

HISTORICAL SIMULATION

JOHN A. DODSON

1. SIMULATION

The objective of this section is to generate an $N \times M$ matrix of N simulated Δt -period simple total returns of M risk factors assumed to be jointly lognormal¹.

Let us assume that we have $T \cdot \Delta T$ years of timeseries history² for the (non-overlapping) ΔT -period simple return for each risk factor, $r_{t,m}$ where $t = 1, \dots, T$ and $m = 1, \dots, M$. If we transform these data according to

$$(1.1) \quad R_{t,m} \triangleq \ln(1 + r_{t,m})$$

the result is assumed to be T i.i.d. multinormal observations. Let us denote this $T \times M$ matrix by \mathbf{R} .

Let us also assume that we possess $N \cdot T$ standard normal variates, for example from a pseudo-random number generator. We can assemble these into an $N \times T$ matrix, \mathbf{Z} . We wish to transform this into an $N \times M$ matrix \mathbf{S} whose rows are independent and whose columns have the same distribution as the columns of \mathbf{R} . Then the reverse-transformed matrix with entries

$$(1.2) \quad s_{n,m} \triangleq \exp(S_{n,m}) - 1$$

meets the objective. We will allow for the analysis horizon for the simulations, Δt , to possibly differ from the sampling frequency of the data, ΔT by way of the usual linear time scaling of returns and covariances.

Before we calculate the mean and covariance of the T observations that make up each column of \mathbf{R} , let us first insulate ourselves from the possibility that these parameters may in fact be slowly varying by introducing an exponential weighting scheme³. Let us assume that the observations are ordered from most to least recent, and define a T vector \mathbf{w} with entries

$$(1.3) \quad w_i = \lambda^{i-1} \cdot \frac{1 - \lambda}{1 - \lambda^T}$$

Date: October 4, 2007.

¹ $N = 10,000$ and $M \approx 3,000$ are typical magnitudes for equity portfolios.

² $T < 1,000$ typically.

³This procedure is supported by Kalman filter methods.

for $0 < \lambda < 1$. We call lambda the decay factor⁴. In the implementation, it is useful to also define the closures $\mathbf{w}|_{\lambda=0} = (1, 0, 0, \dots)'$ and $\mathbf{w}|_{\lambda=1} = (\frac{1}{T}, \frac{1}{T}, \frac{1}{T}, \dots)'$. In all cases, the entries sum to one.

We can identify the population mean with the weighted average of the observations.

$$(1.4) \quad \mathbf{m} \cdot \Delta T \triangleq \mathbf{R}' \cdot \mathbf{w}$$

and we can identify the population covariance with the weighted average of the de-means squares.

It turns out we do not actually need to evaluate the covariance: The method defined below shows how to generate random scenarios without requiring the estimation of any volatilities or correlations.

Let us consider the following construction for the simulated results.

$$(1.5) \quad \mathbf{S} \triangleq \mathbf{1}_N \cdot \mathbf{m}' \cdot \Delta t + \mathbf{Z} \cdot \text{diag}(\sqrt{\mathbf{w}}) \cdot (\mathbf{R} - \mathbf{1}_T \cdot \mathbf{m}' \cdot \Delta T) \cdot \sqrt{\frac{\Delta t}{\Delta T}}$$

We show in appendix A that rows of \mathbf{S} are (multi)normal and share first and second moments with the data. Also, the rows are independent in the limit $T \rightarrow \infty$.

This methodology, which was first publicly documented (as far as the author knows) in an article in the RiskMetricsTMMonitor in 1997[1], avoids problems associated with the traditional approach for generating correlated normal variates using Cholesky Decomposition⁵.

⁴ $\lambda = 0.97$ or 0.94 are good choices for daily data.

⁵with the trade-off that the results are not strictly independent

APPENDIX A. PROPERTIES OF THE SCENARIO SET

By inspection, each row of \mathbf{S} is (multi)normal, since entries are linear combinations of normal random variables.

Let us explore the population ($N \rightarrow \infty$) moments.

The expected value of the entries of a row is

$$(A.1) \quad \frac{1}{N} \cdot \mathbf{S}' \cdot \mathbf{1}_N = \mathbf{m} \cdot \Delta t \cdot \frac{\mathbf{1}'_N \cdot \mathbf{1}_N}{N}$$

which equals $\mathbf{m} \cdot \Delta t$ as required, since each entry in \mathbf{Z} has expected value zero.

We will define the covariance of the data to be the weighted average of the de-meaned squares.

$$(A.2) \quad \mathbf{\Sigma} \cdot \Delta T = (\mathbf{R} - \mathbf{1}_T \cdot \mathbf{m}' \cdot \Delta T)' \cdot \text{diag}(\mathbf{w}) \cdot (\mathbf{R} - \mathbf{1}_T \cdot \mathbf{m}' \cdot \Delta T)$$

The covariance matrix of the entries of a row of \mathbf{S} is

$$(A.3) \quad \begin{aligned} \frac{1}{N} \cdot (\mathbf{S} - \mathbf{1}_N \cdot \mathbf{m}' \cdot \Delta t)' \cdot (\mathbf{S} - \mathbf{1}_N \cdot \mathbf{m}' \cdot \Delta t) \\ = (\mathbf{R} - \mathbf{1}_T \cdot \mathbf{m}' \cdot \Delta T)' \cdot \text{diag}(\mathbf{w}) \cdot (\mathbf{R} - \mathbf{1}_T \cdot \mathbf{m}' \cdot \Delta T) \cdot \frac{\Delta t}{\Delta T} \end{aligned}$$

which equals $\mathbf{\Sigma} \cdot \Delta t$ as required, since $\mathbf{Z}' \cdot \mathbf{Z} \rightarrow N \cdot \mathbf{I}_T$.

Finally, the covariance of the scenarios is

$$(A.4) \quad \frac{1}{T} \cdot (\mathbf{S} - \mathbf{1}_N \cdot \mathbf{m}' \cdot \Delta t) \cdot (\mathbf{S} - \mathbf{1}_N \cdot \mathbf{m}' \cdot \Delta t)' = \frac{1}{T} \cdot \mathbf{Z} \cdot \mathbf{A} \cdot \mathbf{Z}'$$

where

$$(A.5) \quad \mathbf{A} \triangleq (\mathbf{R} - \mathbf{1}_T \cdot \mathbf{m}' \cdot \Delta T) \cdot (\mathbf{R} - \mathbf{1}_T \cdot \mathbf{m}' \cdot \Delta T)' \cdot \text{diag}(\mathbf{w}) \cdot \frac{\Delta t}{\Delta T}$$

is an $T \times T$ matrix whose entries consist of weighted and scaled combinations of the data.

It can be seen that off-diagonal entries of the covariance consist of an average of terms each involving two unequal normal variates. Therefore the off-diagonal entries go to zero as $T \rightarrow \infty$. Strictly speaking, for any finite dataset the rows may in fact exhibit some residual correlation (even for $N \rightarrow \infty$); but we will assume that this is negligible. To the extent that the scenarios are uncorrelated, the fact that they are (multi)normal guarantees that they are independent.

This completes the demonstration that \mathbf{S} has the required properties.

REFERENCES

- [1] Peter Benson and Peter Zangari. A general approach to calculating var without volatilities and correlations. *RiskMetricsTM Monitor*, (2), 1997.

(John A. Dodson) RIVERSOURCE INVESTMENTS, INVESTMENT RISK MANAGEMENT, RESEARCH & DEVELOPMENT