# Risk Management Series

January 1998

# Estimating Covariance Matrices

**Robert Litterman**
**Kurt Winkelmann**

**Acknowledgments**

Understanding the nature of portfolio risk plays an important role in many activities at Goldman Sachs. We have been fortunate to be able to draw on the expertise of many of our colleagues while writing this paper. We are especially grateful to Armen Avenessians, Mark Griffin, Joanne Hill, Robert Litzenberger, Bernard Minsky, Paul Mussche, Victor Ng, Scott Pinkus, Gianluca Squassi, and Mark Zurack for insightful comments on earlier drafts of this paper.

*Robert Litterman*
*212-902-1677*

*Kurt Winkelmann*
*44-171-774-5545*

Robert Litterman is a Managing Director in Firmwide Risk at Goldman, Sachs & Co. and Kurt Winkelmann is a Vice President in Fixed Income Research at Goldman Sachs International.

**Editor: Ronald A. Krieger**

# Estimating Covariance Matrices

## Contents

## Executive Summary

Many practical problems in modern finance require an understanding of the volatility and correlations of asset returns. Examples of such everyday problems include managing the risk of a multi-currency portfolio, optimal asset allocation, and derivatives pricing.

Unfortunately for practitioners, volatility and correlation (or the covariance matrix) cannot be directly observed, and must be estimated from data on daily returns. Thus, how the covariance matrix is estimated can have important implications for the practice of modern finance. We provide examples to illustrate how many practical decisions are influenced by the covariance matrix choice.

Since the covariance matrix must be estimated, the practitioner faces an interesting trade-off between using estimation methods that most closely resemble real-world phenomena and using estimation methods that are not computationally burdensome. This trade-off is intensified for larger covariance matrices, which stand at the core of many risk management and asset allocation problems.

In this paper, we discuss the covariance matrix estimation methods used at Goldman Sachs for large-scale risk management and asset allocation problems. We describe how the methods used at Goldman Sachs account for several regularities commonly observed in financial data. In this context, we show how these methods offer an improvement on the assumption that returns are generated by a stable Normal distribution. We also contrast the methods used at Goldman Sachs with other commonly used approaches.

Covariance matrix estimation can easily be regarded as a topic that will never be closed. In that spirit, we also offer several suggestions for future research, on the premise that a deeper understanding of the covariance matrix can lead to improved practical decision making.

# Estimating Covariance Matrices

## I. Introduction

Investors take on risk in order to generate higher expected returns. This trade-off implies that an investor must balance the return contribution of each investment against its contribution to portfolio risk. Central to achieving this balance is some measure of the correlation of each investment's returns with those of the portfolio. This problem, along with many of the other practical problems in modern finance, requires measures of volatility and correlation. Other examples of such problems include estimating the risk of a portfolio of positions, determining optimal hedges, pricing derivatives, identifying optimal weights for a trade, and finding the optimal asset allocation for a portfolio.

Although the daily practice of finance uses volatility and correlation (or, more precisely, the covariance matrix) of asset returns as inputs, these quantities cannot be directly observed and must instead be estimated, generally from historical observations on financial assets returns.[1] Since different estimation procedures can give rise to different covariance matrices, the choice of estimation method becomes critical. Many financial market participants, recognizing the uncertainty in covariance estimation, have given up on the hope of making informed decisions based on this type of information. But we believe there is no choice. Every decision in finance, as in life, requires that people make choices under conditions of uncertainty. The only issue is whether they make the decisions with more information or with less. For many financial decisions, the relevant information is an estimate of covariances over a future horizon.

To illustrate the role of the covariance matrix in a bit more detail, suppose that we are managing a global government bond trading book. For simplicity, we will assume that the book will take long and short positions in 10-year bonds in the following 10 markets: Canada, France, Germany, Italy, Japan, the Netherlands, Spain, Sweden, the United Kingdom, and the United States. Exhibit 1 presents a summary of these positions.

[1] We address the role of implied volatilities in more detail in Section IV.

**Exhibit 1**

**Government Bond Trading Positions**

*(10-Year Bonds; US$, millions)*

| Market | Position |
|---|---|
| Canada | 100 |
| France | 150 |
| Germany | -500 |
| Italy | 100 |
| Japan | 60 |
| Netherlands | -100 |
| Spain | 100 |
| Sweden | 150 |
| U.K. | 175 |
| U.S. | -175 |

From a risk management perspective, there are several important questions that we would like to address about this portfolio:

- First, we would like to know the overall risk of the portfolio (a question that in principle can be answered with a time series of portfolio returns).
- Second, we would like to know each position's contribution to the overall risk.
- Third, we would like to identify hedges that make the portfolio neutral with respect to market moves.
- Finally, we would like to identify the portfolio's "implied views."[2] Each of these issues depends on the volatility and correlation of asset returns for all assets in the portfolio.

Exhibit 2 contrasts the risk and risk decomposition of this portfolio using covariance matrices estimated with two alternative methods. The figures in the middle column are derived from the covariance matrix used in the Goldman Sachs risk management system as of December

[2] For further discussion of these topics, please see Litterman (1996) and Litterman and Winkelmann (1996). [Note: Full reference information on sources mentioned in this report appears in the Bibliography, page 43.]

## Exhibit 2

## VaR and Risk Decomposition

| Market | GS Risk System Risk (%) | Monthly Data Risk (%) |
|---|---|---|
| Canada | -2.49 | 5.35 |
| France | 6.25 | 7.01 |
| Germany | 20.37 | 6.07 |
| Italy | 15.19 | 18.65 |
| Japan | 2.64 | 1.07 |
| Netherlands | -0.73 | -2.32 |
| Spain | 13.67 | 18.78 |
| Sweden | 19.89 | 25.40 |
| U.K. | 6.43 | 15.11 |
| U.S. | 18.77 | 4.87 |
| VaR | $1.64 million | $2.41 million |

31, 1996.[3] By contrast, the figures in the right-hand column are derived from nine years of monthly returns.

Exhibit 2 shows that measurement of risk and identification of the primary sources of risk in the portfolio depend on which covariance matrix is used. For example, the Value-at-Risk (VaR), measuring the amount of capital that would be expected to be lost once in 100 two-week intervals, increases from $1.64 million when we use the Goldman Sachs risk system covariance matrix to $2.41 million when we use monthly data. Furthermore, the primary source of risk changes from the German 10-year bond position to the Swedish 10-year bond position. (Appendix A, page 41, presents both covariance matrices.)

Covariance matrix estimation is also important for the asset allocation problem. Conventional wisdom holds that good expected return forecasts will overpower the effects of any errors in risk measurement (i.e., the covariance matrix). However, the following example provides an interest-

[3] Goldman Sachs takes many different approaches to managing its risk. We start by using a covariance matrix estimated with decayed daily data to find VaR. Our covariance matrix estimation employs the procedures discussed in this paper. In particular, we use a decay rate that is consistent with a two-week rebalancing horizon. In addition to a covariance matrix, we also use historical simulation and Monte Carlo simulation for risk management purposes. No one approach is "best" or "correct": each provides a unique set of insights into risk.

ing illustration of the performance effects of more or less accurate covariance matrices.

This example simulates the performance from January 1992 through March 1997 of two fund managers who hold identical views on expected returns, but who use different covariance matrices. Each manager selects an optimal portfolio against a currency-hedged, capitalization-weighted benchmark consisting of U.S., German, and Japanese bonds. We have chosen deviations from the benchmark weights to give a predicted tracking error of 100 basis points, and we assume that each manager rebalances quarterly. Thus, there are 21 separate rebalancings.

Of course, deviations from the benchmark allocations are driven by views on expected returns and the covariance matrix. For simplicity, suppose that each manager's quarterly expected return projections are centered at that quarter's actual returns.[4] The manager-specific covariance matrices at each quarterly rebalancing are created as follows: for manager A, the covariance matrix includes only the data for the returns in the upcoming quarter. For manager B, the covariance matrix uses equally weighted weekly data from January 1991, but rather than including all of the upcoming quarter's returns, the covariance matrix for each date uses only data that could be known at the time of rebalancing. How well does each manager perform?

Portfolio managers are often judged by their "information ratio," i.e., the ratio of their excess return to their actual tracking error. We can compute information ratios for each manager by first finding the average performance relative to the index, and then finding the standard deviation of the relative performance. We will call this second quantity the actual tracking error. Taking the ratio of the actual average

[4] The quarterly excess returns (in percent) for the Goldman Sachs German Government Bond index from January 1, 1992 through March 31, 1997, were: .-0.20, -0.60, 1.82, 1.67, 2.25, .21, 1.75, 2.24, -3.04, -2.16, -1.62, 0.27, 3.26, 1.20, 2.62, 3.31, -0.81, 0.20, 2.55, 1.66, -0.05. Similarly, the quarterly excess returns for the Goldman Sachs Japan Government Bond Index over the same period were: 1.16, 1.38, 3.89, 1.65, 1.54, 0.72, 4.60, 4.29, -4.17, -0.23, -0.38, 0.77, 5.68, 4.85, 1.26, -0.10, 0.10, 0.60, 2.65, 1.92, 2.10. Finally, the corresponding excess returns to the Goldman Sachs U.S. Government Bond Index were: -2.83, 2.97, 4.20, -0.87, 3.82, 2.15, 2.52, -1.15, -3.92, -2.04, -0.77, -1.01, 3.32, 4.79, 0.34, 3.23, -3.70, -0.96, 0.21, 1.51, -2.29.

performance to the actual tracking error produces each manager's information ratio.

The performance differences between the two managers on this dimension are quite striking. On average, each manager outperforms the benchmark by roughly 59 basis points (bp) per quarter. However, while manager A has an actual quarterly tracking error of 60 bp (or $60 \times \sqrt{4} = 120$ bp on an annualized basis), manager B has a quarterly tracking error of 69 bp (or 138 bp annualized). The information ratio of manager A is 0.98, 14% better than manager B's information ratio, which is 0.86. Since the managers are identical in all respects except the covariance matrix, it must be the case that the performance differences are a consequence of the choice of covariance matrix. Contrary to the presumption of conventional wisdom, covariance matrix estimation is as relevant for fund management as expected return forecasting. Indeed, covariance matrix estimation can be easily viewed as a forecasting problem in its own right: Rather than forecasting a portfolio's central tendency (expected return), a covariance matrix is essential for forecasting the range of possible performance outcomes.

This paper discusses the covariance matrix estimation methods used at Goldman Sachs for large-scale risk management and asset allocation problems. We try to produce covariance matrices that are consistent with the regularities that are observed in financial data and that form the basis for forecasting future volatilities and correlations. Also, we rely on algorithms that are easy to compute and can be easily understood and compared with alternative estimation methods.

We do not believe there is one optimally estimated covariance matrix. Rather, we use approaches designed to balance trade-offs along several dimensions and choose parameters that make sense for the task at hand. One important trade-off arises from the desire to track time-varying volatilities, which must be balanced against the imprecision that results from using only recent data. This balance is very different when the investment horizon is short, for example a few weeks, versus when it is longer, such as a quarter or a year. Another trade-off arises from the desire to extract as much information from the data as possible, which argues toward measuring returns over short intervals. This desire must be balanced against the reality that the structure of volatility and correlation is not stable and may be contaminated by mean-reverting noise over very short intervals, such as intraday or even daily returns.

The paper is organized as follows: The next section (Section II) reviews regularities that are commonly found in financial time series. Section III presents the covariance matrix estimation framework used at Goldman Sachs, while Section IV contrasts our methods with other frequently used techniques. We discuss the relationship between covariance matrices and structural models in Section V, and offer some concluding comments in Section VI.

## II. Empirical Regularities in Financial Data

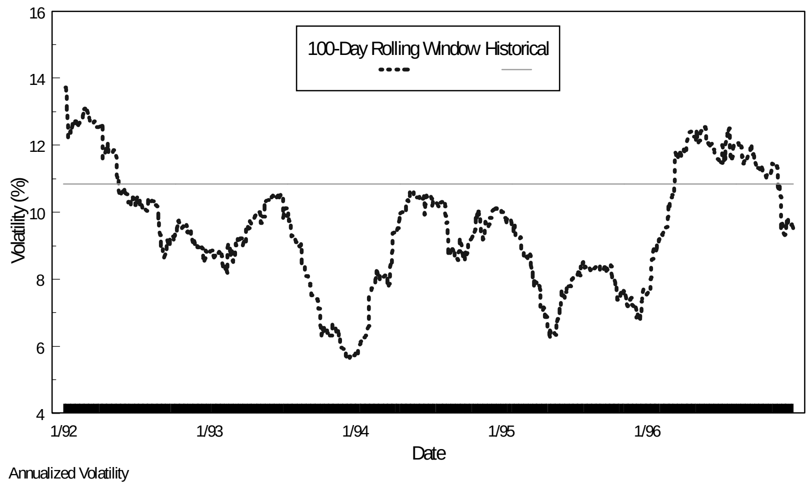
Unfortunately, a common yardstick for discussing regularities in financial time series is the multivariate Normal distribution. If asset returns followed a constant multivariate Normal distribution 24 hours a day, the following regularities should be observed: First, the volatility of each asset's daily return would be roughly constant over time. Second, the correlation of asset returns would also be approximately constant across time. Finally, asset returns would not be "fat-tailed." In other words, approximately 66% of the observed daily returns would be within one standard deviation of the average return, and roughly 95% of the observed daily returns would be within two standard deviations of the average return. As we will see below, asset returns, for the most part, do not conform at all well to this yardstick. Volatilities and correlations are time-varying, and the distributions of asset returns are fat-tailed.

### Volatility Is Time-Varying

Exhibit 3 shows two estimates of the annualized daily volatility for excess returns on the S&P 500 index from January 1, 1992, through December 31, 1996. (We measure daily excess returns as the total return to the S&P 500 less returns to cash.) The first estimate uses the entire history, and weights each day's return equally. Using this method, we estimate S&P 500 index volatility to have been 10.88%, as represented in the chart by the horizontal line at 10.88%.

**Exhibit 3**

## S&P 500 Index Volatility




The second method illustrated in Exhibit 3 follows a common industry practice of calculating volatility through a rolling window of history. With this method, volatility is calculated as the standard deviation of a fixed period of historical daily returns. Since the period (or "window") is fixed, as a new day is added to the window the most distant day in the window is deleted. In Exhibit 3, the window is fixed at 100 days. As the chart illustrates, a rolling window provides volatility estimates that are far from constant: Over the period illustrated, volatility appears to have fluctuated between 5.75% and 13.70%.

Of course, our inferences on time-varying volatility could in part be due to our small sample sizes and the nature of the rolling window. For example, suppose that the first day in the sample is an extraordinarily large observation. Suppose further that the observation that we add is a relatively small observation. Over a two-day period, the volatility estimate will drop as the small observation replaces the relatively high observation. Consequently, even if the distribution is stable and Normal, we should still see some time variation in the volatility estimates because of the small sample sizes.

A simple Monte Carlo simulation can address the issue of whether our observed time-varying volatility is due to a shift in distribution or merely to small samples. We can do this by sampling 1,000 “paths” of 1,250 observations each (i.e. 1,000 “paths” of five years of daily returns), comparing the results of our simulation with actual experience. To set up the simulation, suppose that we assume that the distribution is Normal, with a mean of zero and a volatility of 10.88%. More specifically, we can simulate a time series of volatility, estimated with a 100-day rolling window over a five-year period, by drawing 1,250 observations from this distribution and then calculating the standard deviations over overlapping intervals of 100 observations. Each successive interval adds one new observation and deletes an old observation. For the simulated volatility time series, we can calculate the average absolute deviation between the simulated volatility and the “true” volatility (which in our case we take to be 10.88%). We can then calculate the distribution of the average absolute deviation between the simulated volatilities and the “true” volatility by taking repeated samples of size 1,250. In this case, we sampled 1,000 separate “time series,” getting a mean absolute deviation of 0.61%, with a standard deviation of 0.10%. In other words, even when the distribution of returns is a stable Normal distribution, estimating volatility with a 100-day rolling window will give an average deviation from 10.88% of 0.61%. The largest average absolute deviation for a single 1,250-observation sample in our simulation was 1.08%.

Now, how does the actual history of the S&P 500 compare with the Monte Carlo simulation? We can easily compare the two by calculating the average absolute difference between the actual 100-day rolling window and the long-term historical average, and then finding the average of the resulting time series. This calculation shows that the average absolute difference is 1.87%, with a standard deviation of 1.23%. These figures do not compare especially favorably with our simulated distribution. In fact, since the largest average absolute deviation in our simulation was 1.08%, and since we ran 1,000 simulations, we can conclude that there is less than a 0.001 chance that the actual rolling window would have been observed if the true distribution were a stable Normal distribution with a 10.88% volatility. Consequently, it is hard not to reject the hypothesis that S&P 500 returns are characterized by a distribution with constant volatility.

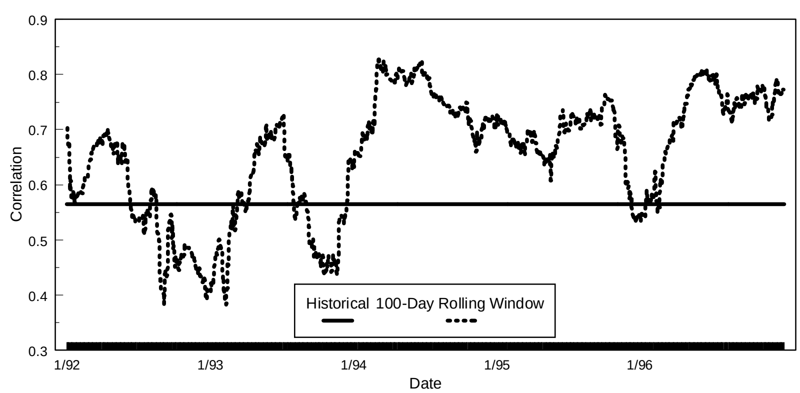
## Correlations Also Vary Over Time

The yardstick of a stable multivariate Normal distribution also predicts that correlations will not vary over time (in addition to predicting that volatility is not time-varying). Exhibit 4 explores this proposition by plotting two estimates of the correlation between French and German 10-year bond returns. We calculate correlations in the chart using daily currency-hedged excess returns on French and German 10-year bonds from January 1, 1992, through December 31, 1996.

Similar to Exhibit 3, the first estimate of correlation uses the entire history and assigns an equal weight to each daily return. With this procedure, we estimate the correlation between French and German 10-year bond returns to be 0.57; it is represented in Exhibit 4 by the horizontal line. We contrast the constant correlation estimate with correlations estimated using a 100-day rolling window. Under the second approach, the correlation between French and German 10-year bond returns ranges between 0.37 and 0.82. As with Exhibit 3, it is reasonable to ask whether the data in Exhibit 4 reflect a small sample or a distribution of returns that is time-varying. Once again, we can answer this question by looking at a Monte Carlo simulation. Rather than assume that the volatility is 10.88%, we assume that the correlation between French and German bond returns is drawn from a distribution with a stable correlation of 0.57, and then follow the same procedure to simulate the distribution of the average absolute deviations. This time, the mean of the average absolute deviations is 0.0542 and the standard deviation is 0.0090. In other words, when the true

**Exhibit 4**

**France/Germany 10-Year Bond Correlation**




Daily Correlation

correlation is 0.57, a five-year rolling window of 100-day correlations will, on average, produce a deviation of 0.0542 from the true correlation of 0.57. The largest average absolute deviation in the simulation is 0.1081.

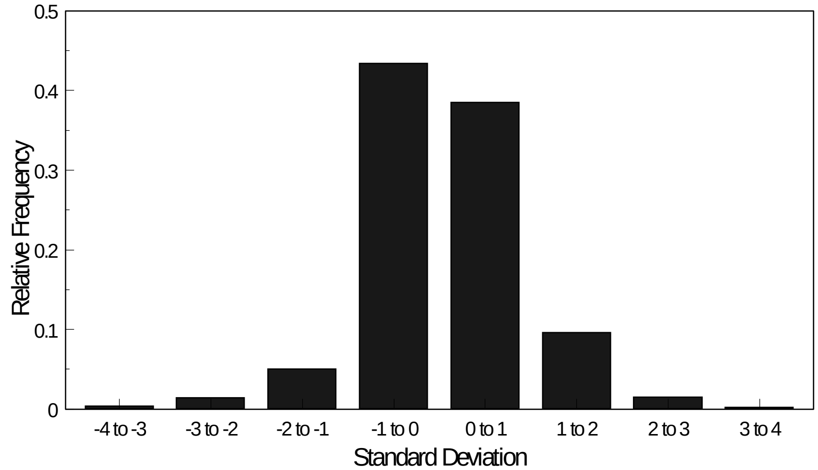
Once again, we can calculate the average absolute deviation for the actual time series of French/German bond correlations. In this case, the average absolute deviation is 0.1243, with a standard deviation of 0.0678. Again, the actual absolute average deviation is larger than the largest simulated deviation, which suggests that there is less than a 0.001 chance of observing this time series of correlations if the distribution of French and German bond returns is truly generated by a stable Normal distribution. As in the case of S&P 500 volatility, this example illustrates the fact that it is hard not to reject the hypothesis that correlations are constant over time.

**Financial Data Have 'Fat Tails'**

The final prediction of the multivariate Normal yardstick is that financial time series do not have "fat tails." To illustrate, we use daily returns; the "tails" of the distribution refer to the percentage of observed daily returns that are outside a fixed band, where the band is defined by the overall volatility (as measured by the standard deviation). When data follow a Normal distribution, 68.3% of the observations are within one standard deviation of the long-term average, and 95.4% are within two standard deviations of the long-term average. In this case, the tails of the distribution refer to the 4.6% of the observations that are more than two standard deviations away from the long-term average; they represent observations that can occur but with less frequency than other possibilities. A distribution is said to have "fat tails" (relative to a Normal distribution) when more than 4.6% of the observations are more than two standard deviations away from the long-term average.

**Exhibit 5**

**Histogram of S&P 500 Returns**




Daily Returns, 1/1/92 to 12/31/96

For a practical example of fat tails, let's look again at the volatility of the S&P 500 index. We already saw that if we use all the data from January 1, 1992, to December 31, 1996, the annualized volatility of daily excess returns works out to be 10.88%. This figure is quite easily converted to 0.68% daily excess return volatility. Consequently, if there are 250 trading days in a year and if returns are Normally distributed, we would anticipate around 11 days a year when returns exceeded 1.37% in absolute value. In addition, we would anticipate that returns would be within -0.68% and +0.68% in 171 of the days.

Exhibit 5 revisits the experience of S&P 500 index returns from January 1, 1992, through December 31, 1996. The chart summarizes the return experience with a histogram.

Exhibit 5 summarizes 1,250 trading days. Over this period, if daily returns are normally distributed, we would anticipate roughly 55 days when returns exceeded 1.37% in absolute value and 855 days when returns were between -0.68% and +0.68%. However, practical experience has been quite different, as illustrated by Exhibit 5. Rather than 55 days of returns outside two standard deviations, the practical experience has been that S&P 500 index returns have had 65 days outside the two-standard-deviation range. Similarly, over the five-year period there were 1,148 days of returns within one standard deviation, rather than the 855 predicted by a Normal distribution. Even though the chart does not provide a definitive test for fat tails, the results do suggest that S&P 500 returns are indeed fat-tailed.

Should we be surprised that financial time series have fat tails? Not really. The Normal distribution arises often in nature, in particular whenever the total uncertainty in a distribution is a sum of many independent sources. Unfortunately this independence does not exist in financial markets. Most investors pay attention to the same economic fundamentals. Large surprises are observed by all participants in the market at more or less the same time. In some cases, the large surprise may be a market move itself, as was the case in the equity market crash of October 19, 1987. As a consequence, it is not a surprise that actual securities returns are fat-tailed.

## III. Estimating the Covariance Matrix

This section discusses the covariance matrix estimation methods used at Goldman, Sachs & Co. Our methods use historical data to account for the empirical regularities discussed in Section II. In estimating covariances, most approaches, including ours, amount to taking a weighted average of products of past returns. For volatilities, the historical returns of an asset are squared and averaged; for correlations, the products of the returns of two assets are averaged. The main choices are how to compute the historical returns and how to weight the products.[5]

To introduce a little formalism, let us suppose that $r_{it}$ is the daily return on the *i*th asset at date *t* and that $w_t$ is the weight applied at date *t*. Furthermore, suppose that the investment horizon is *M* days and that the variance of returns of asset *i* at time *T* over the horizon *M* is denoted $\sigma_{ii}^{t}(M)$. Assuming that mean returns are zero, our estimator of $\sigma_{ii}^{T}(M)$ is given as:

[5] Computation of historical returns is also an important issue in estimating the covariance matrix. For example, differences can arise when total returns are used instead of excess returns (at Goldman Sachs, we use excess returns). Differences can also arise when total bond returns are approximated by the product of a bond's duration and the corresponding yield change.

(1) $$\sigma_{ii}^{T}(M) = (\sum_{s=0}^{T} w_{T-s} r_{iT-s}^{2}) \;/\; (\sum_{s=0}^{T} w_{T-s})$$

We can define an estimator of the covariance similarly. By letting $r_{jt}$ represent the return on asset $j$ at date $t$, we can write the covariance over the horizon $M$ between $r_{it}$ and $r_{jt}$ at date $T$ as:

(2) $$\sigma_{ij}^{T}(M) = (\sum_{s=0}^{T} w_{T-s} r_{iT-s} r_{jT-s}) \;/\; (\sum_{s=0}^{T} w_{T-s})$$

Clearly, there are many possible choices for the weights $w_t$. For instance, if we choose to give equal weight to every observation in the sample, then $w_t = \frac{1}{T+1}$ for all $T$ observations. At Goldman Sachs, however, we choose $w_t$ to be a declining function of time: We give more weight to observations that occurred more recently than to observations that occurred in the more distant past. For example, if $w_t$ is a daily weight and 100% weight is given to the most recent observation, then $w_T$ = 1.0. Now suppose that each day back in history receives 90% of the weight of the following day. Then $w_T$ = 1.0, $w_{T-1}$ = 0.90, $w_{T-2}$ = 0.81, etc. Weighting data in this manner is also known as decaying data, and [1- $(w_{t-1}/w_t)$] is called the decay rate. In our example, if each day back in history receives 90% of the weight of the day that follows it, then the decay rate is 10%.

Representing the variance and covariance as in equations (1) and (2) raises a number of questions: First, what are the differences between using daily data and sampling less frequently? Second, what is the value in using declining weights? Third, how is the decay rate chosen? Fourth, how do the estimators in (1) and (2) address the empirical regularities described in the previous section? Finally, what is the relationship between the decay rate and the investment horizon (i.e., $M$)? We address each of these questions in the discussion that follows.

**What Is the Best Sampling Frequency?**

In finance, we often use historical data to estimate mean returns and the volatility of returns. Statisticians often refer to these quantities as the first and second *moments*. Such lumping together of these moments tends to hide a fundamental difference in their estimation. Mean returns are a function of the beginning and ending value of an asset. Knowing how the value changed from the start to the end is irrelevant. Volatility, on the other hand, is all about knowing how the value has changed. In fact, if you could know enough values over any short time interval, you could get as precise an estimate of volatility as you might want — at least in theory. In practice, there are limits as to how much information you can obtain from any price series, but it is worth remembering the basic truth that measuring returns over shorter intervals provides more information about volatility. On the other hand, we have found that as the return data are taken over shorter and shorter intervals, problems with the quality of the data arise. Prices are sometimes “noisy.” The price for a small transaction may not reflect what would obtain for a transaction of larger size. Over very short intervals, the price changes may reflect bid/offered spreads. Also, daily returns in different time zones are not synchronous. Monday’s returns on Japanese assets will reflect information revealed Friday in New York, while Monday’s returns in New York may reflect information that is not revealed until after the Japanese markets are closed on Monday. Finally, there are very significant intraday changes in volatility reflecting, for example, the release of economic data, the time of day, or trading hours for futures exchanges.

The approach we take at Goldman Sachs to alleviate some of these effects on daily returns is to experiment with a parameter we call the *overlap*. Although wherever possible we start with returns measured at least as often as daily, using overlapped data with parameter $k$ means that the returns used in the covariance matrix are $k$-day overlapping returns rather than one-day returns: For example, when $k$ is two, we use an average of two days’ returns. This averaging of the returns reduces the impact of effects that persist for only a short period. By experimenting with the overlap parameter, we can investigate the trade-off between the loss of information that occurs when using returns sampled less frequently and the contamination that occurs when using returns sampled too frequently.

Exhibit 6 illustrates the differences that occur when different values of the overlap parameter and different amounts of data are used to estimate volatility. The charts plot time series of rolling estimates for the volatility of excess returns for the U.S. Treasury 10-year note from January 1, 1992, through December 31, 1996. Each estimator uses a rolling window of some period of data and equally weights all observations in the window. We show the effect of different periods of data and different overlaps. The estimator based on an overlap of 22 in effect uses monthly data, while the estimator based on an overlap of 5 in effect uses weekly data. We also show estimators based on overlap parameters of 2 and 1 (which corresponds to daily data). We show estimators using rolling windows of two, 12, and 24 months.

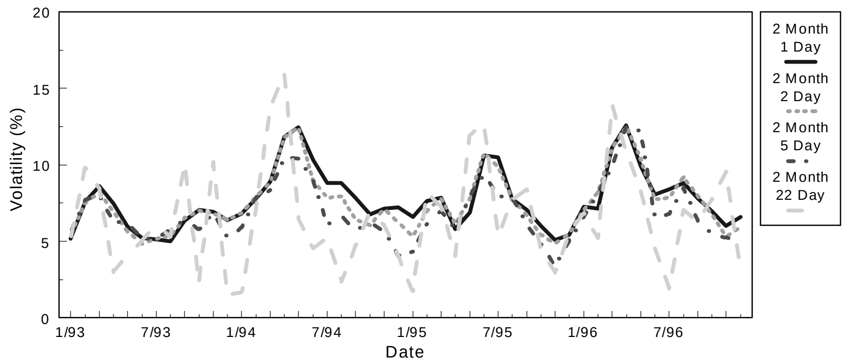
What is obvious in Exhibit 6 is that for long windows with equally weighted observations, the overlap parameter does not matter much. For short windows, the estimators based on larger overlaps tend to be badly behaved; they are noisy because the averaging involved in computing returns over longer periods reduces the overall amount of information.

Rather than using equal weights, as in Exhibit 6, we usually use a decay factor. We do so because we feel that volatilities and correlations tend to vary over time, and thus the older the returns are at any time, the less relevant they are for revealing what the covariance structure is at that time. If we knew something about the process by which the covariance structure is changing, then we could perhaps more carefully set the weights for older data. However, given the large number of assets for which we want to compute volatilities and correlations, we prefer the simpler, though perhaps less exact, approach of searching over various decay rates.

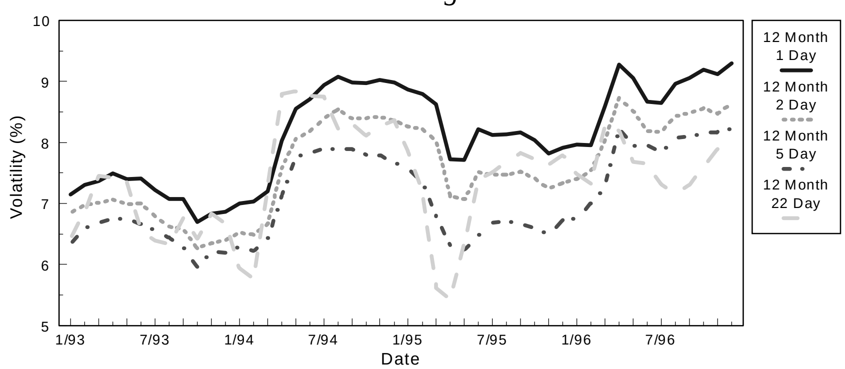
## Exhibit 6

## Volatility Varies With Overlap and Window

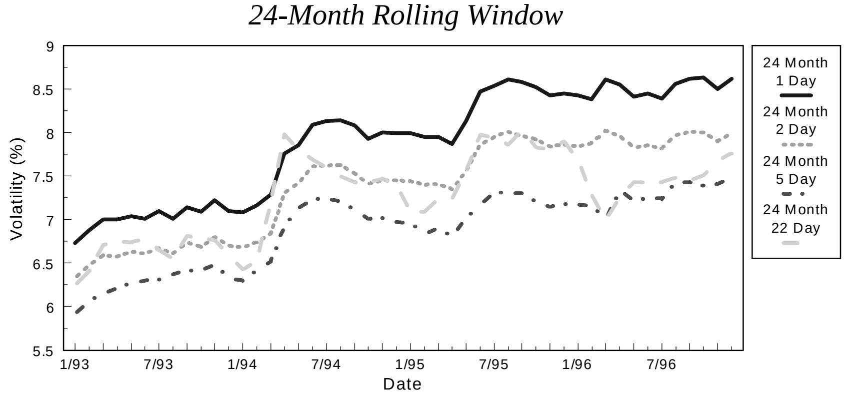
*Two-Month Rolling Window*




*12-Month Rolling Window*




*24-Month Rolling Window*



Using a decay rate can be deceptive. We may fool ourselves into thinking we are using a long history of data. Even with a rapid decay rate, we include lots of old data in our estimator, but the problem is that the old data will get very little weight. Thus, the effect of using a rapid decay rate is very similar to that of using a short data window; the covariance estimator will be noisy. As we shall see, rapid decay rates can easily lead to problems arising from not having enough information to get well-behaved covariance estimators. Just as in Exhibit 6, such a problem will be compounded by having a larger overlap parameter.

**How Can We Choose Decay Rates?**

Equations (1) and (2) show that variance and covariance estimators depend on historical data and the decay rate. As discussed, the choice of decay rate will influence the variance and covariance estimates. Consequently, we need an objective to help us to choose between many different decay rates. Ideally this objective would also be helpful in exploring other issues, such as the relationship between the decay rate and the investment horizon or the choice of distributional assumption. We will assume that our objective at any time is to best explain the distribution of subsequent returns. We prefer a decay rate for which the distribution is most consistent with subsequent returns.

Statisticians give a name to a function that returns the probability of an outcome; they call such a function a “likelihood function.” We will define a likelihood function for return vectors that is itself a function of a covariance matrix. The covariance matrices we consider are functions of decay rates. Thus, we will search over decay rates to find the one for which the subsequent return data is most consistent based on the probabilities defined by the likelihood function.

One key assumption in this approach is the form of the distribution for future returns. We have to allow for fat tails. If, instead, we were to assume the distribution is multivariate Normal, then the occasional very large return vector would be viewed by the likelihood function as being very unlikely, and the decay rate that maximizes the likelihood of those few observations would be chosen. We choose a distribution that makes fat-tailed returns more likely in general, since our objective is to be consistent with all the returns. One simple fat-tailed distribution that we use is a mixture of Normals. That is, we assume that most of the time return vectors are drawn from a low volatility state; but every so often returns are drawn from a high volatility state. We assume the correlations in both states are the same (this assumption may be relaxed in future research), and we assume that all the volatilities in the high volatility state are a constant multiple of the volatilities in the low volatility state. For a given decay rate, we estimate a probability of being in the high state, the ratio of the volatilities in the two states, and the two resulting covariance matrices. Given these, the likelihood of any sample can then be computed (since we've assumed a particular distribution for future returns). The sample we look at is the *k*-day returns during an investment horizon subsequent to the data period over which the covariance matrices are estimated. The likelihood measures the probability that the given sample was generated by the assumed distribution and the estimated parameters (i.e., the variances and covariances). Since the estimated variances and covariances depend on the decay rate, the likelihood function also depends on the decay rate. Thus, the objective is to pick a decay rate that maximizes the likelihood function. Since changing the investment horizon changes the specific return sample of the optimization problem, the optimal decay rate can be expected to change as well. (Appendix B provides mathematical detail on the likelihood function.)

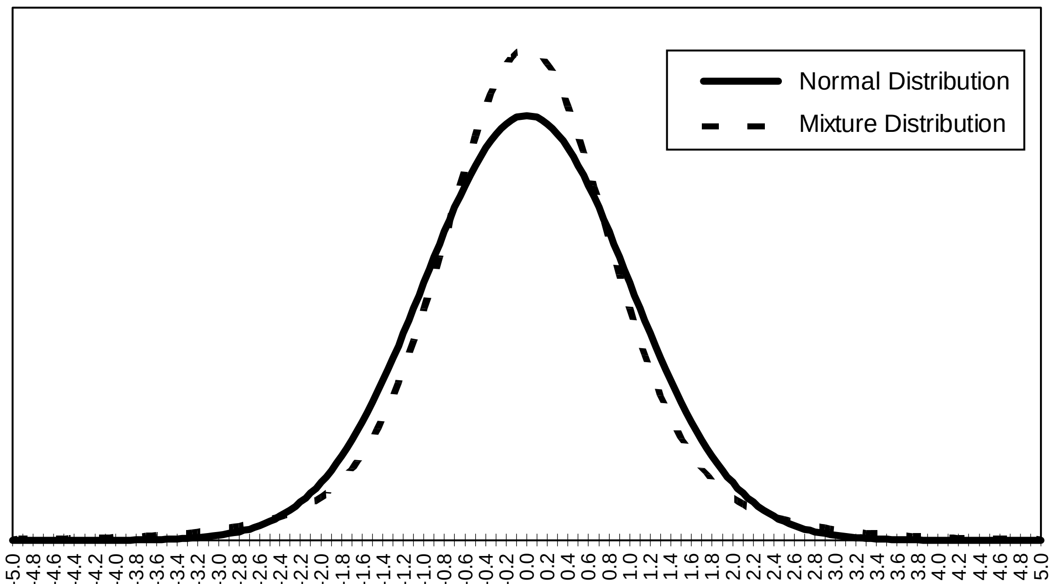
## A Mixture Distribution Produces Fat Tails

At Goldman Sachs, we assume that daily returns follow a mixture of Normal distributions rather than a single Normal distribution. As discussed above, we assume that returns are generated by a low volatility regime coupled with periodic episodes of high volatility. The range of projected daily returns will depend on the volatilities in the low and high volatility regimes and the probability of each regime's occurrence.

**Exhibit 7**

## Comparing a Mixture Distribution With a Normal Distribution

*(Both Distributions Have a Standard Deviation = 1.0)*




In Exhibit 7, we contrast a Normal distribution with a mixture distribution. In each distribution, the standard deviation is equal to one. (Of course, the volatility of the mixture distribution will depend on the volatility in each of the regimes and the probability of each regime's occurrence). Fat tails are created by shifting "probability mass" from the one- to two-standard-deviation range (in absolute value terms) to the other regions of the distribution. Relative to a Normal distribution, the mixture distribution has more probability mass between zero and plus or minus one standard deviation, and beyond plus or minus two standard deviations. (Exhibit 5 provides a practical example).

How does a mixture distribution help us find optimal decay rates? When returns are really distributed with a mixture distribution, but a Normal distribution is assumed, the decay rate will be suboptimal. In other words, the decay rate

**Exhibit 8**

**Decay Rates for German 10-Year Bond**

*One-Day Horizon*

| Distribution | Decay Rate | Likelihood Function |
|---|---|---|
| Normal | 0.05 | -267.41 |
| Mixture | 0.03 | -246.10 |

that maximizes the log-likelihood function under a Normal distribution will be different from the decay rate when returns really follow a mixture distribution, since the Normal distribution cannot account for the large number of returns outside the two-standard-deviation range.

Exhibit 8 illustrates this point by contrasting the optimal decay rates for German 10-year bond returns. We found decay rates by maximizing the likelihood function for a one-day horizon under two distributional assumptions: The first optimal decay rate assumes that returns are Normally distributed, while the second assumes that returns are generated with a mixture distribution. The decay rates were estimated using daily returns on German 10-year bonds from January 1, 1992, through December 31, 1996. As shown in the table, the optimal decay rate is 5% when a Normal distribution is assumed versus 3% when a mixture distribution is used.

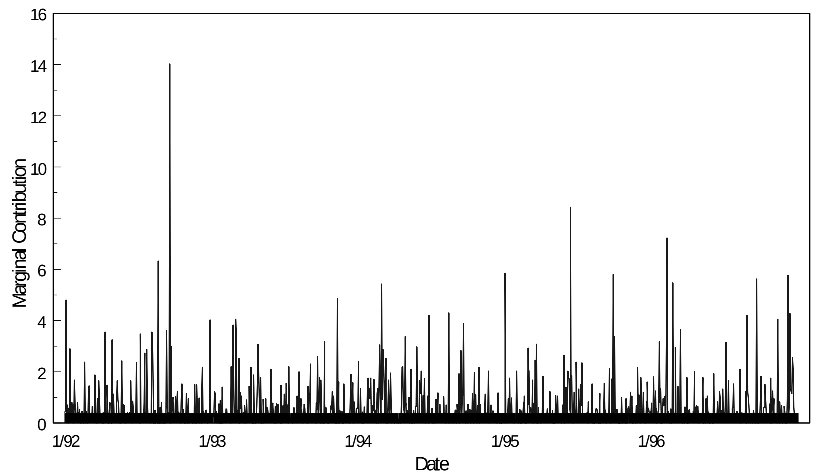
Exhibit 8 also shows the corresponding values of the log-likelihood function. With the Normal distribution, the value of the log-likelihood function is -267.41, whereas with the mixture distribution, the value of the log-likelihood function is -246.10. The two log-likelihood values can be compared directly, as the single Normal distribution is the limiting case of the mixture distribution.[6] Since the log-likelihood function takes on a higher value in Exhibit 8 when the mixture distribution is used, we can conclude that in this example, the Normal distribution assumption does not

[6] We can test for the significance of the differences by using a likelihood ratio test. This test shows that we can reject the hypothesis of no difference at the 1% level.

**Exhibit 9**

## Marginal Contributions to Likelihood

*German 10-Year Bond Returns*




Note: Assumes Normal distribution and 5% daily decay.

account for the extraordinary number of “explosive” returns.[7]

Exhibit 9 illustrates the effect of an extraordinary number of “explosive” returns. The chart plots the marginal contribution of each day’s return on the value of the log-likelihood function from January 1, 1992 through December 31, 1996 (giving 1,302 daily observations), using a Normal distribution and the optimal decay rate of 5%. When the distribution is Normal, the marginal contribution of each day’s return on the value of the log-likelihood function is simply the squared daily return divided by the predicted variance

[7] How does decaying the data compare with using a rolling window? In principle, each method is attempting to estimate the volatility of asset returns by giving more weight to more recent data. The difference is that when we decay the data, we use all of the asset’s history of returns. By contrast, a rolling window actually eliminates some of the return history and therefore eliminates one source of information. The contrast between using a rolling window and decaying the data can be seen by looking at the likelihood function. For the German 10-year bond, the optimal decay rate for a one-day horizon is 5%, and the value of the logarithm of the likelihood function (log-likelihood) is -267.41. By contrast, the optimal window for the German 10-year bond is 110 days, and the corresponding value of the log-likelihood function is -285.62. (We found the optimal window size by searching through window sizes to find the size that maximized the one-day-ahead likelihood function). Clearly, there is a potential gain in estimating volatility by using a decay rate rather than a rolling window.

(or volatility squared): as the squared daily return increases, the marginal contribution to the log-likelihood function increases. As the chart illustrates, there are 60 days that are outside two standard deviations, which is 16 more than would be predicted if German 10-year bonds actually followed a Normal distribution.

Now suppose that this sample's 27 largest daily returns equaled the average of the 27 smallest returns (in absolute value). In other words, 27 daily returns are shifted from outside two standard deviations to inside one standard deviation. Under this scenario, the optimal decay rate (assuming a Normal distribution) is now 4%, and the value of the log-likelihood function is -127.94. Thus, in this example, the effect of assuming a Normal distribution when returns are really fat-tailed is to "overdecay" the data. (Cases could also exist where the Normal distribution assumption leads to "underdecaying" the data.)

The implication of suboptimal decay rates is to induce more uncertainty in our volatility estimates. In a portfolio setting, additional uncertainty in our volatility estimates implies less confidence in our estimates of VaR. Consequently, portfolio managers (a) will not get maximum leverage from their portfolio management decisions (e.g., position sizes for trading portfolios or deviations from benchmark for funds managed against indexes) and (b) will still run the risk of larger-than-expected performances outside a two-standard-deviation range.

### The Decay Rate Decreases as the Investment Horizon Increases

A second important element that influences the optimal decay rate is the choice of investment horizon — i.e., how long the investor expects to hold the portfolio or position. Typically, as the length of the investment horizon increases, more of the historical data are used to calculate volatility (and correlations). The optimal decay rate for a one-day investment horizon is higher than the optimal decay rate when the portfolio is rebalanced every three months. Thus, the optimal estimator of volatility for a three-month horizon is not merely a scaled-up version of the optimal estimator for a one-day horizon.

Suppose that asset returns were independently distributed across time with a constant variance. Under these assumptions, one strategy would be to determine the optimal decay rate for a one-day horizon and then scale the daily volatility estimate up by the square root of time for longer horizons. For example, the volatility estimator for a one-month horizon would simply be the volatility estimator for a one-day horizon multiplied by the square root of 22 (assuming 22 trading days in one month).

As discussed in the preceding section, most financial time series do not have a constant variance. Indeed, most financial time series can be characterized by time-varying daily volatility that seems to mean-revert to some longer-run volatility. For daily rebalancing horizons, our inclination is to give more weight to previous daily volatility and less to the longer-run horizon. Under this scenario, we would be led to assign more weight to more-recent daily returns. By contrast, when the rebalancing horizon is infinitely long, we would naturally be inclined to give more weight to the longer-run volatility and less to particular daily volatilities. We can achieve this end by applying more weight to the past returns (in other words, less weight to more-current returns).

We can illustrate the relationship between the investment horizon and the optimal decay rate by returning to our example of German 10-year bond returns. We have already seen that the optimal decay rate for a one-day horizon is 3% when we assume that returns are generated by a mixture distribution (recall that the mixture distribution helps us account for fat tails). What happens to the decay rate when our investment horizon increases?

Exhibit 10 shows the optimal decay rates for the German 10-year bond at one-day, one-week, and one-month investment horizons. As with the previous examples, the sample consists of daily excess returns on German 10-year bonds from January 1, 1992, through December 31, 1996. As expected, the optimal decay rates depend on the investment horizon. Over the shortest time horizon, when less historical information is likely to be useful, the decay rate is 3%. However, as the investment horizon increases, the information content in past returns also increases, and the decay rate decreases. As shown in the table, at a one-month horizon the optimal decay rate is 1.5%.

**Exhibit 10**

**Decay Rates for the German 10-Year**

*(Assumes Mixture Distribution)*

| Horizon | Decay Rate |
|---|---|
| One Day | 0.030 |
| One Week | 0.023 |
| One Month | 0.015 |

Does the choice of investment horizon make a practical difference for managing large portfolios? Quite clearly, the answer is yes. Investors or traders who do not calibrate their volatility and correlation estimates according to the projected rebalancing period run the additional risk of misestimating volatility and consequently will not take the proper portfolio management decisions when they are required. For example, scaling up a daily volatility estimate by the square root of time during periods of low volatility implies that insufficient weight is given to historical periods of high volatility. For longer horizons, the projected VaR will tend to underestimate the "true" VaR.

**What Happens When We Add Assets?**

The final issue that we need to confront is what to do about adding more assets. In other words, how should the extra information provided by co-movements between assets be handled. The reason adding assets presents a challenge is that the dimension of the covariance matrix increases by the factor [N(N+1)/2], where N is the number of assets, rather than linearly. For example, a two-dimensional covariance matrix (two assets) has three separate parameters of interest, a three dimensional covariance matrix (three assets) has six separate parameters, etc. As we increase the size of the covariance matrix by adding assets, we introduce an additional element of uncertainty into the estimation process. Indeed, for large-scale problems, there will typically be insufficient data to construct a well-behaved covariance matrix. This implies that analysts must reduce the dimensionality of the problem (for instance, by using factor models). However, a potential price for solving a more parsimonious problem could be a reduction in the ease of computation.

At Goldman Sachs, we maintain the assumption that returns are generated by a mixture of Normal distributions, but we replace the assumption of a univariate distribution with that of a multivariate distribution and add terms that account for co-movements between asset returns. Of course, when we add these terms, we also need to describe their co-movements in each regime. As discussed above, we assume that the correlation structure is the same in each regime (an assumption that will be the subject of future research). Nothing else fundamental really changes, though; we still want to find decay rates that maximize the log-likelihood function at different horizons.

Let's look at the effects of including more assets by expanding our German 10-year bond example to include the Italian 10-year bond. As in the previous examples, the data set includes daily returns from January 1, 1992, through December 31, 1996. Once again, the objective is to find a decay rate that maximizes the likelihood function.

Exhibit 11 contrasts decay rates for German and Italian 10-year bonds at horizons of one day, one week, and one month. The table shows decay rates for each market under univariate and bivariate assumptions. As in the univariate case, the decay rate decreases as the investment horizon increases. However, as the table makes clear, the optimal decay rate depends on the number of assets. The optimal decay rate for the bivariate problem is different from that of either univariate problem.

An alternative possible approach would be to find optimal decay rates for portfolios of assets. In this approach, a one-dimensional time series of returns is developed from a portfolio of asset returns, and an optimal decay rate is found for the portfolio returns. An optimal decay rate (in some sense) can be found by experimenting with many such portfolios.

## Exhibit 11

### Decay Rates for Italian and German Bonds

| | Decay Rate | | |
|---|---|---|---|
| **Horizon** | **Germany** | **Italy** | **Combined** |
| One Day | 0.030 | 0.040 | 0.020 |
| One Week | 0.025 | 0.023 | 0.018 |
| One Month | 0.015 | 0.015 | 0.010 |

This approach has the advantage of reducing the problem to a manageable size, particularly when the only variable of interest is the overall portfolio volatility (VaR). However, this approach gives no useful information on the covariance terms, which is a drawback for broader risk management and asset allocation problems. Indeed, knowledge of the covariance terms is crucial for an understanding of the risk structure of a portfolio, as well as for determining an optimal set of portfolio weights.

At Goldman Sachs, we are continuing to explore alternative methods for estimating covariance matrices with a large number of assets. As discussed, the issue is to find methods that are easy to compute yet provide answers to risk control and optimization problems.

In this section, we have discussed the covariance matrix estimation procedures used at Goldman Sachs. We have also discussed some of the trade-offs involved in selecting various parameter values, Of course, other estimation methods have been proposed by practitioners. In the next section, we discuss some of these methods and their relationship to the approach taken at Goldman Sachs.

## IV. Alternative Covariance Matrix Estimation Methods

The driving force behind all the techniques that have been proposed for estimating the covariance matrix is the desire to account for the empirical regularities described in Section II. This section discusses some of these alternatives and compares them with the method used at Goldman Sachs. The four principal methods we discuss are implied volatilities, GARCH, Markov chains, and identification of stationary and transitory components.

### Implied Volatilities

One popular method for estimating volatility is to use the volatilities implied by options markets. Rather than use historical data on asset returns, this method uses actual option prices with an option pricing model to infer an asset's implied volatility. This approach has the apparent advantage of giving indications about volatility that are consistent with the expectations of market participants. Since implied volatilities are calculated with market prices for the underlying security and the option, they can also be regarded as being "forward-looking" estimates of volatility. However, several distinct disadvantages exist in systematically applying implied volatility to risk management and asset allocation problems.

The first issue that is important in the context of both risk management and asset allocation is the range of financial products. If implied volatilities are to be useful, they must cover the range of products likely to be of interest to risk managers and asset allocators. Although sufficient liquidity across a broad product range may exist in the future, such is not the case at present. Even for the simplest of asset allocation problems (e.g., managing a global bond fund), there are too few derivatives products with sufficient liquidity for implied volatility to be useful. From the perspective of broader risk management, the problem is compounded. At Goldman Sachs, we view our portfolio in the context of more than 2,000 risk factors, each of which represents a specific financial product. Actively traded derivatives are available for only a handful of these products.

A second, and related, issue is the estimation of covariances. As the name suggests, implied volatility provides a snapshot of market participants' beliefs regarding volatility. However, there are too few products with prices sensitive to correlation to provide a similar snapshot on covariance.

Finally, it is important to bear in mind that implied volatility is the result of applying an options pricing model to observed options prices. Consequently, practitioners who rely on implied volatility expose themselves to modeling risk. In other words, practitioners will need to evaluate whether their implied volatility figures are being generated by a reasonable model (which is a good practice for any model).[8]

The issues described above suggest that a broad role may not (yet) exist for implied volatilities in risk management and asset allocation. However, a narrower role may exist. Specifically, implied volatilities can be used as a diagnostic tool for models based on historical data. Thus, covariance matrix estimation and risk management and asset allocation procedures can be improved by using implied volatilities as a diagnostic tool.

**GARCH Models Are Similar to Decayed Daily Data**

The GARCH (generalized autoregressive conditional heteroscedastic) method for estimating volatility assumes that the current level of the variance (i.e., squared volatility) depends on past variances and current and past squared returns. Using notation, we can write this idea as:

$$(3)\ \mathrm{var}(R_t) = A_0 + \sum_{i=1}^{I} A_i \,\mathrm{var}(R_{t-i}) + \sum_{i=i}^{J} B_j R_{t+1-j}^2$$

Assuming that the values of $A_i$ and $B_i$ are positive, equation (3) shows that the current variance will be above its long-run mean during periods of high volatility. Furthermore, periods of abnormally large returns (either positive or negative) will increase the current level of volatility. In addition to accounting for fat-tailed distributions of asset returns, the GARCH representation also captures the "volatility clustering" that is typically observed with financial time series.

To implement a GARCH model, the analyst must establish the amount of history to be used (i.e., the value of $I$) and then estimate values for $A_i$ and $B_i$. These parameters can be estimated with maximum likelihood methods.

[8] For a detailed discussion of these considerations, see Derman (1996).

Suppose that $I$ and $J$ are known to be one, and that $A_0$ is known to be zero. Under these assumptions, equation (3) can be simplified to:

$$(4)\ \mathrm{var}(R_t) = A_1 \mathrm{var}(R_{t-1}) + B_1 R_t^2$$

Notice that equation (1) — the equation that represents the variance as an exponential decay — can be conveniently rewritten as:

$$(1')\ \mathrm{var}(R_t) = (1 - w_t)\,\mathrm{var}(R_{t-1}) + w_t R_t^2$$

Comparing equations (4) and (1′) shows that estimating the variance by decaying daily data is equivalent to using a very simple GARCH model. Alternatively, equation (1′) can be interpreted as a restricted GARCH model. In other words, by decaying daily data, we are implicitly using a GARCH model but imposing a restriction on how the data are generated.[9]

One of the features of the GARCH modeling procedure is that it provides an exact equation for the conditional volatility over multiple-period horizons. That is, once we know the parameters for a GARCH process estimated on daily data, we can find the conditional volatility for both one-day and multiple-period horizons. Thus, such a model would provide a more consistent way to estimate volatilities over longer time horizons than estimating separate covariance matrices with different decay rates. The fact that we find different decay rates to be optimal for different horizons suggests that their use is best viewed as a simple and computationally less intensive way to represent a more complex model that would build in mean reversion of volatility.

It is important to bear in mind that GARCH models and mixtures of Normals are attempting to account for the same empirical regularities. Since we can never know the "true" data generating process, we are forced to use some sort of model and thus expose ourselves to potential specification error. Consequently, for practical purposes, we must consider the trade-off between using a more complicated model (one that could potentially bring us closer to the

[9] In principle, the restriction can be tested. However, we have not done so for the purposes of this paper.

"true" data generating process) and computational ease. In the context of GARCH model, this issue comes up when more assets are added.

Suppose that instead of a single asset, we now have two assets. In this case, we need to expand the model to describe the behavior of the variance of returns for the second asset and describe the covariance between the returns on the second asset and the first asset. Let's continue in the simple environment and assume that each asset's current variance depends on its most immediate past variance and the asset's current squared return. Furthermore, let's assume that the current covariance depends on the most immediate past covariance and the product of the returns on both assets. Using these simple assumptions, we will need to estimate six parameters to implement a GARCH model.

Notice that the number of parameters does not increase linearly. For example, if we continue to restrict $I$ to be one but add another asset, we now have 12 parameters to estimate. Indeed, if we have $N$ assets in the covariance matrix, and $I$ is set to one, we will need to estimate $(N+1)\times N$ parameters to implement a GARCH model. Since an unrestricted GARCH model has so many parameters, it stands to reason that the estimation process can be quite computationally burdensome.

### Markov Chains Add Volatility Clustering

An alternative approach to generating fat tails and volatility clustering is to expand on the idea of mixture distributions by adding more structure to the probability of each regime's occurrence. In the previous section, we assumed that returns are drawn with some probability from either a low- or a high-volatility regime. We also assumed that each day's return is independent of previous returns.

With a Markov chain, we relax the idea that regimes are independent across time. More specifically, we add an assumption that makes the probability of drawing from each regime dependent upon the current regime. When we have two possible volatility regimes (low and high), there are now four additional probabilities to be considered: the probability of drawing next period from the low volatility (or high volatility) regime when we are currently in the low volatility regime, and the probability of drawing from the low volatility (or high volatility) regime when we are currently in the high volatility regime. These additional probabilities are called the *transition probabilities*.[10]

As with other methods, including transition probabilities has costs and benefits. Recall from the discussion in the previous section that using mixture distributions gives us fat-tailed distributions. Including the transition probabilities has the advantage of generating volatility clustering. However, the disadvantage is that including the transition probabilities means that we increase the number of parameters that must be estimated. We are continuing to explore the use of Markov chains for risk management and asset allocation problems.

**Stationary and Transitory Components**

A final technique for estimating the covariance matrix is to build on the idea of volatility regimes by adding an economic interpretation. More specifically, returns are viewed as coming from a combination of stationary and transitory components. In this framework, asset returns are largely driven by the stationary component. However, there are periods when asset returns drift away from the stationary component. These periods are called the transitory component.[11]

During the periods when returns are generated by the stationary component, the distribution of asset returns is assumed to be Normally distributed with a constant mean and a constant variance. Consequently, the volatility of asset returns over longer time horizons when returns are generated by the stationary component is merely the volatility of the daily return times the square root of the time horizon.

However, returns are not always generated by the stationary component. In the transitory component, asset returns are assumed to be generated by a distribution that also has a constant daily variance, but whose expected return is mean-reverting. In this setup, the volatility of an asset's return over longer horizons is no longer merely the volatility of the daily return times the square root of the horizon. Now, the volatility over longer horizons depends on the parameters of the mean reversion process. In other words,

[10] For additional information on Markov chains, see Hamilton (1994).

[11] See, for example, Chou and Ng (1995).

volatility in the transitory component depends on how far away the observed return is from the long-run mean in the transitory component.

By combining the stationary and transitory components, we can see that the unconditional covariance matrix depends upon the investment horizon. The dependence on the investment horizon is a consequence of mixing distributions with different variances and the dependence of the transitory component on the mean-reversion parameters. Notice that the distinction between this approach and the mixture distribution approach discussed earlier is the addition of an economic structure to one of the volatility regimes.

The analysis of stationary and transitory regimes can be further enhanced by adding assumptions about the transitions between regimes. The easiest such assumption is that the regimes are independent across time. In other words, knowledge of the current regime provides no information about the regime most likely to occur next period. Under this assumption, analytic solutions for the covariance matrix can be easily derived and shown to depend on the parameters of the mean-reversion process and the investment horizon.

Alternatively, we could assume that regimes are not independent across time and formulate explicit transition probabilities. As discussed above, including explicit transition probabilities has the advantage of additional realism but also imposes additional computational burdens.

At Goldman Sachs, we are continuing to explore the applicability of this technique — modeling stationary and transitory components — to large-scale risk management and asset allocation problems. Our research looks at various assumptions about the structure of the transition probabilities.

## V. Covariance Matrices and Structural Models

Covariance matrix estimation also plays an important role in identifying parameters for structural models. Structural models relate the sensitivity of asset returns to changes in underlying fundamentals. They are used to provide signals about expected returns based on projections of these same fundamentals. Since the parameters in structural models are also related to the composition of the covariance matrix, how the covariance matrix is estimated has important implications for the identification of structural parameters.

To consider a simple illustrative example, suppose that we model German bond returns as functions of two observable variables: changes in the level of German 10-year interest rates and changes in the slope of the German curve. We will model the slope of the German curve as the yield spread between 10-year and two-year bonds. By relating the two interest rate variables to underlying economic fundamentals (e.g., German GDP growth), this structure allows us to develop optimal portfolio strategies, given views on these same fundamentals. Of course, to implement any such strategy, we will need to determine the responsiveness of German bond returns to our two interest rate variables (Note that we could add even more structure to our model by making explicit assumptions about the dynamics of our interest rate variables).

The relationship between German bond returns and our two interest rate variables can also be used when we have no views on fundamentals. For instance, suppose that we have no view on underlying fundamentals, but that we wish to exploit mispricings in the German bond market. Under this "view," we may wish to find a portfolio of securities that is neutral with respect to changes in level or slope. Once again, this portfolio is dependent on the sensitivity of German bond returns to the two interest rate variables.

One popular approach to identifying these parameters is to regress individual bond returns on yield changes and changes in the slope of the curve — using, say, least squares methods.[12] For example, if we restrict our attention to the six benchmark bonds (two-year, three-year, five-year, seven-year, 10-year, and 30-year) in Germany, we have a set of six regressions that we can use to obtain a set of sen-

[12] See, for example, Singleton (1994).

sitivities. We can explore this idea a bit further by introducing some simple notation.

Let $R_i$ denote the return on the *i*th German benchmark bond, *dl* denote the change in the level of the German 10-year bond, and *ds* denote the change in the slope of the German curve. We will assume that the level and slope effects are independent of each other. We denote the sensitivity of the bond's return to changes in level and changes in slope as $b_i^l$ and $b_i^s$. Finally, suppose that there is an error term for each bond that has a zero mean, has some variance, and is independent across time. With this notation, we can write each bond's return as:

$$R_i = b_i^0 + b_i^l \, dl + b_i^s \, ds + e_i \tag{5}$$

We can also use equation (5) to represent the variance of any bond's return in terms of the underlying structural parameters $b_i^l$ and $b_i^s$ and the residual variance. Equation (5) implies that:

$$\text{var}(R_i) = b_i^{l^2} \, \text{var}(dl) + b_i^{s^2} \, \text{var}(ds) + \text{var}(e_i) \tag{6}$$

By applying equation (5) to two bonds, we can easily see that the covariance between them depends on the level and slope parameters. If we denote the covariance between bonds i and j as cov($R_{i,} R_{j,}$) and apply equation (5), we see that:

$$\text{cov}(R_i, R_j) = b_i^l b_j^l \, \text{var}(dl) + b_i^s b_j^s \, \text{var}(ds) + \text{cov}(e_i \, e_j) \tag{7}$$

Equation (7) tells us that the covariance between any two bonds depends on the variance of the level and slope effects, the relative exposures of each bond to these effects, and the covariance between the two error terms. (Notice that if this covariance is zero, then equation (7) implies that no information is lost by estimating the sensitivities independently of each other).

Our discussion in the previous section argued that it is more efficient to use decayed daily data than equally weighted daily data in estimating the variances and covariances between asset returns. Equations (6) and (7) show that the covariances between asset returns depend on the covariances between the two observable factors and the estimated sensitivities. These relationships suggest that we can improve upon the estimation of the underlying struc-

## Exhibit 12

## Factor Coefficients

| | Entire Sample | | Final 90 Days | | Daily Decay | |
|---|---|---|---|---|---|---|
| | Level | Slope | Level | Slope | Level | Slope |
| Two-Year | -1.6277 | 1.4641 | -1.6847 | 1.3850 | -1.6053 | 1.1843 |
| Three-Year | -2.5596 | 1.1949 | -2.8702 | 1.6318 | -2.7971 | 1.4592 |
| Four-Year | -3.7807 | 1.1883 | -4.7283 | 1.8111 | -4.7752 | 1.6199 |
| Five-Year | -4.2372 | 1.1795 | -4.9028 | 1.7269 | -4.8973 | 1.5408 |
| Seven-Year | -5.4415 | 0.5432 | -6.2186 | 0.9418 | -6.3687 | 0.9203 |
| 10-Year | -6.9760 | 0.1118 | -7.0690 | 0.1325 | -7.2167 | 0.2586 |

tural parameters by applying a similar decay rate to the two observable parameters.

Exhibits 12 and 13 explore this point by contrasting three estimates of the correlation matrix of German benchmark bond returns. In Exhibit 12 we show the level and slope coefficients estimated using three different procedures, with daily data from January 1, 1992, through December 31, 1996. All three procedures use least squares regression and differ only in how the data are treated. The first procedure finds the coefficients by equally weighting all of the data in the sample. In the second procedure, we estimate the coefficients using equally weighted data over the final 90-day period. (This procedure can be interpreted as the last window in a time series of rolling 90-day windows). The third procedure estimates the level and slope coefficients by applying a 3% daily decay rate to all of the data (returns, change in level, and change in slope).

Exhibit 13 illustrates the effects of the differences in estimation methods. In this exhibit, we use the relationships between the observable variables and the covariance matrix to find the volatilities and correlations of German bond returns implied by equations (6) and (7). Each correlation matrix shows volatility of bond returns down the main diagonal, with correlations shown on the off diagonals. Matrix A in Exhibit 13 shows the correlation matrix consistent with equally weighted data, matrix B is consistent with the final 90 days of history, and matrix C is consistent with a 3% daily decay rate. As Exhibit 13 illustrates, the differences among the correlation matrices are quite striking.

## Exhibit 13

## Three Covariance Matrices

*A. Entire Sample*

| | 2-Year | 3-Year | 4-Year | 5-Year | 7-Year | 10-Year |
|---|---|---|---|---|---|---|
| **2-Year** | 1.56 | | | | | |
| **3-Year** | .96 | 2.02 | | | | |
| **4-Year** | .91 | .99 | 2.83 | | | |
| **5-Year** | .90 | .99 | .99 | 3.15 | | |
| **7-Year** | .81 | .94 | .98 | .99 | 3.91 | |
| **10-Year** | .75 | .91 | .96 | .97 | .99 | 4.99 |

*B. Final 90 Days*

| | 2- Year | 3-Year | 4-Year | 5-Year | 7-Year | 10-Year |
|---|---|---|---|---|---|---|
| **2-Year** | 1.24 | | | | | |
| **3-Year** | .97 | 2.55 | | | | |
| **4-Year** | .96 | .99 | 3.08 | | | |
| **5-Year** | .96 | .99 | .99 | 3.17 | | |
| **7-Year** | .90 | .98 | .99 | .99 | 3.91 | |
| **10-Year** | .85 | .96 | .96 | .97 | .99 | 4.41 |

*C. Decayed Sample*

| | 2-Year | 3-Year | 4-Year | 5-Year | 7-Year | 10-Year |
|---|---|---|---|---|---|---|
| **2-Year** | 1.22 | | | | | |
| **3-Year** | .99 | 2.00 | | | | |
| **4-Year** | .96 | .99 | 3.26 | | | |
| **5-Year** | .96 | .99 | .99 | 3.33 | | |
| **7-Year** | .92 | .96 | .99 | .99 | 4.22 | |
| **10-Year** | .88 | .94 | .97 | .98 | .99 | 4.76 |

The importance of these differences can be seen in the context of an example. Suppose that our objective is to hedge a long position in the five-year bond with short positions in the two- and 10-year bonds. Furthermore, suppose that we choose the two short positions so as to match exposure to the level and slope effects. The short positions can be found by applying equation (5).[13] Clearly, as we increase the accu-

[13] Assume a long position of 1.0 in the five-year bond, and denote the short positions as $X_2$ and $X_{10}$. The short positions are given as:

$$X_2 = \left\{ b_s^{10} b_l^5 - b_l^{10} b_s^5 \right\} A$$

$$X_{10} = \left\{ b_l^2 b_s^5 - b_s^2 b_1^5 \right\} A$$

$$A = 1 / \left\{ (b_l^2 b_s^{10} - b_1^{10} b_s^2) \right\}$$

### Exhibit 14

## Historical Hedge Performance

| Regression Assumption | Daily Volatility |
|---|---|
| Equally Weighted | 0.079 |
| 90-Day Rolling Window | 0.075 |
| Decayed Daily Data | 0.073 |

Note: All regressions use daily data. All performance results are out of sample. The data range is January 1, 1993, through December 31, 1996.

racy of the estimated structural parameters, we will reduce the variance of the resulting portfolio.

Exhibit 14 shows the annualized daily portfolio volatility using each of the three methods outlined above. Each volatility was estimated by first finding the appropriate daily hedges using only the information available at that date, and then calculating the actual one-day-ahead return on the portfolio. As the table indicates, when more weight is given to more recent data, the portfolio volatility decreases. The model that equally weights the entire history performs worse than the 90-day rolling window or the 3% daily decay. However, the 90-day rolling window also performs worse than the 3% daily decay.

This example illustrates how the choice of covariance matrix estimation method matters for finding hedges. That the choice matters for this example should not be surprising; the 3% daily decay rate was initially chosen to make the covariance matrix between German bond returns consistent with empirical regularities. The premise behind the model in equation (5) is that we can relate bond returns to two observable factors. For actual bond returns to generate the types of empirical regularities we have discussed, then according to equation (5), the source of the empirical regularities must be in either the two observable factors, the error term, or both. The example further illustrates a more general principle: Any structural model should produce a predicted covariance matrix that resembles the actual covariance matrix of asset returns.

Of course, the model shown in equation (5) is a very simple one. We could gain additional insight by imposing additional structure. For example, we could add more structure by using an equilibrium term structure model; our parameters of interest would be those generating the underlying state variables. As in our simpler example, our estimated parameters will also imply a covariance matrix, and in fact we can use the sample covariance matrix of asset returns as part of the procedure for finding the structural parameters.

What advantages accrue from using equilibrium models? One obvious advantage is that equilibrium models can add a "smoothing" to the covariance matrix. This issue becomes more acute when we face large-scale risk management or asset allocation problems. As assets are added to the covariance matrix, its dimension increases and more noise is introduced in the sample covariance matrix. The addition of noise to the sample covariance matrix introduces more uncertainty into the estimation of risk, risk decomposition, and optimal asset allocations. A structural model (or factor model) allows us to represent the covariance matrix of asset returns in terms of a parsimoniously chosen set of underlying factors and has the potential of reducing the uncertainty surrounding risk estimation and optimal portfolio selection.

## VI. Conclusions

Covariance matrices are important for coping effectively with many day-to-day problems in modern finance. However, since covariance matrices cannot be observed directly, they must be estimated. A well-estimated covariance matrix has important implications for calculating a portfolio's Value at Risk and identifying its major risk contributors. Since covariance matrices are critical for finding sources of risk, they are also crucial for constructing hedges (in either the cash or derivatives markets). Finally, since optimal asset allocations are functions of variances and covariances, a well-estimated covariance matrix can lead to more-efficient portfolio allocations.

This paper has described the approach currently used for large-scale risk management and asset allocation problems at Goldman Sachs, and contrasted it with some of the other methods that are commonly used in the industry. While no comprehensive method exists to estimate covariance matrices, the methods used at Goldman Sachs do have the advantages of providing covariance matrices that are consistent with empirical regularities and computational ease.

This paper has offered four contributions. First, it has shown how a likelihood function can be used to find appropriate decay rates. Using a likelihood function means that we can find decay rates for entire sets of asset returns rather than on the basis of individual asset returns. A further advantage of using a likelihood function is that we can compare the effects of different distributional assumptions (e.g. Normal distributions versus a mixture of Normals).

Second, this paper has shown that a mixture of Normals is a more suitable distributional assumption for finding decay rates. The assumption of a mixture of Normals leads to volatility and correlation estimates that are consistent with empirical regularities such as fat tails and time-varying volatility (and correlation). As a result, the optimal decay rates when a mixture of Normals is assumed are different from those found when returns are assumed to be generated by a Normal distribution.

Third, this paper has shown that the optimal decay rate depends on the investment horizon. More specifically, we have shown that as the period between major rebalancings increases, the distant past becomes more important for covariance matrix estimation. We view this result as suggesting that volatilities and correlations are mean-reverting. This finding also suggests that the mixture distribution approach should be amended to include a structure that allows the rebalancing period to affect the likelihood of drawing returns from the alternative distribution.

Finally, this paper has shown that the decay rate depends on the dimension of the covariance matrix. When the number of assets increases, the decay rate also decreases (i.e., more weight is put on the distant past). More of the distant past is used because more information is necessary to estimate the covariance matrix parameters.

We have left several issues as open research questions, which we are continuing to explore. Within the context of the procedures described in the paper, there are four main research topics. First, additional attention should be given to the effects of adding more assets. As discussed, adding assets implies that the dimension of the covariance matrix increases at a nonlinear rate. For portfolios with many assets, a very real restriction on the covariance matrix estimation is implied by the amount of available data.

A second open issue is the estimation of covariance matrices through the use of factor models. The use of factor models can potentially resolve some of the issues that arise when large numbers of assets are included in the covariance matrix.

A third question that we continue to explore is the assumption about correlations in different regimes. In our work to date, we have assumed that covariances in the high volatility regime are scaled up to keep the correlation constant between regimes. We could relax this assumption and let the correlations vary between regimes. One potential advantage of relaxing the constant correlation assumption is a model with higher explanatory power (as measured by a higher value of the likelihood function). However, a potential disadvantage is the actual increase in the number of parameters that will need to be estimated.

The final research question that we continue to examine is the explicit estimation of the transition probabilities. While our current procedures assume that returns are generated by a mixture of Normals, we do not actually estimate the probability of one regime following another (i.e., the transition probabilities). As with our other open research questions, the potential gains in explanatory power must be balanced against the costs associated with the increased computational burden. ■

## Appendix A. Covariance Matrices for Trading Portfolio Example

Below we present the covariance matrices for the trading portfolio example in the Introduction. The initial covariance matrix was estimated using daily data from February 1, 1988, through December 31, 1996, applying a 15% monthly decay. The second covariance matrix uses monthly data over the same period. Both covariance matrices show variances and covariances for currency-hedged 10-year bonds.

### Daily Data

| | Canada | France | Germany | Italy | Japan | Nlg | Spain | Sweden | UK | US |
|---|---|---|---|---|---|---|---|---|---|---|
| **Canada** | 64.60 | | | | | | | | | |
| **France** | 16.33 | 23.88 | | | | | | | | |
| **Germany** | 20.95 | 18.56 | 26.73 | | | | | | | |
| **Italy** | 18.20 | 22.82 | 20.11 | 59.74 | | | | | | |
| **Japan** | 0.54 | 6.27 | 5.45 | 6.94 | 31.82 | | | | | |
| **Nlg** | 18.80 | 20.53 | 21.26 | 22.19 | 6.25 | 24.23 | | | | |
| **Spain** | 14.58 | 23.44 | 21.56 | 45.02 | 2.88 | 22.83 | 60.02 | | | |
| **Sweden** | 11.86 | 22.60 | 19.88 | 33.17 | 11.84 | 22.00 | 32.57 | 50.24 | | |
| **UK** | 19.50 | 18.83 | 23.41 | 21.58 | 4.26 | 21.52 | 22.93 | 20.53 | 43.46 | |
| **US** | 38.48 | 15.77 | 21.18 | 13.53 | 0.50 | 19.14 | 12.95 | 10.89 | 21.87 | 51.90 |

### Monthly Data

| | Canada | France | Germany | Italy | Japan | Nlg | Spain | Sweden | UK | US |
|---|---|---|---|---|---|---|---|---|---|---|
| **Canada** | 61.41 | | | | | | | | | |
| **France** | 24.08 | 41.24 | | | | | | | | |
| **Germany** | 22.35 | 28.91 | 31.74 | | | | | | | |
| **Italy** | 28.07 | 39.11 | 22.48 | 102.31 | | | | | | |
| **Japan** | 25.38 | 16.23 | 16.64 | 11.26 | 38.78 | | | | | |
| **Nlg** | 21.48 | 30.17 | 24.47 | 26.26 | 13.97 | 28.83 | | | | |
| **Spain** | 34.49 | 41.13 | 23.68 | 79.81 | 12.37 | 27.54 | 91.37 | | | |
| **Sweden** | 34.86 | 34.90 | 23.52 | 57.81 | 15.43 | 28.57 | 65.59 | 94.90 | | |
| **UK** | 29.66 | 26.89 | 24.53 | 34.29 | 21.27 | 25.70 | 38.83 | 36.71 | 55.85 | |
| **US** | 38.97 | 18.51 | 19.78 | 6.67 | 19.25 | 18.16 | 12.15 | 19.02 | 19.65 | 47.72 |

## Appendix B. The Likelihood Function When Returns Follow a Univariate Mixture Distribution

In this appendix, we discuss the estimation of the decay rate when asset returns are distributed as a mixture of Normals. For ease of exposition, we will assume that there are two possible regimes. We assume that returns are Normally distributed in each regime, with a zero mean. Consequently, we can characterize the regimes as low and high volatility regimes. We will also assume that returns are independent across time: Each day's return is uncorrelated with previous and future returns; i.e., the probability of drawing from a particular volatility regime in the future does not depend on the current volatility regime.

Using notation, we denote the density functions for the low and high volatility regimes as $f(R_t,\sigma_L^2)$ and $f(R_t,\sigma_H^2)$, respectively, and denote by $p$ the probability of drawing from the low volatility regime. Since the density functions are assumed to be Normal and centered at zero in each regime $s$, they will have the following form:

$$f(R_{t+1},\sigma_s^2) = \frac{1}{\sigma_s\sqrt{2\pi}}\exp\left\{-\frac{R_{t+1}^2}{2\sigma_s^2}\right\} \tag{B1}$$

In this setup, on any given day the next day's return will depend on each regime's volatility and the probability of each regime. More precisely, if $\sigma_L^2$ and $\sigma_H^2$ are known on date $t$, the density function for $R_{t+1}$ is given as:

$$g(R_{t+1},\sigma_L^2,\sigma_H^2,p) = pf(R_{t+1},\sigma_L^2) + (1-p)f(R_{t+1},\sigma_H^2) \tag{B2}$$

For a sample of returns that begins on date 1 and continues through date $T$, the likelihood function is written as:

$$L(R_t, t=1,....,T) = \underset{t=1}{\overset{T-1}{\pi}} g(R_{t+1},\sigma_L^2,\sigma_H^2,p) \tag{B3}$$

Suppose that the high variance is always a constant proportion of the low variance — in other words, $\sigma_L^2 = k\sigma_H^2$. Suppose further that we use equation (B1) to estimate $\sigma_H^2$. By substituting into equation (B2), the likelihood function now depends on three parameters: $p$ (the probability of the low regime), $k$ (the constant variance proportion), and $w$ (the decay rate). For any given value of $w$, we can find values of $p$ and $k$ that maximize the likelihood function. We find the optimal decay rate by repeating this process for many values of $w$.

Notice that this framework contains the assumption of a standard Normal as a special case. For instance, if returns are always generated by the low volatility regime, then $p$ is always 1.0.

## Bibliography

Chou, Ray Y., and Victor K. Ng. "Correlation Structure of the Permanent and Temporary Components of International Stock Prices," working paper, January 1995.

Derman, Emanuel. *Model Risk,* Goldman, Sachs & Co. Risk Management Series, July 1996.

Hamilton, James D. *Time Series Analysis,* Princeton, N.J.: Princeton University Press, 1994.

Litterman, Robert. *Hot Spots and Hedges,* Goldman, Sachs & Co. Risk Management Series, October 1996.

Litterman, Robert, and Kurt Winkelmann. *Managing Market Exposure,* Goldman, Sachs & Co. Risk Management Series, January 1996.

Singleton, Kenneth J. "Yield Curve Risk in Japanese Government Bond Markets," *Japanese Journal of Financial Economics,* Vol. 1, No. 1, December 1994, pp 5–31.