The Journal of Financial Engineering -Volume 7 -Number 3/4 -Pages 289-309

Vignettes on VAR

Tanya Styblo Beder
Michael Minnich
Hubert Shen
Jodi Stanton

ABSTRACT

We describe three problem areas that often plague those who implement Value at Risk from scratch: mapping, data filling techniques for data series holes, and backtesting. In the mapping discussion, we reveal how the required sophistication of the VAR mapping methodology is proportional to the complexity of the portfolio being modeled. In the data filling discussion, we consider three methodologies to deal with missing historical data and compare their performance; we also illustrate how to combine data series of different frequencies. In the backtesting discussion, we describe two types of backtests and show their performance on sample data. The three vignettes are drawn from three critical areas of a successful VAR implementation-reference is made to the remaining areas.

Tanya Styblo Beder, Capital Market Risk Advisors, Inc., 565 Fifth Ave., New York, NY 100 17.

Michael Minnich, Capital Market Risk Advisors, Inc., 565 Fifth Ave., New York, NY 10017.

Hubert Shen, Capital Market Risk Advisors, Inc., 565 Fifth Ave., New York, NY 10017.

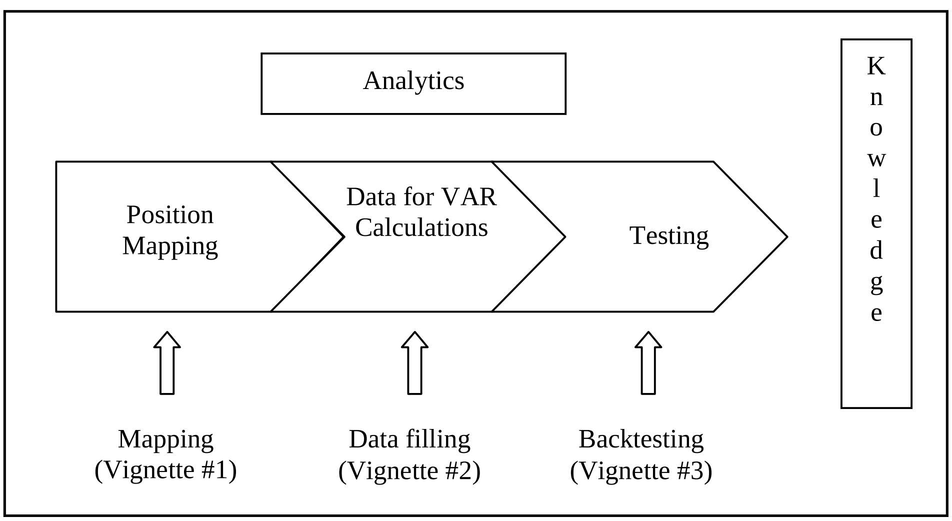
Jodi Stanton, Capital Market Risk Advisors, Inc., 565 Fifth Ave., New York, NY 100 17.

I. INTRODUCTION

Despite continued debate over Value at Risk (VAR)-in particular, whether its strengths outweigh its weaknesses-VAR remains a major advance in risk measurement. Its wide endorsement by regulators, accountants, rating agencies, and key industry groups such as the Group of Thirty, the Derivatives Policy Group, and the International Swaps and Derivatives Association ensure that VAR will continue to be around for years to come. Yet, while much of the theory is clear (i.e., what math to use, what data, etc.), VAR implementation poses many practical problems. These practical problems not only reduce the quality of the risk view but also stymie the ability to use the theoretically preferable solutions. This article focuses on three such VAR challenges and some thoughts on what VAR users have done to address them.

Illustrated in Exhibit 1, the VAR implementation process is composed of five main components. We provide three vignettes to shed some light on the issues and some solutions to problems faced in the three main areas. The first vignette describes what determines whether decisions between mapping at the instrument level or at a broad index level are relevant. The second vignette describes data filling mechanics necessary to build a VAR matrix when data histories are not available or have gaps. The third vignette describes the role of backtesting as an indicator of how well a VAR model works.

Exhibit I



II. MAPPING

Theoretically, instrument-level mapping produces the most robust result. But add in computational intensity, data issues, and operational considerations for a large firm and one turns to a broader mapping framework for simplification. Whether into risk factors, buckets, or indices, however, mapping forfeits resolution. This vignette discusses ways to test the balance between ease and quality of result.

Open any textbook discussion on VAR, and you are likely to see a discussion regarding how a certain instrument can be mapped by individual cash flows or risk factors. Mapping evolved due to the large number of positions, the lack of the necessary instrument-level data, and other computational constraints that are the practical realities of VAR.

To manage these practical realities, trade-offs between theoretical accuracy and computational possibility must be made. In doing so, one must question the accuracy lost, for example, when mapping a bond position based on the bond's duration or maturity rather than on individual cash flows. The answer depends upon the type of position, its sensitivity to yield curve twists, and also its overall concentration in the portfolio. Further complicating matters is that as the size of the portfolio increases, so do the number of assumptions embedded in the calculation itself, which often overwhelms any simple mapping decision.

For several entities we performed simple balance sheet VAR calculations by taking annual report disclosed positions and mapping them into broad indices such as the Lehman Government Bond Index for U.S. Treasuries. Comparisons between these computed balance sheets VARs and those disclosed by the dealers were at times remarkably close. The accuracy depends both on the type of portfolio held and the instruments owned. For some portfolios, namely all long or all short with fairly linear positions and with a good degree of diversification, the balance sheet and disclosed VAR numbers compared well. So well in fact, one may question the usefulness of more sophisticated VAR mapping techniques. When portfolios include both long and short hedging positions, or contain a number of option-like contracts, or become more concentrated, these methodologies begin to diverge. Despite this divergence, there were still a number of firms whose calculations lined up. This could be due to a number of reasons:

1. coincidence
2. portfolio stability
3. law of large numbers
4. similar internal VAR calculation

Investigation offers insights into the origin of the risk and provides a healthy amount of skepticism as to how useful a single VAR number can be given all of the mapping assumptions made.

In order to illustrate the impact of several mapping techniques on the computation of VAR, we present a case applying widely accepted VAR methodologies to four hypothetical equity portfolios. For each portfolio we computed an equally weighted one day VAR at a 95 percent confidence level using daily returns from 3/31/96 to 3/31/98. The mapping techniques used are listed here in increasing order of both accuracy and computational intensity.

Mapping into the S&P index.
Here we bucketed each security into the single risk factor S&P index.

Mapping into the S&P with beta adjustment.
Here we mapped into the S&P index by using the RiskMetrics volatility estimate and then adjusted for each security's beta with the S&P.

Mapping into industry return buckets.
Here we bucketed each security into its industry index.

Mapping the instruments into themselves.
Here we used each stock's own historical returns.

The following portfolios were constructed in order to look at the impact of these four mapping methodologies to calculate VAR.

Portfolio 1

We begin by applying our four methodologies to a theoretically attractive equity portfolio. Portfolio I is a collection of linear, diversified, long positions in twelve Fortune 500 stocks, including Microsoft, Disney, Pfizer, GE, Coca-Cola, GM, DuPont, AT&T, American Express, P&G, Kodak, and Boeing. The net investment in Portfolio I is $1,000,000.[1] By applying our methodologies mentioned previously, we calculated the following I day VARs.

VAR Mapping Method	95% VAR
S&P mapping	$16,160
Beta	$22,694
Industry mapping	$16,482
Instrument level mapping	$17,804

We see that for our diversified portfolio, regardless of the mapping technique applied, all four VAR calculations compare quite well with each other. The results may lead us to question the need for robust VAR methodologies since we obtained similar results with more simplified mapping techniques.

Portfolio 2

Next we show how the VARs begin to diverge as we lose our diversification. We built a portfolio, again with long equity positions, but this time concentrated within the high-tech industry sector. Our equities included Hewlett Packard, Intel, Motorola, Compaq, Dell, Microsoft, Oracle, Gateway, Sun Microsystems, and IBM. The net investment in Portfolio 2 is again $1,000,000[2] By applying our methodologies, we have the following VAR results:

VAR Mapping Method	95% VAR
S&P mapping	$16,160
Beta	$24,205
Industry mapping	$27,554
Instrument level mapping	$31,639

We see that mapping into the S&P 500 index is no longer appropriate when the portfolio is concentrated in one industry sector. Our beta VAR calculation may be inappropriate as well. However, capital at risk calculated by mapping at the industry level and instrument level are somewhat more in line.

Portfolio 3

Next we show that mapping to indices is not necessarily appropriate when we have a mix of long and short positions. Portfolio 3 consists of the same twelve securities as Portfolio 1, but now we have six long positions and six short positions. The gross investment in Portfolio 3 is $1,000,000.[3] Our results from our four methodologies are as follows:

VAR Mapping Method	95% VAR
S&P mapping	$0
Beta	$1,480
Industry mapping	$4,824
Instrument level mapping	$6,665

Notice that in this case we have netted out risk exposures. In our portfolio, the greater our simplification, the greater the netting. S&P mapping, Beta, and industry mapping all indicate less capital at risk than at the instrument level. To the extent that this type of netting is not justified, or if the gross concentration in one bucket becomes too great, one may be understating one's risk considerably, as illustrated above.

Portfolio 4

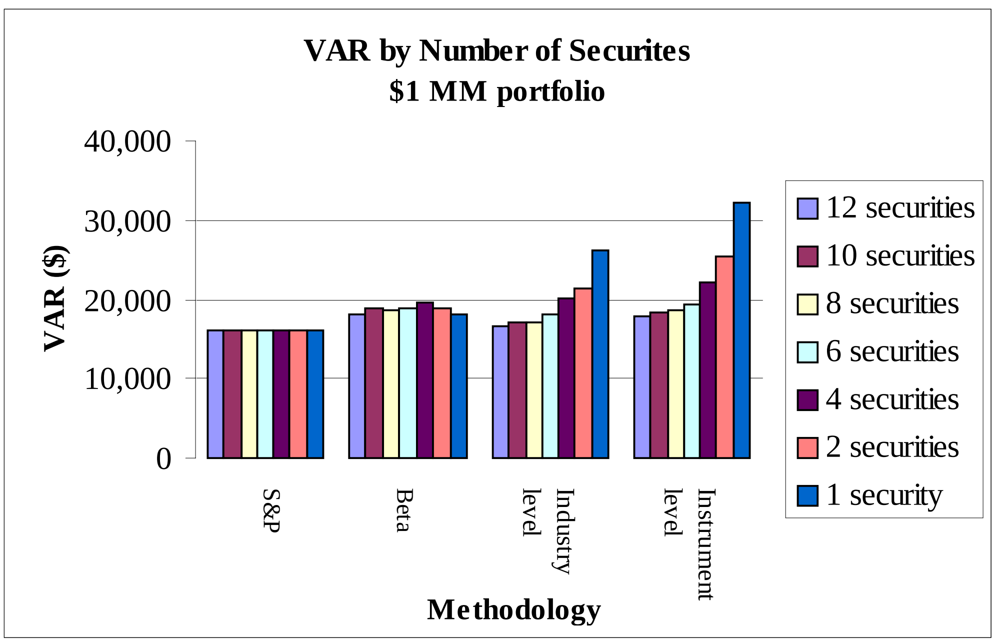
Next we move even further from our preferred world of Portfolio 1 as we look at the combined effect of both increased concentration and a mix of long/short positions. Portfolio 4 consists of the same high tech securities as Portfolio 2. This time, however, we included both long and short positions. The gross investment in Portfolio 4 is $1,000,000.[4] The following results suggest an even greater divergence of VAR:

VAR Mapping Method	95% VAR
S&P mapping	$0
Beta	$2,391
Industry mapping	$0
Instrument level mapping	$8,215

Here the VAR analysis reveals significantly different risk profiles. As in Portfolio 3 we have netted out risk and undervalued our true exposures due to the longs and shorts in not only the S&P mapping VAR, but also in the industry level mapping VAR. The instrument level mapping clearly reveals much more capital at risk than the other methods.

The effects of portfolio diversification on the VAR mapping choice. Finally, we looked at the impact of reducing the number of securities within a portfolio. This is essentially helping us answer the question of how many diversified securities are required in order to use index mapping safely. Let's look again at Portfolio 1, the portfolio of linear diversified long positions. We began with 12 securities as discussed earlier. We then removed two at a time to create portfolios

Exhibit 2



with 10, 8, 6, 4, 2, and 1 security. Each portfolio maintains a $1,000,000 net investment.

The graph in Exhibit 2 displays our results. As the instrument level VAR indicates, using six or more diversified positions yields good results when mapping into the S&P index bucket. Having four or fewer positions leads to a substantial understatement of the risk using the S&P index mapping method. Please note that the example only applies to positions in the same asset class, namely, equities. Further testing should be performed to assess whether other asset classes are similar or to what extent this pattern exists across asset classes.

VAR mapping choices when hedging is involved. When hedge positions exist, the mapping of the hedge must be consistent with the mapping of the hedged instrument in order for the hedge to benefit the portfolio risk in a VAR sense. For example, a short equity total return swap hedging a long S&P portfolio must be mapped such that the hedge offsets the risk as appropriate.[5] On the other hand, caution must be taken as many hedges are not perfect. If the VAR assumes they are, substantial leveraged risk on a correlation assumption can be disguised. For example, a proxy hedge of an equity portfolio with a short S&P future may look like most of the risk is hedged when looking at past correlations. But as the long portfolio and the short hedge grow substantially in size, the dependence on that correlation assumption increases dramatically. The VAR implementation and mapping decisions can hide risks when they mirror the assumptions of the risk takers themselves. Proxy hedged portfolios are extremely hard to abstract beyond position and individual cash flow mapping. In many ways, the presence of hedging is mimicked in Portfolios 3 and 4 with the long and short positions. The problems are similar as well.

The above discussion of how certain hedges can make risk disappear as far as the VAR is concerned brings to mind the potential of VAR manipulation. In any organization, regardless of the VAR mapping technique chosen, one must consider the types of controls necessary to guard against the gaming of the VAR metric. For instance, if S&P mapping or beta adjusted S&P mapping is used, the risk as seen by the VAR model can completely disappear by hedging the S&P. Basis risk between the positions and the index remains but may be hidden from the VAR model. Traders or other position takers can predict the VAR models results and make position changes to disguise the risks.

As this vignette has demonstrated, one needs to carefully consider the mapping technique used by balancing ease with quality of result. For example, Riskmetrics maps all equity positions into the S&P index with a beta adjustment. As illustrated earlier, for a well-diversified, linear portfolio this method may provide an appropriate calculation of VAR. Often, however, it may not. One needs to understand the implications of different mapping techniques and when they can be applied safely. While one technique may be appropriate for a given portfolio,[6] another method may be dangerous.

III. DATA FILLING

One of the first stumbling blocks when calculating VAR is the occurrence of data gaps. Fortunately, techniques exist to overcome this hurdle; however, most classical methods require more history than is typically available (examples are an emerging market security that has been around for one month, or a once-per-month data series on an instrument from a data base, with no sources for the intervening dates). This vignette discusses practical approaches for addressing data gaps and constructing covariance matrices starting from incomplete data series.

Dealing with data gaps is inevitable. Causes for gaps include market holidays, lack of market liquidity, or insufficient data because an instrument or market is too new to have historical data. Data gaps can occur at the beginning, end, or middle of the required series.

In order to compute a firm-wide VAR, such data issues will need to be addressed. For firms who map their positions directly into themselves, the existence of historical data is often a problem since the actively traded positions are in the newer instruments. For instance, on-the-run Treasury bonds are highly liquid, but little price history exists due to their age. Alternatively, short-term fixed income instruments such as Treasury bills have little history due to their short maturity. Most practitioners solve these problems by using yield indices that exist independent of individual instruments. Examples are CMTs for Treasuries and credit yield indices for corporate bonds. The use of yield indices also avoids another fundamental problem of self mapping: Many instruments change over time and exhibit different risk characteristics. As a result, their history may not be reflective of their future risk. For instance, a five year bond behaves more like a three year bond after it is two years old. Using the past two years of its price performance may overstate its riskiness.

Using yield-based indices is often a workable and elegant solution to these problems. One must take account of the fact that the index is often not the same as the position being mapped. For instance, the five year CMT reflects the volatility of a five year Treasury, but not necessarily the volatility of a six year bond. Many practitioners will adjust the dollar amount mapped into the five year bucket by the ratio of the position's duration to that of the bucket. In this way the overall volatility is increased or decreased as appropriate while maintaining the correlations; for instance, the six year bond could be mapped into the five year bucket with a duration adjustment to reflect the increased risk.[7]

The simplest solution to data gaps would be to drop out any dates for which any of the data series did not have a data point. Unfortunately, with large matrices this solution very quickly leads to having a data set of too few data points. Unless the data gap issue can be dealt with, a fundamental trade-off exists between using more data series, which often corresponds to a more finely tuned VAR, and having enough data history.[8]

When dropping points or using an index series is not possible, one must employ statistical data filling mechanisms. Statistical techniques have both pros and cons. In the following table we discuss three common techniques and the implications of each. Note that other techniques exist.

	Method	*Pros*	*Cons*
1.	**EM algorithm.**	Provides maximum likelihood estimates of returns.	Potentially nonpositive definite covariance matrix; may be misleading for small number of risk factors or multicollinearity in returns; subject to bias if missing data is not random; computationally intensive.
2.	**Substitute Return Selection.**	Works best when missing data has strongly correlated series to draw from.	Does not necessarily capture correlation contribution to VAR as described in this article.
3.	**Pairwise Data Calculation.**	Easiest implementation.	Potentially nonpositive definite covariance matrix; may be misleading for small number of pairwise common dates.

RiskMetrics Expectation Maximization (EM) algorithm. This data filling technique involves using an iterative algorithm to fill data gaps in a manner that guarantees the self-consistency of the data set. Specifically, the data holes are filled with the average values that one would expect for them, given the *known* data. Then the conditional means, volatilities, and correlations of the newly generated data are recomputed and the process is iterated until the statistics of old and new data match. This approach is powerful in that data gaps are filled sufficiently to enable construction of a covariance matrix of the gap-filled series. Fairly robust, the technique has potential statistical flaws,[9] but its greatest drawback is the computational time required for any matrix of significant size. While JP Morgan indicates that they run it on as many as 50 data series at a time, other firms may find that running the EM algorithm on 50 data series often exceeds the computational firepower available to them.

Substitute return selection. A second common technique is substitute return selection. This method involves selecting a data series that behaves similarly to series with the missing data and making a substitution where data points are missing.[10] For instance, a missing return for a BB corporate bond would be filled by the most similar history available. If the most similar history is that of a BB+ corporate bond, this method would substitute the daily return of the BB+ for the missing BB return. While "worst case" substitution might use a Treasury return of a similar tenor, care must be taken to ensure that excess data gaps and too few similar indices for data filling do not cause spurious results. From a risk oversight perspective, data fills should be signed off by someone officially, but the process could be automated in part by allowing a computer to find the most statistically correlated (or anti-correlated) series to substitute. Note that this method is similar in spirit to the Riskmetrics EM algorithm, but is not as computationally intensive.

Pairwise data calculations. A third method involves calculating covariances based on the *mutually* available data between any two data series. Unlike the other two data filling methods, no exogenous data needs to be inserted or inferred. To illustrate the method, assume we have two one-year (250-day) data series, A and B. A has a complete set of data, but B is missing 20 data points at the beginning. The pairwise method calls for calculating variances and covariances with the data available. Therefore, the variance for A will use all 250 data points, while the variance for B will use the available 230 data points.[11] The covariance between A and B will use only returns on the 230 dates common to both series. The 20 data points that A has, but B doesn't, are ignored for the covariance calculation. Since this methodology is nothing more than the standard covariance calculation with selective exclusion of data points, extra computational time is not required.

By using all of the available data without relying on statistically generated data, the accuracy of each *individual* matrix element may be improved. However, this practice may lead to spurious *aggregate* results, such as a nonpositive definite covariance matrix,[12] potentially exhibiting variances less than zero or implied correlations[13] greater than 1. Consider the example discussed above: The covariance and the variance of series B were calculated using 230 data points, while the variance of series A was calculated using 250 data points. To see how the correlation could be greater than 1, assume that series A and B are identical for the 230 data points that they have in common. If the volatility in the 20 extra data points in series A is lower than the volatility in the other 230 data points, the variance of A will be smaller than that of the subset of A used to calculate the covariance. As a result, the implied correlation can be larger than one.[14]

When filling data gaps using methods similar to the Pairwise Data method illustrated above, testing must be performed to guard against nonpositive definite matrices. A major benefit of this methodology is improved computational time, but computational efficiency should not be given priority over having a well-behaved matrix.

Example. To illustrate the three methods, we consider an equally weighted portfolio of five currencies (JPY, GBP, FRF, ITL, IDR) and compute a variance covariance VAR. Assume we have one year's worth of data for the five exchange rates, as well as data for 10 more risk factors[15] (Thai baht, Australian dollar, Philippine peso; I month LIBOR rates for the U.S. dollar, Japanese yen, Italian lira and British pound; the S&P 500; the Dow Jones industrial average; and the USD five year swap rate). In our example, we deliberately create data gaps in the data histories of the five currencies and use each of the above methods to fill in those gaps. We then compute a VAR for each of the portfolios of filled-in returns. We then compare the resultant VARs to the actual VAR computed without data gaps.

The data gap scenarios which we consider are: (1) sporadic gaps in the data, (2) a six-month gap at the beginning of the year, and (3) the combination of the above gaps. The results are summarized in the following table.

Percent Deviation of Data-gap Computed VAR from Nondata-gap Computed VAR

	Data Filling Method	*Scenario I* (Sporadic gaps)	*Scenario 2* (6-month gap)	*Scenario 3* (Combined gaps)
1	EM algorithm	***	3.1%	***
2	Substitute Return Selection	-1.5%	0.2%	-0.9%
3	Pairwise Data Calculation	-0.2%	-0.3%	-0.5%

*** Indicates that 15 risk factors[16] were insufficient to resolve problems with multicollinearity which manifested itself as a nonpositive definite covariance matrix (leading to such spurious results as divergent correlations or negative variances). The addition of five more risk factors also did not resolve these problems. The continued addition of further risk factors would eventually overcome the multicol linearity problem, but would impose a computational burden.

As the above chart indicates, the Pairwise Data method provides surprisingly consistent and accurate results for our examples and offers a viable alternative to firms seeking an operational shortcut to handling data gaps, despite its lack of mathematical rigor. Please note that the above results were provided for illustrative purposes only. Further testing with other data series and other types of data gaps would be required to generalize this conclusion.

Combining weekly and daily data. Even if data gaps can be resolved, another type of data problem arises from simultaneously using datasets that have different data frequencies, (e.g., weekly versus daily data). Working with disparate frequency datasets is often necessary due to lack of daily data in many instruments. Another pitfall is that what appears to be daily data is in fact weekly data in disguise; one notices this magic act when daily prices stay unnaturally constant over set intervals (dampening the volatility[17]). This often occurs for illiquid instruments.

One solution to disparate time series is to convert all of the more frequent data to the less frequent time scale. This is very straightforward. For instance, daily returns can be converted to weekly returns by simply compounding the five daily returns in the week. This solution is rarely workable, however, as often not enough data points exist to merit converting disparate data frequencies to the longer frequency basis. For example, if there is only one year of data, there are only 50 weekly returns with which to work.

On the other hand, rather than converting all of the data series to weekly, one can compute variances and covariances using weekly data when one or both of the data series being combined are weekly, but use daily data series for those calculations involving only daily datasets. Combining weekly and daily data is accomplished by compounding the daily returns over the preceding five days to obtain the weekly return for that week as described.[18] The resulting variances and covariances for these pairs are divided by five to account for the frequency of data used.[19]

The following diagram illustrates how to combine weekly and daily data in a matrix. RB_1 through RB_{12} illustrate Risk Buckets. Covariances between Risk Buckets which both have daily data are denoted "D," while covariances between Risk Buckets in which at least one has weekly data are shaded and denoted "W." The process to create a covariance matrix uses daily data for all variances and covariances between groups that both have daily data, while only using weekly covariances between groups for which at least one of the groups has weekly data. In the illustration, RB_5 and RB_6 both have weekly data, so the columns and rows of covariances in which they are included use weekly data.

	RB_1	RB_2	RB_3	RB_4	RB_5	RB_6	RB_7	RB_8	RB_9	RB_{10}	RB_{11}	RB_{12}
RB_1	D	D	D	D	W	W	D	D	D	D	D	D
RB_2	D	D	D	D	W	W	D	D	D	D	D	D
RB_3	D	D	D	D	W	W	D	D	D	D	D	D
RB_4	D	D	D	D	W	W	D	D	D	D	D	D
RB_5	W	W	W	W	W	W	W	W	W	W	W	W
RB_6	W	W	W	W	W	W	W	W	W	W	W	W
RB_7	D	D	D	D	W	W	D	D	D	D	D	D
RB_8	D	D	D	D	W	W	D	D	D	D	D	D
RB_9	D	D	D	D	W	W	D	D	D	D	D	D
RB_{10}	D	D	D	D	W	W	D	D	D	D	D	D
RB_{11}	D	D	D	D	W	W	D	D	D	D	D	D
RB_{12}	D	D	D	D	W	W	D	D	D	D	D	D

Note that this method for dealing with disparate data frequencies suffers from the same theoretical nonpositive definite problems discussed in the pairwise method. Care must be taken when using these methods so that logical impossibilities, such as a negative VAR and implied correlations greater than 1, are not allowed for any portfolio.

IV. BACKTESTING VAR

To meet regulatory VAR disclosure requirements, it is not surprising that some firms threw up their hands and performed VAR on their revenues. This vignette will look at the backtesting challenges ahead after the first public disclosures of VAR.

Backtesting will be a widely debated topic this year as many firms plan to include it in their regulatory disclosure. The goal is to provide an assessment of how well your VAR measure has done its job of measuring risk. This allows one to answer the all-important question of whether, when you take more risk, you in fact achieve more return.

Backtesting is the key to validating the use of VAR. As such, the regulators now require backtesting. Essentially what one does in a backtest is to look at the profit and loss over a historical period and validate that it is consistent with what the VAR model would have predicted. Backtesting is theoretically elegant, simple to understand, and a powerful argument to support a VAR model's accuracy. Unfortunately, it is also difficult to perform.

One simple backtest is to determine the percentage of time that losses exceed the VAR estimate. This percentage should correspond to approximately 100% minus the VAR confidence level selected. For example, if one is performing a 95% confidence VAR, one would expect to have losses that exceed the VAR number 5% of the time.

Consider the following table based on two portfolio VAR backtests over the three year period from 1995 to 1997. The first portfolio (S&P) consists simply of the S&P 500 index. The second portfolio (EM) is composed of a basket of emerging market currencies. The VARs were computed using both equally weighted data and also exponentially[20] weighted data at a 95% confidence level. Hence, one would expect 5% of the losses to exceed the 95% VAR amount.

Backtest 1 Results: A *positive indicator of the VAR method working for a particular portfolio is 5% of the returns falling below the 95% VAR amount.*

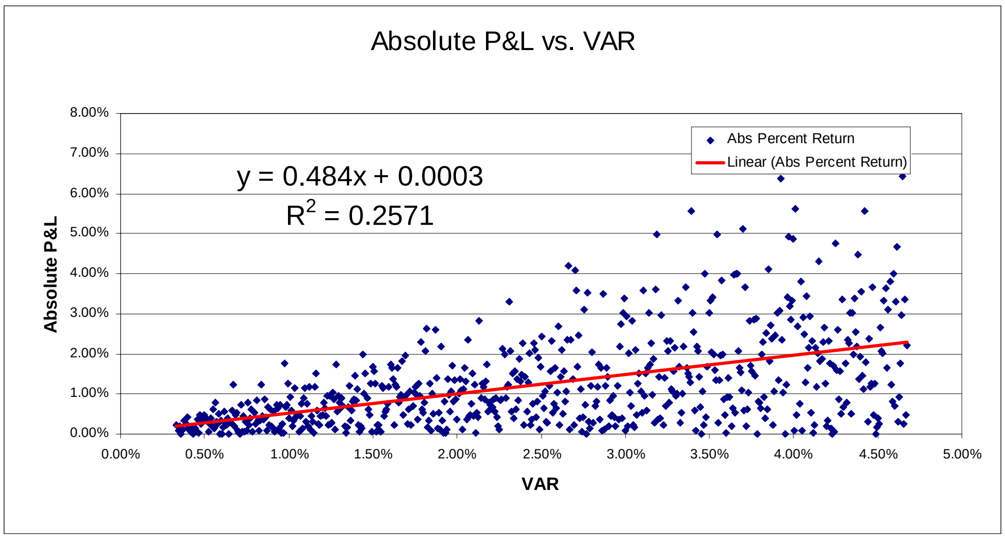
Percent of Returns below 95% VAR

VAR Method / Portfolio	Equally Weighted VAR	Exponentially Weighted VAR
S&P Portfolio	6%	5%
EM Portfolio	3%	5%

As the table indicates, the exponentially weighted VAR appears to work better than the equally weighted VAR for the two sample portfolios chosen. Please note that these results may not hold for your particular portfolio. Performing a backtest using your own VAR model on portfolios that you hold is the one way of validating its ability vis-a-vis the types of instruments held. For instance, the exponential VAR appears to work better than the equally weighted VAR for the above portfolios, but the opposite may be true for a different set of portfolios. For our two chosen portfolios, the equally weighted VAR model should be reevaluated given its poor performance, underestimating VAR in one case and overestimating in the other.

A more advanced backtest incorporates risk return relationships. This backtest checks to see if in fact the expected absolute value of the profits and losses (P&L) is higher when the risk is higher. Specifically, we plot the absolute value of each daily P&L versus the previous day's VAR estimate to determine whether the VAR does in fact predict on average the size of the P&L swings. Exhibit 3 shows what one might expect when plotting risk (VAR) versus absolute P&L. Please note that the slope of the least squares regression line 0.484 is very

Exhibit 3



close to 0.485, which is the theoretical slope that should result[21] if VAR is to serve as a good predictor of P&L volatility.

One can assign error bars around this slope given a certain number of data points in the regression. Using over 600 data points as in the graph in Exhibit 3, we would expect the slope to fall within the range of 0.481 to 0.489.[22]

To see how well this backtest performs on the two portfolios presented earlier for both equally and exponentially weighted data, we produce similar scatter plots for each. Please note the slopes for each portfolio consolidated in the following table.

Backtest 2 Results: A *risk vs. return regression slope in the range of 0.481 to 0.489 is a positive indicator of the VAR method working for a particular porffolio.*

Risk vs. Return Regression Slope

VAR Method / Portfolio	Equally Weighted VAR	Exponentially Weighted VAR
S&P Portfolio	**0.6789**	**0.3424**
EM Portfolio	**0.0410**	**0.2499**

Notice that none of the sample slopes falls within the expected range. This indicates that neither of the two VAR methods is working particularly well for these two portfolios. Please note, however, that the above backtest confirms one of the first backtest's results, namely, that the exponentially weighted VARs perform better relative to the equally weighted VARs for both portfolios. But unlike the first backtest, the second backtest does not give an "all clear" rating to the exponential VAR for these portfolios. Some improvements are definitely possible.

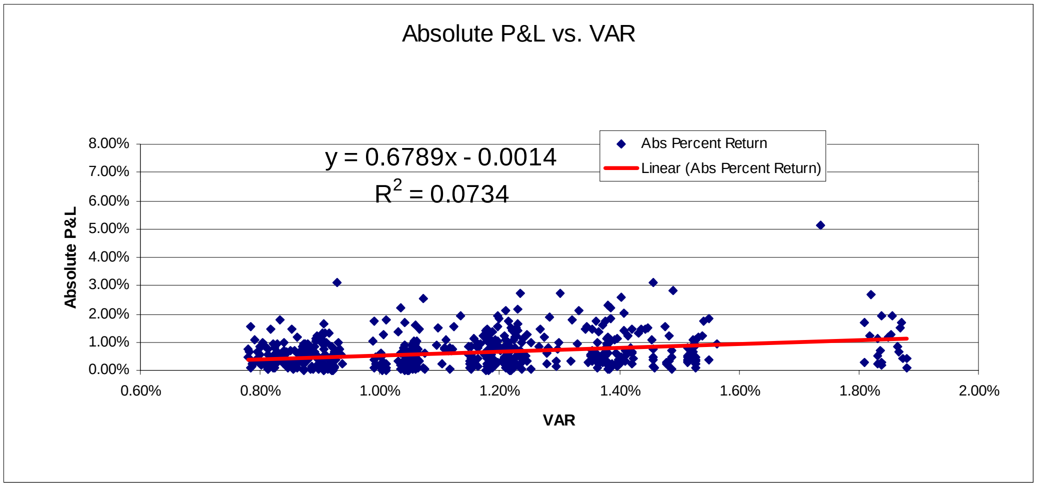
In Exhibits 4, 5, 6, and 7 we include the actual risk versus return regression graphs for each of the above four results in the table. The graphs provide an interesting picture of what is at times behind the poor results. Two reasons are readily apparent:

1. The nonuniform distribution of results across the VAR axis.
2. The presence of outliers.

Next we discuss these two points in more detail using the graphs as examples.

The S&P exponentially weighted graph in Exhibit 5 illustrates the first point very clearly. Notice how there are many, many observations of low VARs, but relatively few observations of large VARs. This nonuniformity of results has a large effect on the slope, in that there are not enough observations at the higher

Exhibit 4
S&P Equally Weighted



VAR levels to arrive at a statistical estimate of the absolute P&L during a high VAR day.

Both of the emerging market graphs in Exhibits 6 and 7 illustrate the second problem of these charts, namely, the presence of outliers. Please note especially on the equally weighted graph the presence of outliers at the lower VAR amounts. These outliers decrease the slope and hence cause it to perform the worst of all four results.

Exhibit 5
S&P Exponentially Weighted

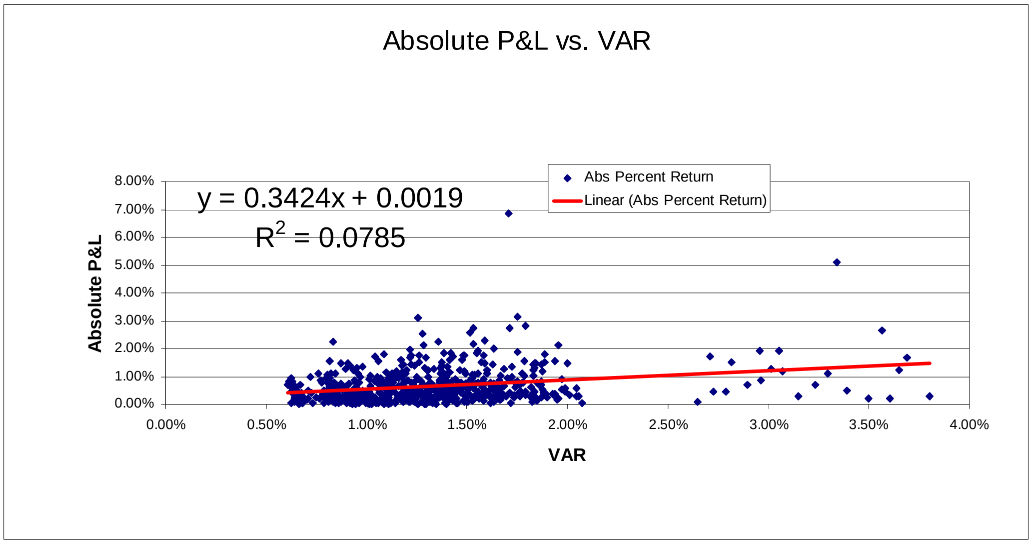
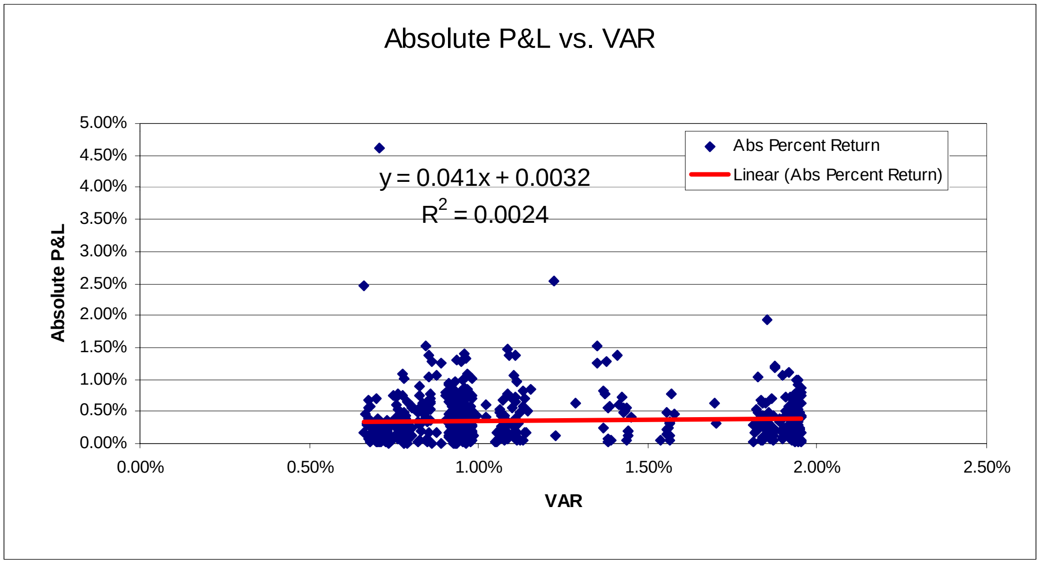
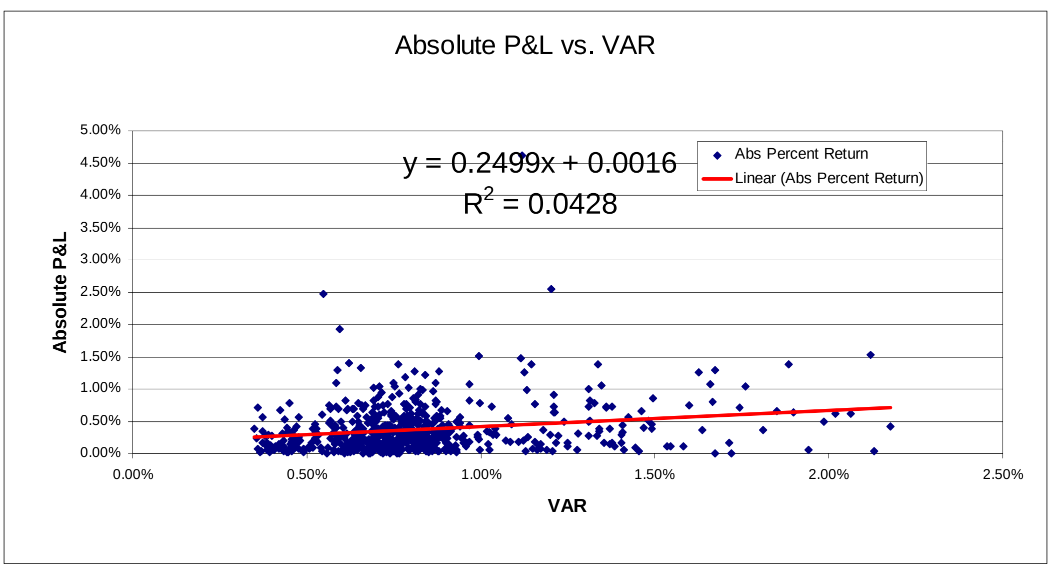


Exhibit 6
Emerging Markets Equally Weighted



Clearly further theoretical work needs to be done on the issue of backtesting. Unfortunately, theory is only part of it; the practical reality often overwhelms the issues raised thus far. Backtesting, simple in concept, has practical realities that are not simple. The previous tables and graphs were computed on idealized portfolios under ideal conditions. The requirements for an ideal portfolio are:

Exhibit 7
Emerging Markets Exponentially Weighted



1. The portfolio is static, i.e., there are no buys or sells.
2. The VAR is calculated as frequently as the VAR horizon.
3. The P&L data captured utilizes real market prices.

The static portfolio requirement is unrealistic for a casual retail investor, let alone a sophisticated financial institution. Unfortunately, VAR is only a snapshot of the risk in the portfolio over a specified period of time, and if instruments were bought or sold during this period, the P&L data would no longer correspond to the VAR horizon. For instance, a one day VAR assumes that the portfolio is held constant over a 24-hour period. But the P&L is due not only to the position captured in the VAR snapshot, but also to the positions that are traded in and out of the portfolio during the following day. There is a P&L effect occuring ing that never even makes it into the VAR calculation. Therefore, the portfolio must remain constant to perform a valid backtest. Fortunately, if buys and sales during the day can be reversed out, the P&L that would have occurred had the portfolio remained constant can be synthetically produced. For some firms, however, the pricing and P&L data is not captured in such a way as to enable P&L normalization. As a result, backtesting becomes a daunting task. There are solutions, some more satisfactory than others, but they depend on the type of VAR performed.

The second requirement is to perform VAR calculations or snapshots at least as frequently as one's VAR horizon. If the VAR calculated is a daily VAR, one must be able to calculate their VAR daily in order to ensure that each daily P&L observation is compared to its corresponding daily VAR. Each daily VAR only describes that day's P&L. Therefore, for each P&L observation there needs to be a corresponding daily VAR.

The third requirement is to use real market prices in one's P&L calculation. This requirement is easily satisfied for many firms who hold highly liquid Treasury or equity securities. For other market participants who hold less liquid instruments such as nonbenchmark emerging market positions, residual mortgage instruments, or thinly traded high yield issues, real market prices are not as easily observable. Consider the historical price series of some of the instruments with prices that often do not move. The lack of movement is often due to illiquidity rather than an indication of a lack of market risk. Thus, the P&L observed on these instruments will be lower than that predicted by VAR.

Some firms fall victim to the third requirement even though they have liquid positions. This is often due to internal reserving policies. There are many different ways in which reserves are taken on trading positions: some reserves are explicitly held out, while others are implicit in their prices marked on the books. Problems arise in the implicit reserves since they impact the P&L calculations. One example of an implicit reserving technique is the conservative marking of positions to the lower of cost or market. While this methodology conservatively delays the recognition of P&L until the P&L is realized, it clouds the real P&L needed to perform a satisfactory backtest. Therefore, the P&L that one observes from these prices can be much less than that predicted by VAR.

The third requirement highlights one of many errors that could be made in the VAR calculation process, namely, using internal pricing for the covariance calculations that have unusually low volatility due to either illiquidity or implicit reserving. This practice can artificially generate lower VAR estimates than appropriate. It is imperative to use liquid data in the VAR generation process, lest the VAR seriously underestimate the risk inherent in a portfolio.

The requirement to perform backtesting is certain to push the limits of many a firm's systems, just as VAR had done previously. As a result, it is little wonder that many firms are opting to use a VAR approach that facilitates backtesting right from the start. Like VAR, backtesting is a very simple and elegant calculation. It is only the practical reality that destroys its theoretical beauty.

NOTES

1. Each individual position has a market value of $83,333.

2. Each individual position has a market value of $100,000.

3. Microsoft, Pfizer, Coca-Cola, DuPont, American Express, and Kodak each had a long market value of $83,333, while Disney, GE, GM, AT&T, P&G, and Boeing each had a short market value of -$83,333.

4. Hewlett Packard, Motorola, Dell, Oracle, and Sun Microsystems each had a long market value of $100,000, while Intel, Compaq, Microsoft, Gateway, and IBM each had a short market value of -$100,000.

5. Whether all of the risk is offset or whether there is some basis risk remaining depends on the terms of the swap. Care should be taken to ensure that the VAR accurately reflects whether all or only part of the risk is hedged.

6. Note that the appropriate method may change as the portfolio changes its mix over time.

7. Another popular method is to map the position into a combination of buckets rather than to just the closest bucket. For instance, the six year bond above could be partly mapped into the five year and seven year buckets instead of just one. This method will not only reflect the volatility differences of the position and the mapped bucket, but also take

account of some differences in correlation between the two buckets it is mapped into.

8. The amount of data needed to measure VAR is debatable. While most methodologies call for a minimum of one year's worth of data, exponentially weighted methodologies such as Riskmetrics actually rely on much less.

9. The EM algorithm suffers if the timing of the missing data points is correlated in some way to the data itself. For instance, if at times of high market volatility data is routinely missing, biases could creep into the data. In addition, the algorithm requires nonmulticollinearity of the time series at not only the initial stage but at every iteration. Nonmulti collinearity is required as the intertemporal covariance matrix must be invertible to calculate each iteration. Adding additional data series can alleviate this problem for the most part, but computational time is increased dramatically.

10. The data points could be weighted by the ratio of the two series volatilities to keep the volatility ratio constant.

11. 250 - 20 data points available for B.

12. A positive definite matrix is one whose eigenvalues are all greater than zero. Without this property, one would be able to construct a portfolio whose return has negative variance or imaginary volatility, a market impossibility.

13. The implied correlation between two data series in a covariance matrix is calculated by dividing the covariance between two series by the respective standard deviations of each.

14. Implied correlation= $\sigma_{AB}/\sigma_A \sigma_B$. As discussed in the text, $\sigma_{AB} = \sigma_B^2$, and σ_A^2 could be less than σ_{AB} due to the extra 20 data points. This combination would imply that the correlation is greater than 1 as per the equation.

15. The extra data series help improve the accuracy of both the Substitute Return Selection method and the EM algorithm since the potential for more similarly correlated data series can improve accuracy. In the case of the EM algorithm, extra data series are *required* due to the potential for a noninvertible covariance matrix as discussed in Note 9.

16. The five portfolio histories plus the extra 10 risk factors.

17. Please refer to the vignette on backtesting for a discussion regarding illiquid security pricing and its effect on VAR and backtesting

18. Adding the logarithms $\log(p_{T+1}/p_T)$, where p_T is the price on day T.

19. Under the assumption of a normal distribution of returns, variance or covariance computed from weekly returns is equal to the daily variance or covariance of returns multiplied by five, the number of days in the week, due to linearity in variance of returns.

20. A decay factor of 0.94 was used.

21. The theoretical slope is calculated by dividing the distribution mean of the absolute value of the P&L in terms of standard deviations by 1.65 (the number of standard deviations corresponding to a 95% VAR).

22. Error range given a 95% confidence level, computed using: $t_{n-1}(\alpha/2)$ / sqrt(n) where n equals the number of data points, t is the student t-distribution, and α is 1 minus the confidence interval.