8.99	5.41	6.12	5.78	6.21	5.86
0.0005	6.01	13.85	8.86	10.76	8.99	10.12	9.74
0.0002	9.66	24.90	18.14	23.86	16.98	20.37	19.93
0.0001	13.99	39.20	32.00	44.51	28.32	35.56	35.41

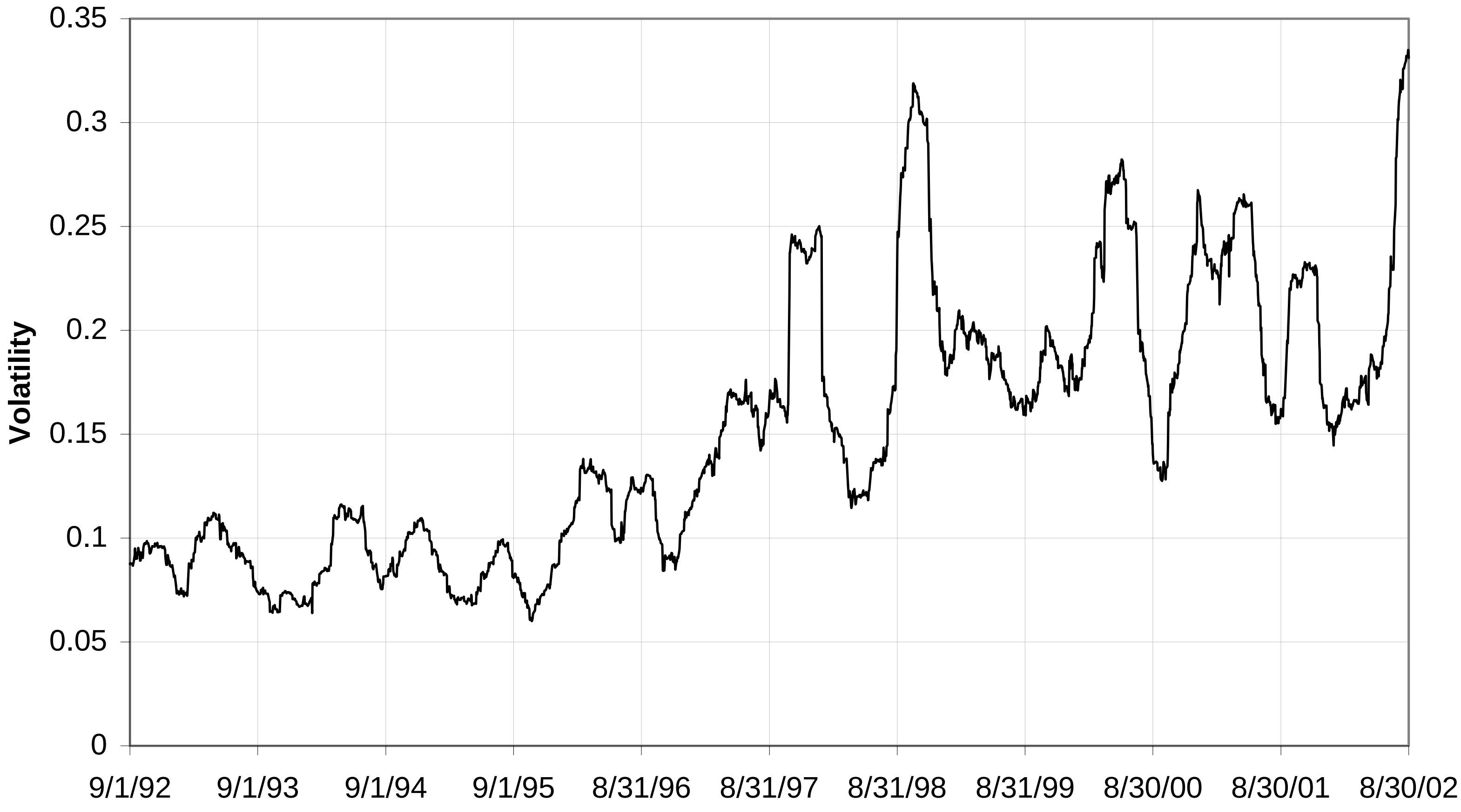
Panel B: 250-Day Sample

	SV	**SV**	**SVJ**	**SVJ**	**SVJ**	**SVIJ**	**SVCJ**
	EJP	**BCC**	**EJP**	**ABL**	**BCC**	**EJP**	**EJP**
RMSE	5.089	8.415	4.76	8.095	7.960	5.991	5.539
α = 0.05	0.99	1.16	0.85	0.48	1.00	0.89	0.89
0.01	1.62	2.40	1.45	0.94	1.84	1.66	1.63
0.005	2.12	3.46	2.02	1.47	2.58	2.40	2.35
0.002	3.13	5.90	3.41	2.93	4.25	4.22	4.16
0.001	4.33	8.99	5.38	5.19	6.46	6.69	6.65
0.0005	6.14	13.85	8.87	9.41	10.09	11.00	11.04
0.0002	9.77	24.90	18.16	21.32	18.90	22.08	22.35
0.0001	14.04	39.20	31.97	40.24	31.34	38.48	39.06

Panel C: Exponentially Weighted Moving Average, 74-Day Sample

	SV	SV	SVJ	SVJ	SVJ	SVIJ	SVCJ
	EJP	BCC	EJP	ABL	BCC	EJP	EJP
RMSE	3.381	4.179	2.989	9.382	5.548	4.102	3.340
α = 0.05	1.00	1.07	0.91	0.75	1.01	0.94	0.92
0.01	1.47	1.75	1.37	1.29	1.57	1.50	1.42
0.005	1.84	2.29	1.85	1.90	2.05	2.04	1.93
0.002	2.58	3.42	3.03	3.50	3.17	3.39	3.24
0.001	3.38	4.78	4.73	5.95	4.60	5.33	5.15
0.0005	4.59	6.86	7.76	10.52	7.07	8.76	8.59
0.0002	7.05	11.43	16.02	23.40	13.23	17.75	17.80
0.0001	9.67	17.10	28.48	43.98	22.26	31.20	31.98

Figure 1
S&P 500 63-Day Volatility, 9/1/1992 - 8/30/2002



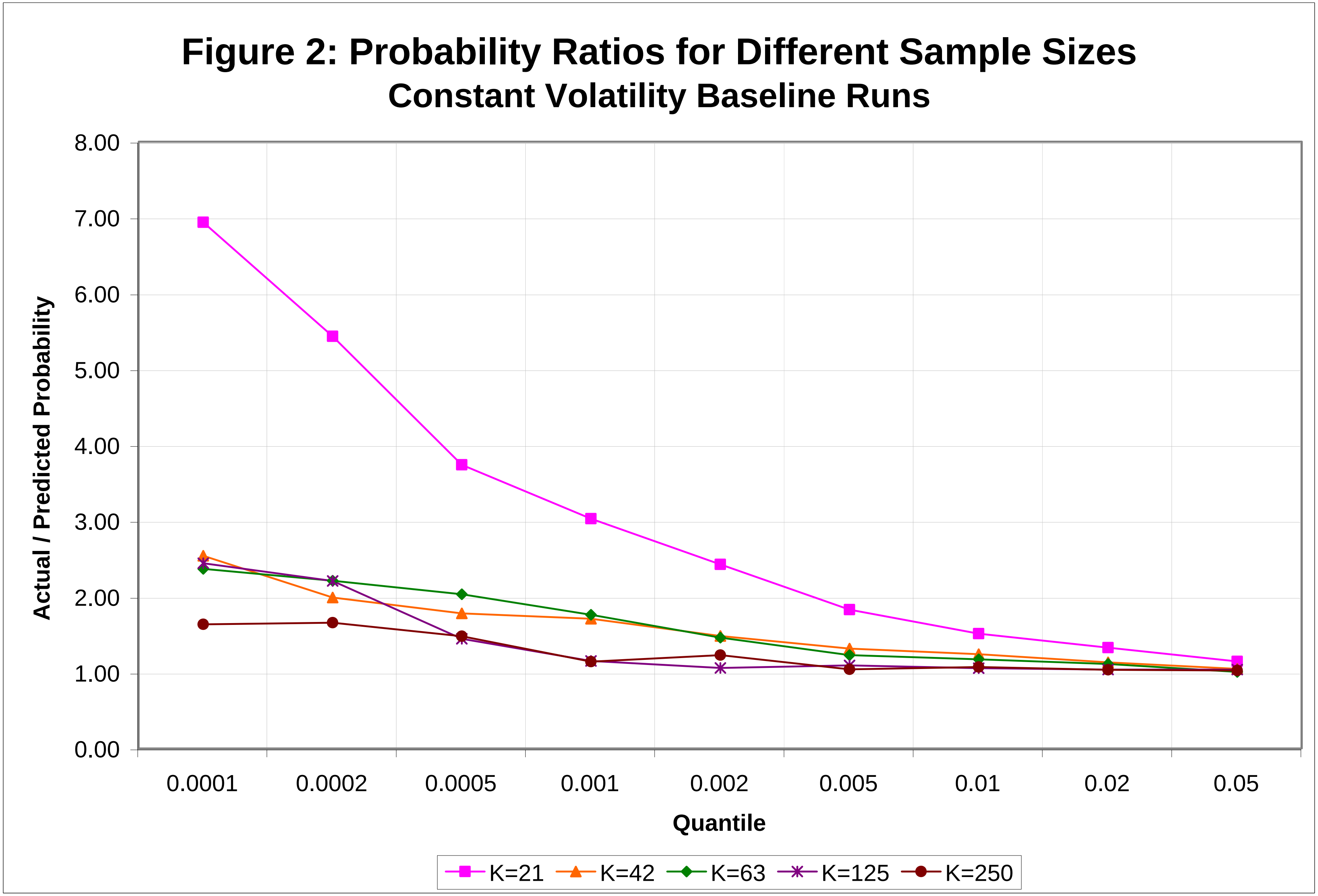


Figure 2: Probability Ratios for Different Sample Sizes
Constant Volatility Baseline Runs
Actual / Predicted Probability
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0.0001
0.0002
0.0005
0.001
0.002
0.005
0.01
0.02
0.05
Quantile
K=21
K=42
K=63
K=125
K=250

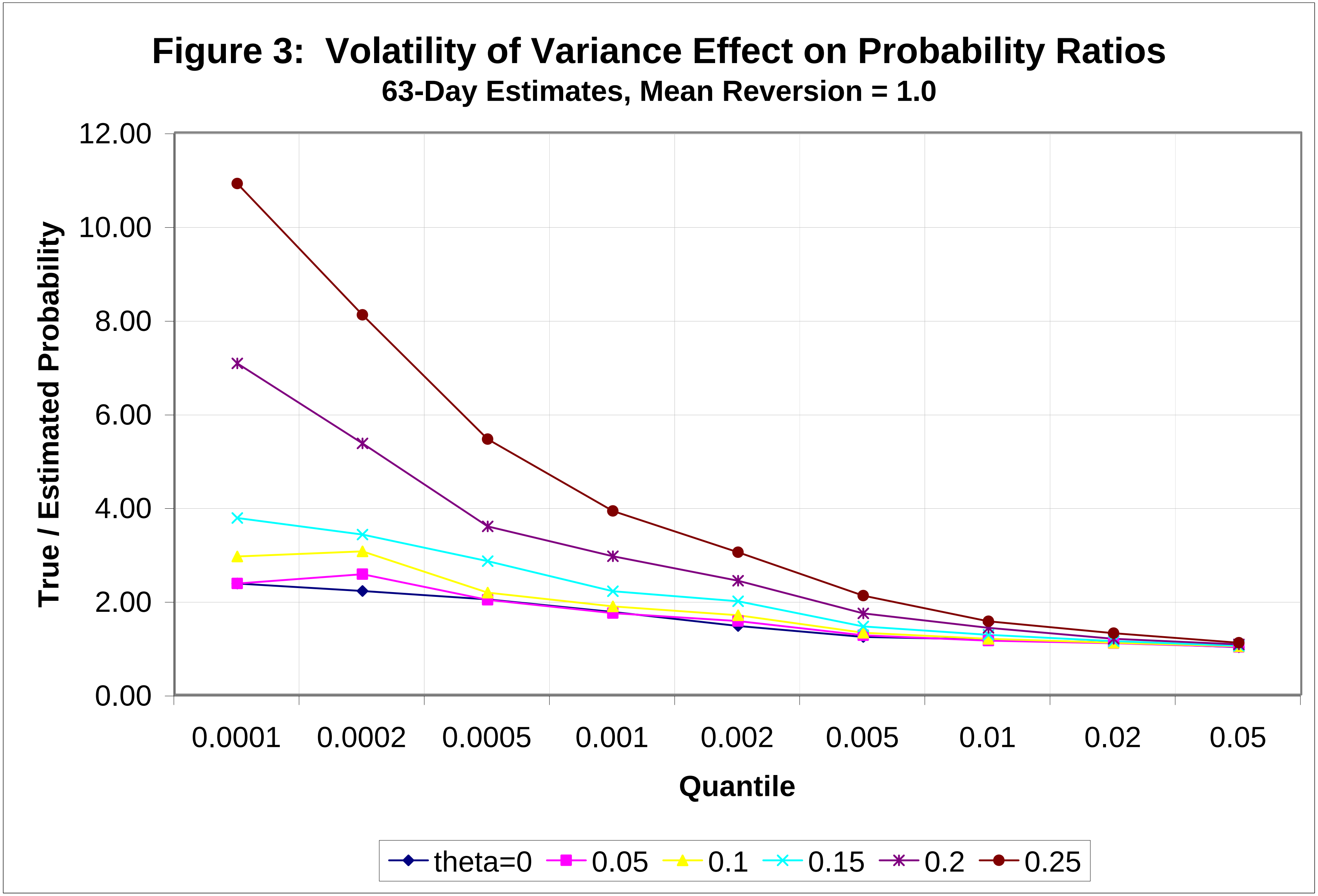


Figure 3: Volatility of Variance Effect on Probability Ratios
63-Day Estimates, Mean Reversion = 1.0
True / Estimated Probability
12.00
10.00
8.00
6.00
4.00
2.00
0.00
0.0001
0.0002
0.0005
0.001
0.002
0.005
0.01
0.02
0.05
Quantile
theta=0
0.05
0.1
0.15
0.2
0.25

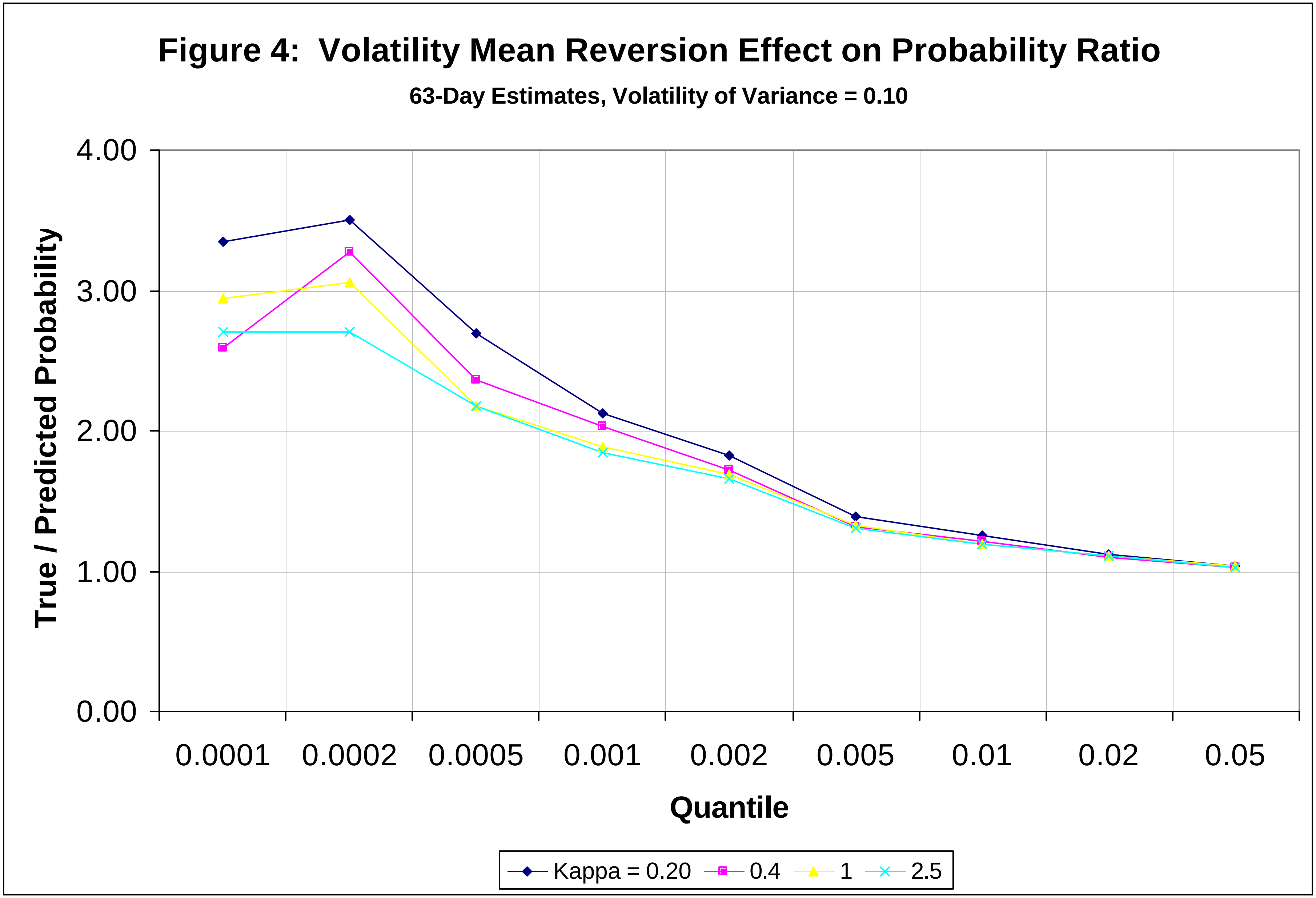


Figure 4: Volatility Mean Reversion Effect on Probability Ratio
63-Day Estimates, Volatility of Variance = 0.10
4.00
3.00
2.00
1.00
0.00
True / Predicted Probability
0.0001
0.0002
0.0005
0.001
0.002
0.005
0.01
0.02
0.05
Quantile
Kappa = 0.20
0.4
1
2.5

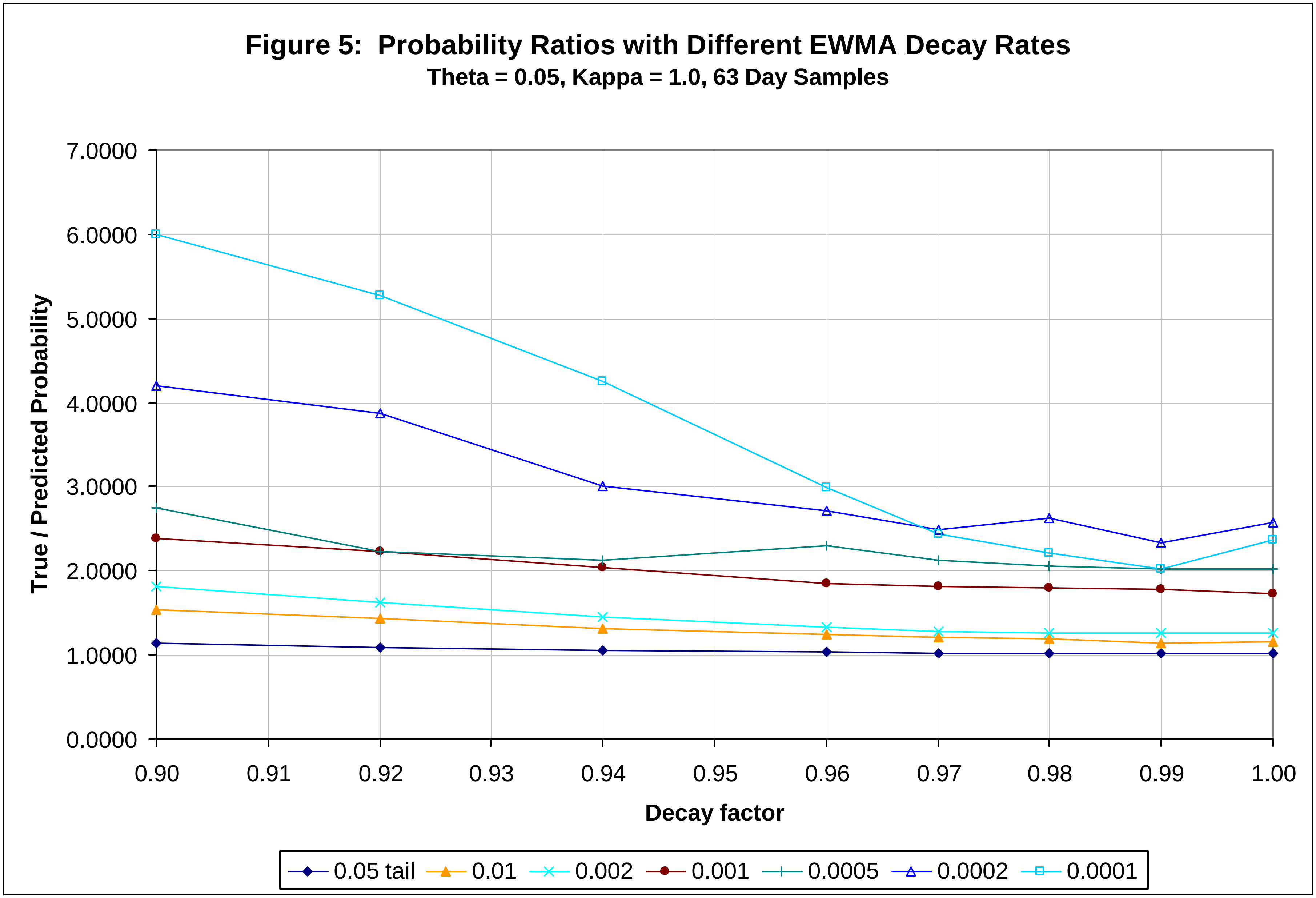


Figure 5: Probability Ratios with Different EWMA Decay Rates
Theta = 0.05, Kappa = 1.0, 63 Day Samples
True / Predicted Probability
7.0000
6.0000
5.0000
4.0000
3.0000
2.0000
1.0000
0.0000
0.90
0.91
0.92
0.93
0.94
0.95
0.96
0.97
0.98
0.99
1.00
Decay factor
0.05 tail
0.01
0.002
0.001
0.0005
0.0002
0.0001

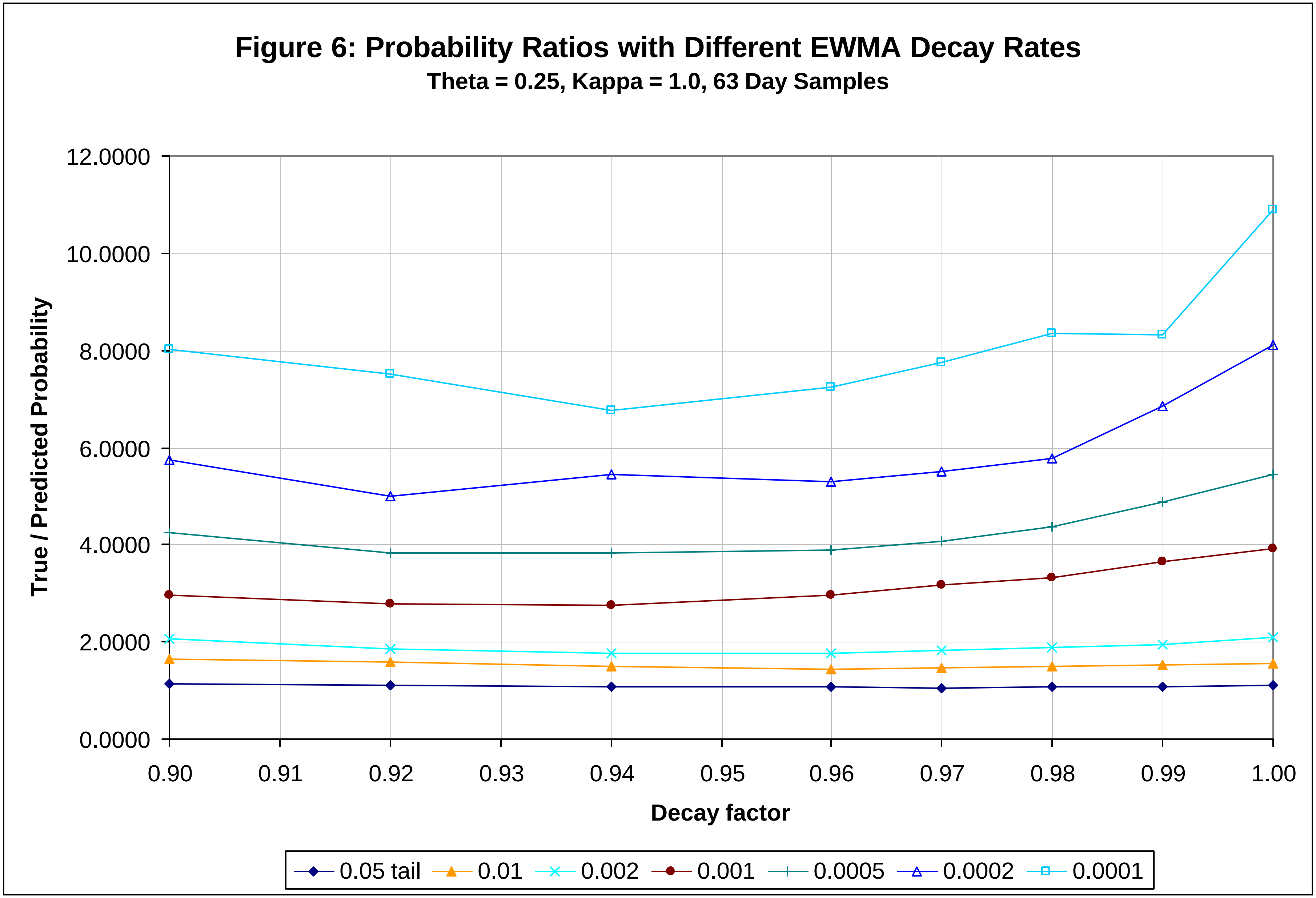


Figure 6: Probability Ratios with Different EWMA Decay Rates
Theta = 0.25, Kappa = 1.0, 63 Day Samples
True / Predicted Probability
12.0000
10.0000
8.0000
6.0000
4.0000
2.0000
0.0000
0.90
0.91
0.92
0.93
0.94
0.95
0.96
0.97
0.98
0.99
1.00
Decay factor
0.05 tail
0.01
0.002
0.001
0.0005
0.0002
0.0001