The previous blog introduced the notion of financial arbitrage and briefly explored the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory (APT) models for pricing an asset (e.g. stock). The CAPM correlates a particular asset with some macroeconomic factor (e.g. inflation or one of the indices) to determine the expected return on the arbitrage. The APT generalizes this 1-dimensional correlation to the case where multiple factors affect the asset price. The applicable formula that covers both cases is

RA = Rfree + β1 ( P1 - Rfree ) + β2 ( P2 - Rfree ) + ...  
&nbsp = Rfree + β1 RP1 + β2 RP2 + ...

where:

RA is the expected rate of return of the asset in question,

Rfree is the rate of return if the asset had no dependence on the identified macroeconomic factors (free rate of return),

βi is the sensitivity of the asset with respect to the *i*th macroeconomic factor, and

Pi is the additional [risk premium](https://en.wikipedia.org/wiki/Risk_premium) associated with the *i*th macroeconomic factor with RPi = Pi - Rfree being the actual risk premium.

Obviously, setting all the βi beyond β1 to be zero in the APT recovers the CAPM.

To use either of these models, the arbitrageur needs to set multiple free parameters (Rfree, RA, βi, Pi) using his judgement based on historical data and some of the aspects of this procedure will be the focus of this post.

For simplicity, we’ll limit the analysis to correlate one stock with one index and we’ll follow the excellent article entitled [CAPM Beta - Definition, Formula, Calculate CAPM Beta in Excel](https://www.wallstreetmojo.com/capm-beta-definition-formula-calculate-beta-in-excel/) by Dheeraj Vaidya for WallStreetMojo. I’ll be adding only a few points here and there just to round out what Vaidya presented but it is worth emphasizing what a fine job he did in his presentation.

The correlation we’ll be exploring is between a company called MakeMyTrip (MMYT ticker symbol) and the NASDAQ Composite (^IXIC ticker symbol). To match, Vaidya’s analysis, we confine our time frame from January 1st, 2012 to October 30th, 2014. Yahoo Finance served quotes under the historical data link that presents itself after entering a ticker symbol (see green ellipse in the figure below)

Graphical user interface, application

Description automatically generated

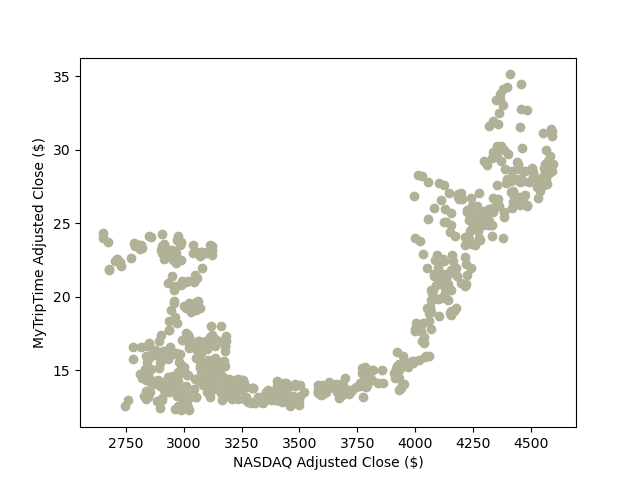
Selecting the time span and downloading the data in CSV format are easy. I read the data for MMYT and ^IXIC in pandas data frames but since the average price of the NASDAQ Composite over that time span was $3563.91 compared to an average of $18.98 for MakeMyTrip, plotting each time series on a common plot won’t work, even with a log scaling. Instead, taking a page from Z-scoring in statistics, I made a plot of the normalized stock price for each listing in which the instantaneous price was divided the average.

Chart, line chart

Description automatically generated

There is no obvious correlation between the two time series. The NASDAQ Composite, more or less, rose steadily during this time span while MakeMyTrip shows a more of a parabolic behavior, with a downward trend during roughly the first third of the time span followed by minimum in the second third, and punctuated by rapid, and often volatile growth, in the third.

These differences in the qualitative evolution of the two assets presents itself even more strongly in a scatter plot showing the adjusted closing price of each asset.



Nonetheless, there is a reasonably good correlation between the two assets in terms of their fractional gain, defined as the difference between two successive days relative to the price of the earlier of the two days (i.e. (pi+1 – pi)/pi where pi is the price of the asset on the *i*th day).

Chart, scatter chart

Description automatically generated

There is a definite but weak positive correlation between the adjusted close of the NASDAQ Composite and MakeMyTrip. A linear regression, computed using numpy’s polyfit routine (order 1), confirmed the same value of 0.9858 for the slope of a linear regression line that Vaidya reported. This value is then the β between MakeMyTrip and the NASDAQ composite for this time span.

But the fun doesn’t stop there. We can use the power of the pandas package to extend Vaidya’s presentation by randomly sampling the data to get an idea of the spread in the value of β based on using different samples due to differences in time span or reporting interval. Running a Monte Carlo with 350 samples each (almost exactly half of the total number of available data points) for N = 10,000 trials gives the following statistics for β:

* the mean was 0.9835
* the standard deviation was 0.1443
* the distribution of β values is normal

Chart, histogram

Description automatically generated

Using the standard techniques of statistical analysis, we might be inclined to report the beta value as β = 0.9835 ± 0.0014 or, said equivalently, β could lie in the range of 0.9807 and 0.9863 with the usual 95% confidence. This uncertainty in the value of β is about 5.7% and this translates directly into a 5.7% uncertainty in the assessment of the assets rate of return. A 5% uncertainty is likely to be a good rule of thumb for the arbitrageur in estimating whether he wants to look further at an asset.

Another source of error that arbitrageur must wrestle with is the value for Rfree, the [risk-free rate of return](https://www.investopedia.com/terms/r/risk-freereturn.asp). According to Investopia.com, while a true risk-free rate of return is only theoretically realizable, the 3-month Treasury note is taken as a good proxy. However, even this ‘sure fire’ investment vehicle sees movement on the secondary market. The [Wall Street Journal has excellent data and plots](https://www.wsj.com/market-data/quotes/bond/BX/TMUBMUSD03M/historical-prices) which show that, at least in recent months, the daily movement of the Rfree can be 5-10%.

The final ingredient in the CAPM model is RP, the additional risk premium associated with the asset. The way this value is set is probably as much an art as a data science question since it not only has to account for the actual financial strengths and weaknesses of the asset but also the market sentiment. If the example in last month’s blog were indicative, values ranging from 2-10% are reasonable. The uncertainty in the estimation of those risk premiums are probably correspondingly larger, maybe in the 20-30% range.

All told, the estimated value for the real rate of return on an asset must account for all of these errors sources. To illustrate this, lets continue on with the comparison of MakeMyTrip with the NASDAQ composite by assuming the following:

* β = 0.9835 with a 1-standard deviation uncertainty of 0.0014
* Rfree = 0.5% with a 1-standard deviation uncertainty of 0.025% (5% of the 0.5% value)
* RP = 2.5% with a 1-standard deviation uncertainty of 0.5 % (20% of the 2.5% value)

With these assumptions, the CAPM rate of return would then be RA = 0.5% + 0.9835\*2.5% = 2.9588%. The corresponding error in that estimation is obtained using the usual propagation of error techniques and takes the value of 0.4943%. This value means that the arbitrageur needs to figure in about 0.5% slop 67% of the time he undertakes this transaction.

All this machinery of linear regression, Monte Carlo simulations, and propagation of error explains the rise of algorithmic trading and the mathematical analysts (so-called ‘quants’) in todays modern market.