Return Volatility Estimates: A Review and Practical Analysis

Grant Holtes

January 3, 2024

Summary

This paper focuses on the estimation of return volatility and addresses challenges that arise where the frequency of available data is shorter than the desired return horizon. A range of estimators are described and evaluated. The evaluation employs both synthetic data, generated to mimic specific data-generating processes, and real-world data from equities. The synthetic data analysis reveals that the log return method, when returns are independent and identically distributed (IID), provides an unbiased estimator with lower variance than other methods. However, in the presence of serial correlation and momentum, the log return method is biased, requiring correction for autocorrelation. Simpler approaches, such as resampling to a yearly frequency or using rolling annual returns, perform well across various scenarios and do not require assumptions on how returns are generated, making them more robust estimators.

1 Introduction

Estimation of return volatility is a foundational task in quantitative finance, with volatility being used as in input in CAPM modelling, portfolio optimisation and risk management models, among other use cases. When volatility is defined as the standard deviation of returns, its estimation seems trivial: Given a time series of returns, simply calculate the standard deviation. This is complicated by two questions:

What return horizon matters: For many tasks, such as portfolio optimisation, we really care about longer term returns. For other tasks, such as risk management, we may be more concerned with shorter term returns. The selection of return horizon is a important step.

How to best use all available data: If your data is reported at a shorter period than the return horizon, how can you use the higher frequency data to get better estimates than resampling at the return frequency?

2 Estimators

This analysis assumes that annual volatility is the estimate of interest, but these methods could be applied to any case where the frequency of interest is longer than the frequency of available data.

2.1 Multiply by squareroot of periods per year

This is the commonly quoted method which appears on most forums and websites, which seems useful as it uses all of the available data to estimate volatility, before converting to an annual figure. Where there are p return periods observed per year, we can calculate annual volatility with Equation 1.

$$\sigma_{annual} = \sqrt{p}\sigma_p \tag{1}$$

For example, to convert monthly return volatility to annual we would use Equation 2

$$\sigma_{annual} = \sqrt{12}\sigma_{monthly} \tag{2}$$

This method uses the properties of the variance estimator, which states that given n independent and identically distributed (IID) random variables with variance σ_p, the variance of their sum is given by Equation 3

$$\sigma^2_{annual} = p\sigma^2_p \tag{3}$$

This makes the assumption that returns for each year are given as the sum of returns in each period, which implies that returns in each period are not reinvested and do not compound. For most use cases this is an incorrect assumption.

2.2 Log return method

The equation in Equation 3 can be used while preserving the effects of compounding if log returns are used, as log returns are compounded through addition rather than multiplication. This method requires the following steps:

1. Convert returns to log returns: $ln(P_t/P_{t-1})$ where P_t is the price or index level at time t
2. Compute the mean μ_{lp} and variance σ^2_{lp} of the log returns
3. Calculate mean annual log return as $\mu = p\mu_{lp}$ and the annual log variance as $\sigma^2 = p\sigma^2_{lp}$.
4. Convert from log returns using Equation 4, which the the equation for the standard deviation of a log-normal distribution.

$$\sigma = e^{\mu+0.5\sigma^2}\sqrt{e^{\sigma^2} - 1} \tag{4}$$

Similar to the previous method, this assumes that the monthly log returns are IID.

2.3 Log return method with serial correlation

We can relax the IID assumption to allow for serial correlation between the period's returns. This requires the variance of the annual log returns to be adjusted for autocorrelation before conversion using Equation 6, as described by Lo, 2002.

$$R_t(q) = \sum_{k=1}^{q-1} R_t \tag{5}$$

Where R_t are the log returns for a single period, $R_t(q)$ is the log returns for q periods. Given σ^2 is the variance of returns and p_k is the correlation between returns at a lag of k.

$$VAR(R(q)) = q\sigma^2 + 2\sigma^2 \sum_{k=1}^{q-1} (q-k)p_k \tag{6}$$

Its worth noting that $VAR(R(q))$ can also be computed as the sum of the covariance matrix between R_t and its q lags.

2.4 Resample to return period

All of the methods described so far use assumptions regarding the nature of how return are generated and combined over time to map from the high frequency measurement of volatility to annual volatility. When these assumptions are incorrect, for example when returns are not IID, the estimators can perform poorly. A simpler approach which places no assumptions on the data generating process is to simply resample the price or index data at a yearly frequency to compute volatility.

2.5 Rolling return

Simply resampling to the return period can seem unwise, as it throws away observations that we seemingly should be able to use to compute a better estimate. One method to still compute annual returns but use all available data is to compute the rolling annual returns, then compute the volatility of this series.

$$R_t = \frac{P_t}{P_{t-p}} - 1 \tag{7}$$

The issue with this method is that each successive return is tightly coupled to the previous return, containing many of the same time periods, which we would expect to cause the series of rolling returns to contain a high degree of serial correlation and not much additional information when compared to the simpler approach of resampling at a yearly frequency.

2.6 Bootstrapped annual returns

Another tempting method is to bootstrap annual returns by randomly sampling from the period level returns to construct a larger sample of annual returns. This suffers the same flaws as the log return method, as it assumes that returns in each observed period are IID. When evaluating this method, random sampling without replacement is used, generating a sample size that is 10 times the size of the input sample.

3 Evaluation approaches

We can test these estimators through two methods:

Synthetic data: Testing on generated data that follows a specific data generating process, to test how the estimators perform when returns both are and aren't IID.

Real world data: Test the ability for these estimators to predict out-of-sample volatility on real world series where the data generating process is unknown.

4 Synthetic data evaluation

4.1 Methodology

The following Monte Carlo method is used to evaluate each estimator on synthetic data:

1. A extremely long series of asset prices is generated that follows a specific data generating process of interest
2. The full series is resampled at a yearly frequency, non-overlapping yearly returns are computed and the "true" volatility is estimated from this series. This will act as the target that each estimator is evaluated against.
3. A small subset of observations from the start of the series is extracted, and each estimator is used to compute the volatility of this series.
4. This process is repeated for a number of trials, allowing the bias and variance of each return volatility estimator to be measured.

4.2 Results on IID returns

As a simple first case, asset prices are generated where returns for each observation period are IID, as in Equation 8 where $R_t \sim N(\mu, \sigma^2)$.

$$P_t = P_{t-1}(1 + R_t) \tag{8}$$

An example of one of the series generated is given by Figure 1.

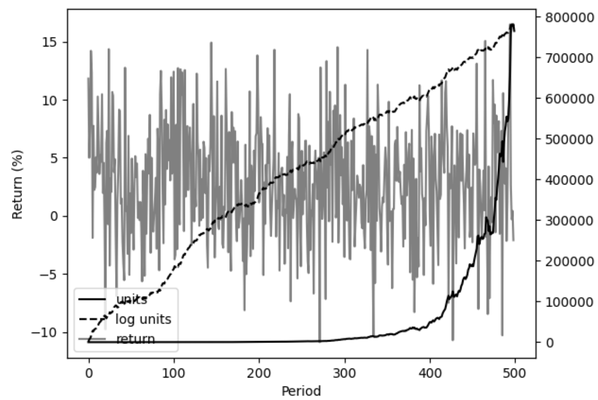

Figure 1: Data generated at a monthly frequency with monthly returns $R_t \sim N(3\%, 5\%^2)$. First 500 periods shown in raw units, log units and return

The results in Table 1 show that the simple method of multiplying monthly volatility by the square root of 12 provides results that are biased towards zero. As the IID assumption is true, the remaining methods provide similar and unbiased results. Notably the log returns and bootstrapped methods provide lower estimator variance, with half the standard deviation of the remaining measures.

Method	Average vol	Mean error	Mean absolute error	Error stdev
multiply by sqrt frequency	0.17258	-0.06871	0.06871	0.00875
log returns	0.24532	0.00404	0.01256	0.01516
log returns with autocorrelation	0.23116	-0.01013	0.02873	0.03422
resample at yearly freq	0.22928	-0.012	0.0362	0.04435
rolling annual returns	0.2304	-0.01089	0.02894	0.03397
bootstrapped returns	0.23429	-0.00699	0.01397	0.01585

Table 1: Error on independent returns series. Data generated at a monthly frequency with monthly returns $R_t \sim N(3\%, 5\%^2)$. 20000 periods generated, 240 (20y) used to estimate volatility. 1000 trials generated.

The correlation in error between estimators is shown in Table 2. As expected, the errors in the yearly resampled method and rolling annual returns method are highly correlated, with a correlation of 0.81, showing that there may be limited benefit to using rolling returns over simply resampling to a yearly frequency.

Method	mult..	log ..	log ..	resa..	roll..	boot..
multiply by sqrt frequency	1.0	0.78	0.3	0.21	0.31	0.62
log returns	0.78	1.0	0.35	0.29	0.36	0.8
log returns with autocorrelation	0.3	0.35	1.0	0.81	0.97	0.25
resample at yearly freq	0.21	0.29	0.81	1.0	0.81	0.21
rolling annual returns	0.31	0.36	0.97	0.81	1.0	0.26
bootstrapped returns	0.62	0.8	0.25	0.21	0.26	1.0

Table 2: Correlation between errors for each method on independent returns series. Data generated at a monthly frequency with monthly returns $R_t \sim N(3\%, 5\%^2)$. 20000 periods generated, 240 (20y) used to estimate volatility. 1000 trials generated.

Similar results are found when using a weekly data frequency - See appendix items Table 6 and Table 7.

4.3 Results on returns with momentum

We can add momentum to the series by generating the returns such that they are positively correlated with returns in previous periods.

$$P_t = P_{t-1}(1 + R_t) \tag{9}$$

$$R_t = \beta_1 R_{t-1} + \beta_0 + \epsilon_t \tag{10}$$

where $\epsilon \sim N(0, \sigma^2)$ and $\beta_0 = \mu(1 - \beta_1)$ to ensure the average return is kept constant at μ.
An example of one of the series generated is given by Figure 2.

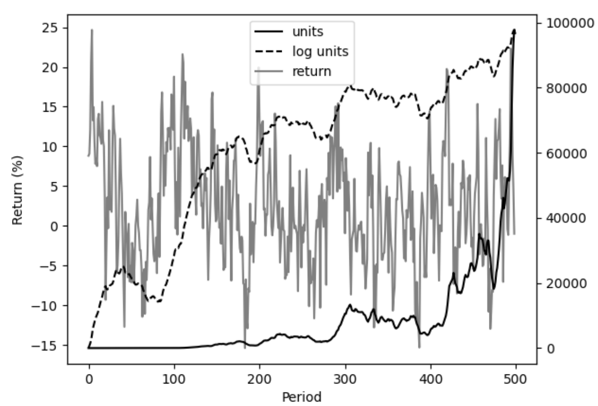

Figure 2: Data generated at a monthly frequency with monthly returns, $\beta_1 = 0.7, \mu = 0.03, \sigma = 0.05$. First 500 periods shown in raw units, log units and return

The introduction of non-IID returns results in the simple square-root multiple, log returns and bootstrap estimators to be biased, as in Table 3. As is implied by Lo, 2002 this bias is negative, resulting in an underestimation of volatility in the presence of positive autocorrelation. The log-returns with autocorrelation compensation estimator performs well, with seemingly unbiased results. Notably, the simple methods of resampling to a yearly frequency and using rolling annual returns were also unbiased and had a similar variance to the log-returns with autocorrelation compensation estimator.

Method	Average vol	Mean error	Mean absolute error	Error stdev
multiply by sqrt frequency	0.2379	-0.55971	0.55971	0.02964
log returns	0.34021	-0.4574	0.4574	0.05403
log returns with autocorrelation	0.77676	-0.02086	0.14935	0.18443
resample at yearly freq	0.73159	-0.06603	0.16496	0.19254
rolling annual returns	0.73856	-0.05905	0.15191	0.17676
bootstrapped returns	0.32386	-0.47375	0.47375	0.05115

Table 3: Error on returns series with return momentum, $\beta_1 = 0.7, \mu = 0.03, \sigma = 0.05$. 20000 periods generated, 240 (20y) used to estimate volatility. 1000 trials generated.

The correlation results in Table 4 show that the presence of serial correlation in returns causes the simple annual return and rolling annual return methods to be even more highly correlated, with a correlation of 0.9.

Method	mult..	log ..	log ..	resa..	roll..	boot..
multiply by sqrt frequency	1.0	0.59	0.51	0.39	0.44	0.61
log returns	0.59	1.0	0.74	0.64	0.71	0.96
log returns with autocorrelation	0.51	0.74	1.0	0.8	0.88	0.7
resample at yearly freq	0.39	0.64	0.8	1.0	0.9	0.59
rolling annual returns	0.44	0.71	0.88	0.9	1.0	0.66
bootstrapped returns	0.61	0.96	0.7	0.59	0.66	1.0

Table 4: Correlation between errors for each method on independent returns series. Returns series generated with return momentum, $\beta_1 = 0.7, \mu = 0.03, \sigma = 0.05$.. 20000 periods generated, 240 (20y) used to estimate volatility. 1000 trials generated.

5 Real world data

We can also test these methods on real world data, although in a more limited fashion. The method used takes a history of equity prices and divides it in half. The first half is used by each estimator to estimate volatility, which can then be compared to the actual volatility in the second half of the series. This aims to simulate the usage of these estimators in a practical use case, where the goal is to predict future volatility. 30 years of data from 1993 to 2023 is used, allowing 15y for each of the subsets. The "target" volatility in the second half of the series is measured using the resampling at yearly frequency method, as with the synthetic data.

series	AAPL	CL	ED	DE	SPY
target	0.46	0.121	0.142	0.293	0.187
multiply by sqrt frequency	0.5	0.256	0.19	0.305	0.152
log returns	0.711	0.302	0.201	0.386	0.167
log returns with autocorrelation	0.914	0.176	0.194	0.298	0.171
resample at yearly freq	0.85	0.192	0.18	0.328	0.157
rolling annual returns	0.757	0.187	0.188	0.264	0.162
bootstrapped returns	0.734	0.294	0.171	0.343	0.168

Table 5: Volatility predictions for annual returns of a selection of equities for 2008 to 2023, based on data from 1993 to 2007. Data from Google finance, unadjusted close price, weekly data frequency

The results in Table 5 shows that the estimators are all poor predictors, with significant errors with all estimators. Interestingly, the simple method of multiplying by the square root of the frequency does not provide results that are consistently biased downwards as was seen in the synthetic data, even when comparing against the "true" annual volatility for the training period (resample at yearly freq estimator). This suggests that the equity prices tested compound in an additive fashion rather than multiplicative, which is a topic that deserves separate investigation. This is less likely to be the case if dividend reinvestment was accounted for. This simple method also had the lowest average absolute error across the five equities tested.

6 Conclusion

The following conclusions can be made from the observations:

1. Theoretically, the method of using the squareroot of the number of periods per year multiplied by the standard deviation of returns per period should be the worst estimator, given it relies on two faulty assumptions: 1) Returns are IID, and; 2) returns add over the year rather than compound.

2. If the data generative process of the returns is known to give IID returns, using the log-return method gives an unbiased estimator with lower variance than resampling to yearly returns or rolling annual returns.

3. If returns are not IID and momentum is present, the log returns method is biased. If the autocorrelation of returns is positive, this bias will reduce the estimated volatility. This can be corrected for through accounting for autocorrelation, which will provide an unbiased estimator in the case of returns with momentum and in the base IID returns case.

4. Simply resampling to a yearly frequency and calculating the standard deviation of returns performs well and is as good as the log returns methods, except for the special case of IID returns where the log return method has lower variance. The log return method with auto-correlation correction does not have this same low variance in the IID case and is no better than yearly returns.

5. Rolling yearly return volatility is highly correlated with the volatility given by resampling at a yearly frequency, and this correlation increases when returns are serially correlated. However, the rolling returns estimator has a slightly lower variance.

6. When used in a predictive fashion on real data, all estimators preformed poorly. This suggests non-stationarity in the returns series used and is deserving of additional investigation.

A higher level conclusion is that knowledge of the underlying data generating process of returns can allow for better estimators to be developed, which are unbiased and have low variance. However, if the assumptions placed on the data generating process are violated, these estimators can have a large bias. Simple methods which directly measure the property of interest, which may seems overly simplistic, provide results that are comparable to these more complex estimators, without the vulnerabilities that come from placing assumptions on the return distribution.

References

Lo, A. W. (2002). The statistics of sharpe ratios. *Financial Analysts Journal*, *58*(4), 36–52. https://doi.org/10.2469/faj.v58.n4.2453

7 Appendix

Method	Average vol	Mean error	Mean absolute error	Error stdev
multiply by sqrt frequency	0.07205	-0.0256	0.0256	0.00393
log returns	0.09822	0.00057	0.00359	0.00448
log returns with autocorrelation	0.09434	-0.00331	0.01075	0.01285
resample at yearly freq	0.09336	-0.00428	0.01331	0.01615
rolling annual returns	0.09426	-0.00339	0.01067	0.0128
bootstrapped returns	0.09517	-0.00247	0.0052	0.00604

Table 6: Error on independent returns series. Data generated at a weekly frequency with monthly returns $R_t \sim N(0.6\%, 1\%^2)$. 20000 periods generated, 1040 (20y) used to estimate volatility. 1000 trials generated.

Method	mult..	log ..	log ..	resa..	roll..	boot..
multiply by sqrt frequency	1.0	0.93	0.14	0.06	0.14	0.69
log returns	0.93	1.0	0.2	0.12	0.2	0.74
log returns with autocorrelation	0.14	0.2	1.0	0.79	0.98	0.14
resample at yearly freq	0.06	0.12	0.79	1.0	0.78	0.07
rolling annual returns	0.14	0.2	0.98	0.78	1.0	0.13
bootstrapped returns	0.69	0.74	0.14	0.07	0.13	1.0

Table 7: Correlation between errors for each method on independent returns series. Data generated at a monthly frequency with monthly returns $R_t \sim N(0.6\%, 1\%^2)$. 20000 periods generated, 1040 (20y) used to estimate volatility. 1000 trials generated.