Combining Non-Constant Weights with Historical Simulation VaR

Riccardo Rebonato[*+^] and Vasant Shanbhogue[*]
[*]Quantitative Analytics Group, Royal Bank of Scotland
[+]Oxford University - OCIAM
[^] Imperial College - Tanaka Business School

November 13, 2007

Abstract

We show how the historical simulation method can be extended to deal with time series that displays different importance weights. The method is shown to be particularly useful to deal with seasonality. We build on the work by Hull and White (1998) about volatility rescaling, and we show some subtle points that arise in the presence of non-constant weights aross time series.

1 Introduction and Motivation

Given a set of time series of risk factors, Historical Simulation (HS) is one of the best-known and most widely used methods to obtain an empirical distribution of profits and losses (P&Ls), from which various risk statistics can be estimated, e.g. Value at Risk (VaR), Conditional Expected Shortfall, and others. Of these risk statistics, VaR is particularly important because it is used to calculate the market risk component of regulatory capital under the Basel I and II regulatory frameworks.

The pros and cons of HS have been discussed at length (see, e.g. Lambadiaris et al (2003) and Jorion (2006)) and will not be repeated here. For the purpose of the present discussion, We will simply recall that it is (almost) non-parametric in nature and that it is capable, at least in principle, of capturing forms of co-dependence richer than simple correlation even when the number of underlying risk factors is high. In reality, there is one extremely important parameter in standard implementations of HS: the length of the data window. All synchronous vectors of market risk changes are implied to have exactly the same weight (i.e. to be exactly equally relevant to today's market conditions), and all data past the window cut-off are deemed to be totally irrelevant. Indeed, supposed difficulties in assigning non-constant importance weights to different 'days' in the data set is often quoted as one of the greatest shortcomings of the HS technique.

We show in this paper that dealing with non-constant weights which are identical across all the time series is trivial (various alternative schemes are for instance discussed in Hull and White (1998)). We show that it is also conceptually easy and computationally straightforward to deal with the more challenging case where different weights have to be assigned to different data series. This situation is of particular relevance when some of the time series display seasonality, as is the case of specific commodities such as natural gas, heating oil, and gasoline. In this latter case there are, however, some subtleties that we explore in Section 6.

The paper is organized as follows: in Section 2 we show in what contexts this problem is of importance; in Section 3 we define the notation and set out the precise terms of the problem; in Section 4 we give an intuitive discussion of the determination of joint weights; in Section 5 we present the algorithm that produces a HS estimation of the profit-and-loss distribution arising from a number of market risk factors when different importance weights are given to different groups of risk factors; Section 6 presents a simple example that brings to light some of the subtler points in calculating joint statistics with non-constant weights; in Section 7 we show using a realistic example that the proposed algorithm indeed delivers what it is designed to and we present some numerical results that show how the method works in practice; Section 8 concludes the paper.

2 The Importance of Different Weighting Schemes

In general, different weights can be applied to the elements of a time series under at least two distinct conditions.

In the first case the underlying phenomenon is assumed to be time-stationary, but data collected at different points in time are deemed to be reliable to different extents because of non-constant experimental precision in their collection. The second situation arises when the precision with which each data point has been collected is constant throughout the series, but the statistical properties of the underlying phenomenon are thought to be non-constant over time scales comparable with the length of the time series. A combination of the two situations is, of course, possible.

In risk management the situation that is by far more common is the second. Heteroskedasticity in the underlying phenomenon poses in general challenges to statistical estimation. There are some conditions, however, that lend themselves well to handling by a suitable weighting scheme.

The first applies when the changes in 'regime' are relatively slow compared with the 'duration' (precisely defined below) of the weighted data. One example could be a gradual change in risk premia for risky asset classes that takes place over many months or even years because of a change in macro-economic conditions (e.g. change in global savings patterns). A weighting scheme that gave more importance to the more recent months can easily handle this situation. However, even rapid changes in market regime can sometimes be handled by an adaptive weighting scheme (albeit less successfully), as long as the change in regime can be detected rapidly, and a reliable procedure exists to assign suitable weights to 'patches' of past relevant history.

Another type of heteroskedasticity that can be handled well, and arguably better, by a weighting scheme is when the change over time of the joint statistical properties of the time series follows a deterministic pattern. Seasonality is an obvious example, of great relevance to some commodities. In the case of many energy commodities, for instance, consumption and production through a calendar year show a predictable pattern. These supply/demand interactions lead to an expected relative price level pattern through a calendar year. The differing extent of supply/demand interactions across seasons can also lead to differing volatilities across seasons. Natural gas in the United States provides a good example. This commodity is extracted from the ground fairly evenly throughout a calendar year. However, consumption has typically been dominated by the winter season in which natural gas is burned as a fuel for heating. Further, natural gas is increasingly being used as a fuel in power plants. As a consequence, the consumption shows a peak both in the summer and during the winter season. Periods of increased consumption are also periods of tightness in supply, and this leads to higher variability of prices in these periods.

We should point out, however, that not all commodities exhibit seasonal tendencies. For example, crude oil does not show seasonality, but it does show non-zero correlation with gas. To handle both crude oil and gas the HS methodology must be able to capture this correlation even when one commodity is seasonal and another is not.

This broad classification can shed some light on the weighting schemes of relevance in market risk management applications. For instance, when the time series of interest are volatilities, the well-documented phenomenon of volatility clustering (often described using GARCH-like techniques - see, eg, Campbell and Lo (1997) and references therein) calls for a rapidly decaying weighting scheme. Models of the market dynamics based on a classification of Markov states (say, using a Viterbi algorithm in a Bayesian framework - see, eg, Bilmes (2006), Cappe et al, (2005)) introduce non-constant weights that try to identify the probability that a past data point belongs to (and is therefore relevant to) the current state.

In all non-trivial cases, the application of a weighting scheme implies the discarding (complete or partial) of some of the available raw data. A measure of how much data has been retained by a scheme is the duration, d, of the weighted data. If the normalization of the weights is chosen so that the highest weight is 1, the duration (in years) is defined as:

$$d \equiv \frac{1}{N_y} \sum_{i=1}^{N_d} w_i$$

where w_i are the weights, N_d the total number of points in the data series and N_y is the number of days per year. This quantity is of obvious intrinsic interest, and is also important because many banking regulators impose conditions for acceptance of internal model approval based not only on the minimum length of the data series (the number N_d), but also on the duration of the weighting scheme.

Fig 1 shows different weighting schemes that make use of two or three years of data, but all have the same duration of 0.5 years. With the exponentially-decaying seasonal schemes depicted in Fig 1 it is assumed that, conditional on the current date being, say, January 1^{st} 2008, data points collected in December, November and October 2007 are relevant but to a smaller and smaller extent; data from the end of September back to the beginning of April 2007 ('summer' data[1]) are conditionally fully irrelevant; data from March 2007 back to end of September 2006 are relevant again, with a local peak of relevance for January 1st, 2007. The periodic continuous line and the periodic curve with crosses show the trade-offs between two weighting schemes with the same duration, one that takes into account more distant data (up to January 1^{st}, 2005), the other with a two-year cut-off .

Fig 1 approximately here

Fig 1: Several weighting schemes with the same duration of 0.5 years: constant weights, exponentially decaying weights, periodic exponentially decaying weights making use of two or three years' worth of data.

In estimating the distribution of P&Ls using HS in the presence of non-constant weights it is useful to distinguish two situations (Cases I and II below).

Case I: *Same weights for all the time series.* This can correspond, for instance, to the simple case of decaying relevance for all time series at an identical decay rate. But the same weights across time series can occur even in the presence of a richer 'structure' than a simple decaying function. For instance, if, as mentioned above, the state of the market on each day is classified using Bayesian techniques (eg, Hidden Markov Model as in Rebonato and Chen (2007)) as belonging with a certain probability to a set of possible states (eg, 'excited', 'normal' and 'quiet'), a weighting scheme common to all the time series can be used to reflect this classification. See, eg, Schaller and Norden (1997). In either case, the important common feature of this class of applications is that, however simple or complex their dependence on the 'day' may be, the weights are the same for all the time series.

Case II: *Different weights for each group of time series.* This situation can arise if it is thought that the relevance of past data decays over time at different rates for the time series belonging to the various groups. As alluded to above, Case II is of relevance in the construction of a overall P&L distribution when some, but not all, of the time series display the type of seasonality encountered, for instance, in some commodities.

[1] The industry definition for which months constitute 'summer' and which constitute 'winter' vary by commodity and region. Our choice of months here is just representative.

3 Terminology and Notation

We assume to have N_s time series, one for each risk factor. Each time series is made up of N_d entries ('days'). For a given day, the associated N_s entries contain the changes in the risk factors observed during that day. So, to each day there is an associated vector, $\mathbf{\Delta r}_i, i = 1, 2, ..., N_s$ of risk changes, with each $\mathbf{\Delta r}_i = (1\text{x}N_s)$. For simplicity of exposition we will assume that all the vectors, $\mathbf{\Delta r}_i$, are arranged in decreasing chronological order, with $\mathbf{\Delta r}_1$ corresponding to the most recent available date ('yesterday'). This choice is only made for clarity of exposition and will not affect the results. The whole set of raw data is therefore made up of a $(N_d\text{x}N_s)$ matrix.

We subdivide the N_s time series into M groups[2]. Let $n_m, m = 1, 2, ...M$, be the number of time series belonging to the mth group. Clearly,

$$\sum_{m=1}^{M} n_m = N_s \tag{1}$$

We associate to each group and to each day a weight, w_k^m, $m = 1, 2, ...M$, $k = 1, 2, ..., N_d$.

Let MP_0 be the set of current market risk positions affected by the N_s risk factors, and let $f_{MP_0} : \mathcal{R}^{N_s} \longrightarrow \mathcal{R}^1$ be the mapping from the changes in the N_s risk factors to the P&L. For the ith day this reads: $f_{MP_0}(\mathbf{\Delta r}_i|MP_0) = P\&L_i$, $i = 1, 2, ..., N_d$. The $(1\text{x}N_d)$ vector $P\&L_i$ contains the hypothetical P&L produced by the *current* market positions (hence the subscript MP_0) if they had been subjected to the changes in the market risk factors experienced on any of the N_d days in the data set.

With a traditional HS technique, the true unknown distribution of P&Ls is approximated by a series of N_d Dirac-delta distributions, each of 'area' $\frac{1}{N_d}$. This implicitly corresponds to assigning identical weights to all the days: $w_k^m \equiv \frac{1}{N_d}$,for any m, k. We would like to construct an approximation for the true unknown distribution of P&Ls when the weights w_k^m are exogenously assigned to be non-constant.

4 Determination of Joint Weights

Consider the case where we are trying to estimate a statistic of a population (say, its variance). We have a finite number of observations. For some reason we have more confidence in some of these observation than others, or, more generally, we have reasons to believe that, conditional on the current state of the world, some observations are more relevant than others to this current state. We can express our degree of confidence by means of normalized weights, w_i. It will help the discussion below if we require the weights to be rational rather than real numbers. To give precise meaning to the concept of 'confidence', we

[2] In practice, N_s will be of the order of a few thousands or of tens of thousands, while M will be of the order of a few units.

can imagine that each observation was repeated in the sample a number of times proportional to its weight[3]. Let n the number of real observations, and N the total number of repeated observations.

Given this way of looking at the problem, let us estimate the simplest population statistics, ie, the mean of the population. Then, in the case of identical weights we have:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2}$$

In the case of non-constant weights we have

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{n} h_i x_i \tag{3}$$

where h_i is the 'repetition factor' for occurrence i and

$$\sum_{i=1}^{n} h_i = N$$

The two expressions can be written in a common format as

$$\overline{x} = \sum_{i=1}^{n} w_i x_i$$

with

$$w_i = \frac{1}{n} \text{ (constant weights)} \tag{4}$$

and

$$w_i = \frac{h_i}{N} \text{ (non-constant weights)} \tag{5}$$

Calculating any statistic for a univariate variable presents no problems. For the variance, for instance, we have the estimator (neglecting the -1 factor for simplicity)

$$var[x] = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{6}$$

and

$$var[x] = \frac{1}{N}\sum_{i=1}^{n} h_i(x_i - \overline{x})^2 \tag{7}$$

for the constant and non-constant-weight case, respectively. Both can be written as

$$var[x] = \sum_{i=1}^{n} w_i(x_i - \overline{x})^2 \tag{8}$$

[3] This was the reason for requiring each weight to be a rational number. Clearly, apart from some rounding, the extension to real weights is straightforward, especially because, as shown below, the additional computational cost assciated with the virtual repetitions is almost zero.

Consider now vectors of observations. For notational simplicity we will only deal in this section with the case of two variables, x and y, but the extension to many variables is obvious. Each 'observation' is now a couple of numbers (x, y). We have n such couplets of observations: $(x, y)_i$, $i = 1, 2, ..., n$. If our 'trust' in any couplet $(x, y)_i$ only depends on the index i, then the extension of the above to statistics involving the joint distribution of the population is straightforward. Consider, for instance the covariance:

$$cov[x, y] = \frac{1}{N} \sum_{i=1}^{n} h_i (x_i - \overline{x})(y_i - \overline{y}) \tag{9}$$

Note that we are giving the same degree of confidence to (we are 'repeating the same number of times') both x and y belonging to the ith couplet. But what if, for each couplet, we had different degree of confidence in x and in y? (This would be the case for past joint observations of changes in, say, commodities and equities risk factors in the presence of seasonality). Suppose that, for the ith observation, x has a very high weight, but y has a very low weight. If h_i^x and h_i^y are the number of 'repetitions' for x and y on a stand-alone basis, respectively, it is clearly neither correct to repeat the couplet h_i^x nor h_i^y times. In particular, if the weight, say, h_i^x were zero one would never want to calculate a statistic that depended on x and y using the ith couplet because, given the chosen weights, the ith value of x would be totally unreliable or irrelevant.

To overcome this problem we observe that the joint weight given by the product $h_i^x * h_i^y$ has better properties as an indicator of how many times one should count a given couplet for the joint statistics. One is therefore led to introduce new weights $\omega_i^{xy} = \frac{h_i^x * h_i^y}{M}$, where

$$M = \sum_{i=1}^{n} h_i^x * h_i^y \tag{10}$$

Recalling the definition of weights we also have

$$h_i^x = w_i^x N \tag{11}$$

$$h_i^y = w_i^y N \tag{12}$$

and where it is assumed by construction that

$$\sum_{i=1}^{n} h_i^x = \sum_{i=1}^{n} h_i^y = N$$

and therefore

$$M = N^2 \sum_{i=1}^{n} w_i^x * w_i^y \tag{13}$$

and

$$\omega_i^{xy} = \frac{h_i^x * h_i^y}{M} = \frac{w_i^x w_i^y}{\sum_{i=1}^{n} w_i^x * w_i^y} \tag{14}$$

The expression for the covariance then becomes

$$cov[x,y] = \sum_{i=1}^{n} \omega_i^{xy}(x_i - \overline{x})(y_i - \overline{y}) \qquad (15)$$

where the same weights ω_i^{xy} are used for the calculation of the means $\overline{x}$ and $\overline{y}$.

If, for any reason, realizations of x were very reliable (relevant) when realizations of y are not reliable (relevant) at all, and vice versa, the error in computing the covariance would be very high (the term M would be very small). If the weights w_i^x were non-zero exactly when the weights w_i^y are zero, one cannot compute the covariance at all (which makes sense, as we have no confidence in any of the joint realizations).

From the discussion above it is clear that Equation (14) can be generalized for the case of n variables to

$$\omega_i^{(x_1,x_2,\ldots,x_n)} = \frac{\Pi_{k=1}^{n} w_i^k}{\sum_{i=1}^{N_d} \Pi_{k=1}^{n} w_i^k} \qquad (16)$$

5 The Algorithm

From the discussion above, the algorithm to deal with Case I (same weights for all groups) is trivial : we just draw synchronously from all the time series a number of times proportional to their weights (possibly modulo some rounding). More precisely, we 'loop' over all the days and give to the Dirac-delta distribution of P&L corresponding to each day k a weight proportional to w_k. In practice, this amounts to populating the P&L distribution by 'repeating' a number of times $h * N_d * w_k$ each contribution to the P&L (with h an integer greater or equal to 1). In the special case of constant weights, $w_k = \frac{1}{N_d}$, h can be chosen equal to 1, and each contribution to the P&L distribution appears exactly once.

We stress that, since the computationally expensive part in creating the P&L distribution is the valuation of the P&Ls corresponding to the different changes in risk factors, the procedure above has virtually no additional computational cost.

For Case II (different weight functions for each group) the following procedure can be applied under the assumption below.

Assumption The P&L of each instrument only depends on the risk factors in one group at a time.

Essentially, this assumption allows us to treat products like, say, quanto swaps that might be subject to risk factors in different groups by breaking apart the sensitivities to the different risk factors in the different groups. Under this assumption, the algorithm is as follows.

1. Consider each group of risk factors to which the same weights apply.

2. For each day, calculate the quantities $P\&L_i^m$, $m = 1, 2, ..., M$ in each group, possibly after scaling the $P\&L$s in each time series if required (see next section).

3. Add the various group-specific $P\&L_i^m$ to obtain the total P&L (summed over groups) for that day: $P\&L_i = \sum_{m=1}^{M} P\&L_i^m$.
4. Repeat for all the days.
5. To obtain the distribution of P&Ls compatible with the chosen weights for the risk factors, count the total $P\&L_i$ on each day a number of times proportional to the weight given in Equation (16) with $n = M$, where the variables $x_1, x_2, ..., x_n$ now stand for the groups.
6. From the distribution of P&Ls, extract the desired statistics (eg, look at the desired percentile of the empirical distribution to obtain statistics to estimate VaR, etc).

The validity of this procedure rests on the fact that, thanks to the Assumption above, one can deal directly with the sum of P&Ls within a group, rather than looking at the risk factors as the random variables of interest. Trivially, this 'group $P\&L$' random variable has the weight associated with all the risk factors in that group.

6 Volatility Rescaling: A Worked-Out Example

The weighting scheme presented in Section 5 appears very intuitive and, indeed, almost obvious. There are, however, some subtle points related to the application of Equation (16) that are best discussed by means of a simple, if artificial, example.[4] Section 7 presents a more realistic application.

Consider two asset classes ('equities', labelled X and 'commodities', labelled Y) each with four observations of price changes, labelled x and y, and non-constant weights, w_x and w_y, as in Tab 17:

$$\begin{bmatrix} \text{Time} & x & y & w_x & w_y & \omega_{xy} \\ 1 & 0.1 & 0.1 & 0.25 & 0.5 & 0.5 \\ 2 & 0.3 & 0.3 & 0.25 & 0.5 & 0.5 \\ 3 & 1 & 1 & 0.25 & 0.0 & 0 \\ 4 & 3 & 3 & 0.25 & 0.0 & 0 \end{bmatrix} \tag{17}$$

The last column shows the weights calculated using Equation (16). We propose to calculate the correlation among the changes in the two asset classes. We also consider a portfolio made up of two long positions with notionals of 1 in equities and commodities, and we calculate the VaR associated both with the two stand-alone portfolios (equity and commodities) and with the combined portfolio. (For simplicity we assume the VaR of the portfolio to be given by 1 standard deviation of the portfolio returns.)

Before starting any calculation, a glance at the contrived example in Table 17 suggests than any reasonable weighting scheme should give 1 for the correlation

[4] We would like to thank Dr Rick Klotz for suggesting this example and for useful discussions on some related points.

between equities and commodities, and a VaR equal to the sum of the VaRs for the individual portfolios. Let's see under what conditions these results are obtained.

Let's calculate first the averages of x and y, $\bar{x}_w$ and $\bar{y}_w$ with their weights w_x and w_y:

$$\bar{x}_w = \sum_{i=1,4} w_x^i x_i = 1.1 \tag{18}$$

$$\bar{y}_w = \sum_{i=1,4} w_y^i y_i = 0.2 \tag{19}$$

(In this section the subscripts w or ω emphasize which weights have been used in calculating the average).

Similarly we can calculate the variances of x and y as

$$\sigma_w^2[x] = \sum_{i=1,4} w_x^i \left(x_i - \bar{x}_w\right)^2 \tag{20}$$

$$\sigma_w^2[y] = \sum_{i=1,4} w_y^i \left(y_i - \bar{y}_w\right)^2 \tag{21}$$

We obtain $\sigma_w^2[x] = 1.315$ and $\sigma_w^2[y] = 0.01$.

We now move to the calculation of a joint statistic such as the covariance between x and y, for which we require the weights ω_{xy}. If we compute

$$covar[x,y] = \sum_{i=1,4} \omega_{xy}^i \left(x_i - \bar{x}_w\right)\left(y_i - \bar{y}_w\right) \tag{22}$$

we obtain $covar[x,y] = 0.01$.The correlation would then be given by

$$\rho_{xy} = \frac{covar[x,y]}{\sigma_w[x]\sigma_w[y]} = 0.0872 \tag{23}$$

very different from the correlation of 1 that we were expecting.

A moment's reflection shows that the problem stems from the inconsistent use of the weights w_x and w_y for *marginal* statistics ($\bar{x}_w$ and $\bar{y}_w$) to calculate a *joint* statistics (ρ_{xy}). If one consistently uses the weights ω_{xy} in calculating the averages and the variances:

$$\bar{x}_\omega = \sum_{i=1,4} \omega_{xy}^i x_i \tag{24}$$

$$\bar{y}_\omega = \sum_{i=1,4} \omega_{xy}^i y_i \tag{25}$$

$$\sigma_\omega^2[x] = \sum_{i=1,4} \omega_{xy}^i \left(x_i - \bar{x}_w\right)^2 \tag{26}$$

$$\sigma_\omega^2[y] = \sum_{i=1,4} \omega_{xy}^i \left(y_i - \bar{y}_w\right)^2 \tag{27}$$

one obtains $\bar{x}_\omega = \bar{y}_\omega = 0.2$, $\sigma^2_\omega[x] = \sigma^2_\omega[y] = 0.01$, $covar[x,y] = 0.01$ and $\rho_{xy} = 1$, as expected.

Let us now move to the calculation of the VaR of the two stand-alone positions and of the combined portfolio. As we have assumed that each portfolio is given by a single long position of notional 1 and that the VaR is just equal to one standard deviation, we immediately obtain

$$VaR\,[x] = \sqrt{\sigma^2_w[x]} = 1.1467 \tag{28}$$

$$VaR\,[y] = \sqrt{\sigma^2_w[y]} = 0.1 \tag{29}$$

Clearly, for the calculation of the stand-alone VaRs (which are marginal statistics) we want to use the weights w_x and w_y. Note, in particular that, if we used the weights $\{\omega\}$ to estimate the VaR of x we would be discarding useful information, as we would effectively 'throw away' the entries for x associated with times 3 and 4.

The changes in value of the portfolio, P, on the four days are shown in Tab (30), together with their weights:

Time	P	ω
1	0.2	0.5
2	0.6	0.5
3	2.0	0
4	6.0	0

(30)

The average and variance of the portfolio, calculated using the weights $\{\omega\}$, are:

$$\bar{P}_\omega = 0.4 \tag{31}$$

$$\sigma^2_\omega\,[P] = 0.04 \tag{32}$$

and its VaR, VaR_P, is therefore $\text{VaR}_P = \sqrt{\sigma^2_\omega\,[P]} = 0.2$. This number is however not equal to the 'intuitive' result given by the sum of the stand-alone VaRs for the two portfolios:

$$VaR\,[x] + VaR\,[y] = 1.1467 + 0.1 = 1.2467. \tag{33}$$

It is easy, however, to see that the 'intuitive' VaR can be recovered if we rescale the changes x by the ratio, R, of the volatilities of the x time series calculated with the old and the new weights:

$$R = \frac{\sigma_w\,[x]}{\sigma_\omega\,[x]} = 11.46 \tag{34}$$

The new rescaled time series and weights are now:

$$\begin{bmatrix} \text{Time} & x & y & w_x & w_y & \omega_{xy} \\ 1 & 1.1467 & 0.1 & 0.25 & 0.5 & 0.5 \\ 2 & 3.4402 & 0.3 & 0.25 & 0.5 & 0.5 \\ 3 & 11.467 & 1 & 0.25 & 0.0 & 0 \\ 4 & 34.402 & 3 & 0.25 & 0.0 & 0 \end{bmatrix} \tag{35}$$

and the new changes in the portfolio after the rescaling, P^r_ω, are given by:

$$\begin{array}{lll} \text{Time} & P^r_\omega & \omega \\ 1 & 1.2467 & 0.5 \\ 2 & 3.7402 & 0.5 \\ 3 & 12.467 & 0 \\ 4 & 37.402 & 0 \end{array} \tag{36}$$

The average and the variance of the rescaled portfolio are now given by $\bar{P}^r_\omega = 2.4934$ and $\sigma^2_\omega[P^r_\omega] = 1.5543$, respectively. This gives a VaR for the portfolio of

$$\mathrm{VaR}_P = \sqrt{\sigma^2_\omega[P^r_\omega]} = 1.24673 \tag{37}$$

that now exactly coincides with the sum of the stand-alone VaRs, as given by Equation (33). We note in passing that volatility rescaling for VaR purpose as been known since 1998 (see, eg, Hull and White (1998)). As we show below, however, its use in the case of non-constant weights introduces some subtle effects.

The rescaling seems to be a simple and neat solution of the paradox: what is achieved by this operation is to retain as much 'good' information about the marginal distribution of x as possible and to use it in the combined statistic. There is, however, no unique mathematically correct procedure to do this, and whether the rescaling is reasonable (and, indeed, useful) or not depends on the nature of the problem. For instance, if, in the stylized example above, the zero weights to times 3 and 4 for the series y had been given simply because our data series was faulty or corrupted, then the rescaling procedure is perfectly justifiable: there is no reason to throw away good information about x just because our y data are found wanting. More precisely, we expect the change over the next step to be 'small' for Y, (because the 'large' values for y of 1 and 3 are considered unreliable), but a change in X could be reasonably expected to be anywhere in the range 0.1 and 3.0 (because we believe that both large and small changes x are equally valid.) This will give rise to a 'rather large' VaR for the combined portfolio, that takes into account all we know about the volatility of x.

But this is not the only situation that can arise. If the zero weights had been given because we believe that, good as they may have been when collected, the y data for times 3 and 4 are now irrelevant (perhaps because of seasonality, or because they belong to a Markov state different than the currently prevailing one), the situation is different. Suppose, for instance, that, conditional on the present time, we believe that the magnitude of the moves in y can only be small (of the order of 0.1 and 0.3, not of 1 and 3) over the next time step *and that the nature of the co-dependence between x and y is such is that x can only be large if y is large* (this is fully compatible with the data). Then rescaling the x time series is no longer necessarily appropriate: the correct VaR should be 'low', the current best estimate of the variance of the combined portfolio over the next time step is indeed given by $\mathrm{VaR}_P = \sqrt{\sigma^2_\omega[P]} = 0.2$, and it was our original intuition that the total VaR had to be necessarily the sum of the two stand-alone VaRs that was faulty.

The example in this section shows that rescaling the magnitudes of the changes in the data series used for the portfolio VaR as suggested by Equation (34) can be a simple and useful device to retain some information that would otherwise be lost. The appropriateness of the rescaling, however, ultimately depends on the nature of our model of the underlying data-generating process.

7 A Realistic Example

We now move to a more realistic example by considering three series of risk factor changes ($\mathbf{\Delta r}_i$) over 500 business days (with the assumption of 250 business days per year). The first series is expected to show seasonality, the second series is not expected to show seasonality but is from the same asset class as the first, and the last one is from a different asset class. We choose to use the same weights for each asset class, and so we divide the three series into two groups ($M = 2$), use the periodic exponentially decaying weights with two years of data from Figure 1 for the seasonal series group, and use constant weights for the non-seasonal series group. Some of the data statistics are summarized below (P95 and P99 indicate the 95^{th} and 99^{th} percentile, respectively):

	Mean	Stdev	P95	P99
Gas Series	-0.0486	0.3646	-0.6820	-1.3480
Oil Series	0.0260	1.0273	-1.6800	-2.3900
Eurodollar Series	-0.0003	0.0496	-0.0300	-0.0650

We now consider a risk position of 1,000 Gas Futures, 1,000 Oil Futures, and 100,000 Eurodollar Futures. The mean and standard deviation for the one-day $P\&L$ (in thousands of dollars) in the position in each risk factor is then obtained by multiplying the numbers in the above table by 10,000 for Gas (1 Future = 10,000 MMBtu), 1000 for Oil (1 Future = 1000 barrels), and by 250,000 for Eurodollar futures (1 Future is for one million dollars deposit over three months, with a \$25 sensitivity per basis point). The $P\&L_i$ for each day i is obtained by adding the individual $P\&L$ numbers for the three positions. By doing this, we get the following:

	Mean P&L	Stdev P&L	P95	P99
Gas Series	-486	3,646	-6,820	-13,480
Oil Series	26	1,027	-1,680	-2,390
Eurodollar Series	-74	12,388	-7,500	-16,250

Now we compute the joint weights applied to each risk factor for each day using Equation (16). With these new weights, we can again calculate the means and standard deviations of the series:

	Mean P&L	Stdev P&L	P95	P99
Gas Series	-486	3,646	-6,820	-13,480
Oil Series	26	1,027	-1,680	-2,390
Eurodollar Series	-791	10,738	-5,000	-23,750

As can be noted, in the presence of different weights for the different groups the statistics obtained using the new joint weights do not in general match all the statistics obtained using the initial weights. Now, a decision has to made as to which is the important statistic of interest that is appropriate to match exactly. For this example, we assume that the stand-alone 95^{th} percentile VaR for each series is this statistic. We then calculate scaling factors for each series, scale up the returns, and then calculate the statistics again. We can also now calculate the statistics for the summation of the scaled P&L series.

	Mean P&L	Stdev P&L	P95	P99
Gas Series	-486	3,646	-6,820	-13,480
Oil Series	26	1,027	-1,680	-2,390
Eurodollar Series	-1,186	16,107	-7,500	-35,625
Sum P&L	-1,647	16,238	-9,840	-36,033

The scaling now ensures an exact match of the VaRs using initial weights and joint weights. One aspect of the scaling is that clearly not all statistics of interest can be jointly matched, so it is important to select the statistic relevant for a given application.

8 Conclusions

Historical Simulation is a powerful technique to estimate the VaR arising from synchronous vectors of market risk factor changes using as few assumptions as possible about the joint distribution of the changes in the risk factors. Usually, the HS technique is presented as a non-parametric method. This is, however, not correct, because the number of days in the vectors is actually an extremely important parameter, that implicitly generates a digital weighting structure: past days are either considered to be perfectly and identically relevant to today's conditions, or totally irrelevant.

This paper addresses the problem of generalizing the HS simulation technique when the weighting function is not of such binary nature, and when it may be different for different time series.

We have shown that a fairly intuitive weighting scheme for these synchronous vectors allows the estimation of distributional properties under these circumstances. We have also introduced a volatility rescaling scheme that can be of assistance in dealing with different weighting schemes for different risk factors. We have stressed that there is no single mathematically exact solution, and that the appropriateness of the volatility rescaling scheme depends on the specific problem at hand.

We conclude by noticing that the scheme we have presented can be of use in the application of HS methods for computing VaR in such a way that regulatory requirements (say, about the duration of the underlying time series) are met.

Bibliography

Bilmes, J., *What HMMs Can Do*, (2006), IEICE Trans. Inf. & Syst., Vol E89-D, No. 3, 869–891

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

Cappe, O., Moulines, E., and Ryden, T., (2005), *Inference in Hidden Markov Models*, 2005, Springer

Hull, J, White, A, (1998), *Incorporating Volatility Updating with the Historical Simulation Method for Value at Risk*, Journal of Risk, Fall 1998, 115-119

P. Jorion, *Value at Risk*, 2006, McGraw-Hill, 262-265.

Lambadiaris, G, Papdopoulos, L, Skiadopoulos, G, Zalis, Y, (2003), *VaR: Historical or Simulation?*, Risk, September, 124-126

Rebonato R, Chen J, (2007), *Evidence for State Transition and Altered Serial Codependence in US$ Interest Rates*, submitted for publication in Quantitative Finance

Schaller, H. and Norden, S.V., *Regime Switching in Stock Market Returns*, 1997, Applied Financial Economics, 7(2), 177–191(15).