The Use of Historical Data in Calculating the Expected Market Risk Premium

This note summarizes different approaches in using historical market return and interest rate data to estimate the expected market risk premium, such as might be used in the capital asset pricing model (CAPM). Some practitioners use arithmetic averages of historical data (Ibbotson Associates SBBI 1996 Yearbook as an example) while others discuss the geometric average Mark Kritzman (FAJ May-June 1993). The question is which method is correct, arithmetic or geometric? The answer is that it depends on what investment holding period assumption you make.

At least nine estimators exist:

- Arithmetic
- Weighted
- Adjusted
- Simple
- Overlap
- Wrap around overlap
- Combinatorial
- Rebalancing
- Geometric

Essentially, these nine methods can be reduced to three general procedures: arithmetic, geometric, or some combination of the two.

First, let's establish some terminology: N represents the assumed investment holding period measured by the number of data points (60 months, 5 years, etc.) in a holding period and T represents the number of data points in the practitioner's data set (such as 840 monthly data points or 70 yearly data points). Both N and T must be denominated in the same units, i.e. data can be monthly, yearly, etc, but both N and T must be measured in the same units. Let's assume for the purposes of a running discussion that the practitioner's data points are monthly. If the practitioner assumes N=1 then a monthly holding period is assumed. If N=12 then a one-year holding period is assumed, etc. Let's also assume that the practitioner has 840 months of data to work with.

Marshall Blume (JASA September 1974) documented the biases in using sample arithmetic or geometric means of one-period returns to assess long-run expected rates of return. According to Blume, this phenomenon had been previously observed by Marshall Blume and Irwin Friend (RES August 1974), Lawrence Fisher (JOB January 1966), and Pao Cheng and M. King Deets (JFQA June 1971). Blume assumed that actual returns (in return relative form) are equal to expected returns plus a surprise element, ε (and $\varepsilon \sim NID[0,\sigma^2]$ such that the surprises are expected to be equal to 0 and are independent and identically normally distributed with a fixed variance). Blume demonstrated the following:

- If N = 1, then the arithmetic average is unbiased, while the geometric average is biased downwards.

- If N>1 but N<T then the arithmetic average is biased upwards and the geometric average is biased downwards.

- If N=T then the geometric average is unbiased, while the arithmetic average is biased upwards.

He then described four unbiased estimators for use when $N>1$ but $N<T$, the most relevant situation to a practitioner:

- The simple estimator

When T is an integral multiple of N, multiply the first N monthly relatives together to get a compounded wealth relative, then compound the second N monthly relatives together, and so on until all T data points are used. Then take the arithmetic average of the compounded wealth relatives (the practitioner will have T/N of them) to obtain an unbiased estimate of the expected N-period wealth relative.

- The overlap estimator

Multiply the first one through N-period monthly relatives together to get the first compounded wealth relative, then multiply the 2^{nd} through $N + 1^{st}$ relatives together to get the second compounded wealth relative, etc. until all T data points are used. Then take the arithmetic average of the compounded wealth relatives to obtain an estimate of the expected N-period wealth relative. Blume points out that this method is probably less efficient than the simple method.

- The weighted estimator

The weighted estimator is simply a weighted average of the straight arithmetic mean compounded over N periods (A^N) and the straight geometric mean compounded over N periods (G^N). The weighting is given by:

$$\hat{E}\,(W_N) \approx \frac{T-N}{T-1} x A^N + \frac{N-1}{T-1} x G^N$$

where $\hat{E}(W_N)$ is the estimate of wealth after N periods. Notice that when $N = 1$, the weighted estimator of the expected wealth relative reduces to the compounded arithmetic mean and when $N = T$ the weighted estimator reduces to the compounded geometric mean. Blume indicates that the expected wealth relative is approximated by the equation, indicating that it is approximately unbiased.

- The adjusted estimator

The final estimator Blume proposed adjusts the straight arithmetic mean compounded over N periods (A^N) by an adjustment factor he calculated using his data and multiple regression. The adjustment factor is :

$$Adjustment\ Factor = -0.9174 + 1.0058x \ln \sigma(R) + 1.0441x \ln(N) - 0.9989x \ln(T)$$

where $\ln\sigma(R)$ is the log of the standard deviation of the monthly return relative. Blume found that for his assumed ranges of the true monthly market return relative and its true standard deviation, and for $N \leq 80$ and $T \leq 100$ and $N \leq T$, his regression fits the bias calculations "well." The practitioner would then divide A^N by the adjustment factor to yield an approximately unbiased estimator of $E(W_N)$. This adjustment factor depends on Blume's own multiple regression results and the practitioner's assumptions, or estimates, of true market return and volatility.

Blume concluded, based on Monte Carlo tests, that the simple estimator is more efficient than the overlap estimator, but less efficient than the weighted or adjusted estimators. If one cannot assume stationary independent normal distributions than one cannot use the adjusted estimator, and if one cannot assume independence of successive one-period relatives then one cannot use the weighted estimator. Both of those two non-linear estimators yield only a modest increase in efficiency even under ideal conditions. Therefore, the simple estimator is the preferable estimator of the four choices if $N>1$ but $N<T$.

Joel Hasbrouk (JFQA December 1983) developed explicit formulae for the standard errors of the simple estimator and the overlap estimator, because Blume relied on estimates from Monte Carlo analysis. He confirmed Blume's empirical finding that the simple estimator was more efficient than the overlap estimator, using plausible assumptions in his formulae. Hasbrouk also introduced two new estimators and their variances: the wrap-around overlap estimator and the combinatorial estimator.

•The wrap-around overlap estimator

The wrap-around overlap estimator is simply the overlap estimator, but instead of calculating the last compound wealth relative ending at T, the practitioner treats the data as if it wrapped around from the last data point back to the first data point. The practitioner continues to calculate wealth relatives from the assumed continuous data until all the data are used symmetrically. In our example, if N=12 (one year) then instead of stopping at compounding data points 829 through 840, the practitioner would compound data points 830 through 840 and data point 1. The practitioner would go on to compound data points 831 through 840 and 1 to 2, etc. until the last possible consecutive wealth relative is calculated using data point 840 and data points 1 through 11. Hasbrouk found that, using assumed inputs, that the wrap-around overlap estimator was more efficient than the simple or overlap estimators.

•The combinatorial estimator

The idea of resampling can be taken further, assuming that the one-period return relatives are independent. Instead of compounding data points that are consecutive, Hasbrouk proposed taking all possible combinations of return relatives to calculate N-period compound wealth relatives. In plain English, this means that, if N equals 60, the practitioner would compound the first data point with 59 other data points out of the entire data set, and repeatedly do that for every possible combination of data points. The process is repeated for data point 2, etc. The number of possible combinations (not permutations since order does not matter) that can be concocted is given by: $T!/(N![T-N]!)$. If N=60 and T=840 then the number of compounded wealth relatives the practitioner would have to calculate is approximately 3.972 times 10^{92} before taking an arithmetic average! A serious question arises if the practitioner is willing to trade off greater work for increased efficiency. A recent case before the OPUC indicated that calculations for N=9 and T= 69 took 28 hours to calculate on a Hewlett Packard Apollo 710 and that effort involved only about 57 million possible combinations.

Hasbrouk concluded that the overlap estimator should not be used, the combinatorial estimator is computationally costly, and that the wrap-around estimator yields an incrementally small gain in efficiency at a greater computational cost. The simple estimator will in most cases perform adequately.

Pao Cheng (JFQA December 1984) seemed a bit frustrated that the maximum efficient estimator (the combinatorial) was computationally out of reach, so he identified the complete set of unbiased estimators (but not examining the weighted or adjusted estimators). He presented explicit formulae and demonstrations of their efficiencies, and set out to develop another estimator that was close in efficiency to the combinatorial estimator but at far less computational cost. He proposed the rebalancing estimator.

•The rebalancing estimator

The rebalancing estimator compounds derived data points and requires that T be an integral multiple of N. The derived data points are arithmetic averages of the original data points. If N = 60 and T = 840, then the practitioner has 840/60 or 14 sets of holding periods. The rebalancing estimator takes the arithmetic average of the first 14 data points as the first derived data point, the arithmetic average of data points 15 through 29 to calculate the second derived data point, etc. until all the T data points are used. The practitioner will end up with 60 such derived data points, which are then compounded. Cheng presents efficiency estimates of the various estimators, and concludes that the rebalancing estimator is relatively closer in efficiency to the combinatorial estimator, yet even easier to compute than the wrap-around overlap and the overlap estimators. He also concluded that even if T/N were not an integer, it can pay to throw out data to get T/N to equal an integer, and still the practitioner will get a more efficient estimator than the wrap-around estimator. Cheng's efficiency estimates generally showed that the simple estimator was also more efficient than the wrap-around overlap estimator.

The conclusion one draws from this literature is that if N=1, the practitioner should use the arithmetic average of the historical observations to estimated the expected return. If N=T the practitioner should use the geometric average of historical observations to estimate the expected return. If N > 1 but N < T the practitioner should, in a manner of speaking, use a combination of arithmetic and geometric methods by arithmetically averaging some form of compounded wealth relative. The most efficient estimator is also computationally excessive, but other estimators do exist as alternatives.

Now, how does the practitioner use these concepts to estimate the expected market risk *premium* for use in the CAPM? One possible method would be to stop short of making final estimates of expected stock market returns, call it $\hat{E}(W_{S,N})$, and US Treasury market returns, call it $\hat{E}(W_{B,N})$, and to first calculate market risk premium estimates before taking the arithmetic average. Let's denote the market risk premium by γ. The estimator described above is the following:

$$\hat{E}(\gamma_N) = \frac{1}{T/N} x \sum_{i=1}^{T/N} \frac{\hat{E}(W_{S,i})}{\hat{E}(W_{B,i})}$$

Assuming T = 840 and N=60, the practitioner will have T/N or 14 stock market wealth relatives over the assumed N-period investment horizon and 14 US Treasury security wealth relatives. The practitioner then divides the stock market wealth relative by the US Treasury market wealth relative for each of the 14 holding periods. Finally, the practitioner takes the arithmetic average.

Then, convert the holding period market risk premium back into its monthly equivalent:

$$\hat{E}(\gamma_{Monthly}) = \hat{E}(\gamma_N)^{\frac{1}{N}}$$

Finally, convert the monthly market risk premium into an annual equivalent:

$$\hat{E}(\gamma_{Annual}) = (\hat{E}[\gamma]_{Monthly} - 1)x12$$

One other possible use of the data is to complete final estimates of $\hat{E}(W_{S,N})$ and $\hat{E}(W_{B,N})$, and then take the difference to calculate the expected market risk premium. Fuller and Hickman (FPE Fall/Winter 1991) take this approach and recommend:

$$\hat{E}(\gamma_{Annual}) = [\hat{E}(W_{S,N})^{1/N} - 1] - [\hat{E}(W_{B,N})^{1/N} - 1]$$

Their data are annual to begin with so they do not need to re-annualize the market risk premium estimate into annual terms. They had T=60 annual data points and assumed N=30, for a final annual market risk premium estimate of 5.731%.

However the practitioner finally uses wealth relative calculations in market risk premium estimates, the essential point remains that wealth relatives must be calculated if the assumed holding period is greater than one data point in length.

The discussion above is also applicable to other compound growth processes, such as dividend growth rates. If the practitioner had 20 years of annual growth rates and wanted to estimate expected growth next year without any other information, the methods discussed above would apply as well. This means that dividend growth for the next year would be estimated by multiplying the arithmetic average of the annual growth rates (plus 1) times the current dividend, but that the dividend 20 years out would be estimated by multiplying the geometric average of the annual growth rates (plus 1) by the current dividend. Dividend growth in between 1 and T would be estimated using one of the more complicated estimators described above.

Citations

Blume, Marshall. "Unbiased Estimators of Long-Run Expected Rates of Return." *Journal of the American Statistical Association*, Vol. 69, No. 347 (September 1974), pp. 634-638.

Blume, Marshall, and Irwin Friend, "Risk, Investment Strategy and the Long-Run Rates of Return." *The Review of Economics and Statistics*, Vol. 56 (August 1974).

Cheng, Pao L., and M. King Deets. "Statistical Biases and Security Rates of Return." *Journal of Financial and Quantitative Analysis*, Vol. 6 (June 1971), pp. 977-994.

Cheng, Pao L. "Unbiased Estimators of Long-Run Expected Returns Revisited." *Journal of Financial and Quantitative Analysis*, Vol. 19, No. 4 (December 1984), pp. 375-393.

Fisher, Lawrence. "Some New Stock Market Indexes." *Journal of Business*, Vol. 39 (January 196), pp. 191-225.

Fuller, Russel J., and Kent A. Hickman. "A Note on Estimating the Historical Market Risk Premium." *Financial Practice and Education*, Fall/Winter 1991, pp. 45-48.

Kritzman, Mark, "What Practitioners Need to Know." *Financial Analysts Journal*, May- June 1993, pp. 14-17.

Hasbrouk, Joel. "On Estimates of Long-Run Rates of Return: A Note." *Journal of Financial and Quantitative Analysis*, Vol. 18, No. 4 (December 198