Yewno

Quantitative Finance Assignment

Bernardo Bravo

August 16, 2017

Introduction

This assignment covers various topics and common questions in the context of financial engineering and econometrics applied to option valuation, portfolio optimization, algorithmic trading, and risk management. The tasks were both fun and challenging since some of them are problems with several possible solutions and modelling choices.

1 Normal distributions for financial returns

How good is the assumption of normal distributions for financial returns?

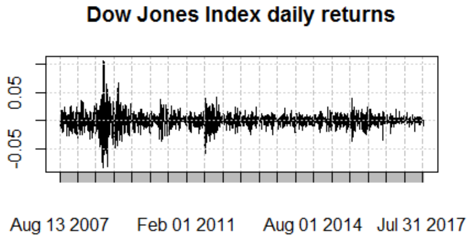
The short answer is: a normal distribution is a bad assumption. Long answer: The distribution returns on financial time series has been widely studied; consensus supported by empirical evidence is that financial time series returns have fat-tailed distributions rather than a lighter-tailed normal distribution, see for example [7]. On this regard in [8] the authors argue that fat tails in financial time series arise because of the long-range volatility correlations.

In the study presented here I use the Dow Jones Industrial Average Index log return data as an example to test the hypothesis of normality. Figure 1 depicts the volatility of the log returns, one can observe the volatility clustering phenomenon, which implies a strong autocorrelation in the underlying volatility. Also, the quantile plot in figure 1 shows a deviation from normality.

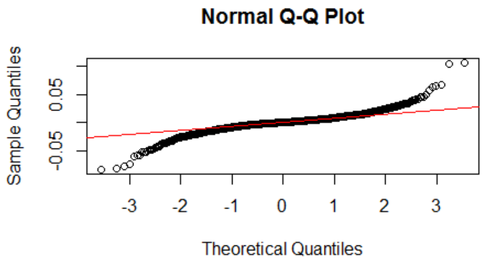
Furthermore, by conducting a Shapiro-Wilk and a Lilliefors test, one can formally reject the null hypothesis of normality supported by the small p-values observed in table 1.

Shapiro-Wilk normality test	Lilliefors (Kolmogorov-Smirnov) normality test
$W = 0.89, p - value < 2.2e - 16$	$D = 0.11, p - value < 2.2e - 16$

Table 1: Normality tests



(a) Stock index returns



(b) Quantile plot

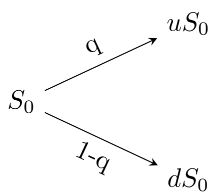
Figure 1: Dow Jones return data and quantile plot

2 Option Pricing

Simulate options pricing using the binomial model. Why would you use this model instead of the Black-and-Scholes model?

Under the binomial option pricing model[1] we assume that at every time step the stock value can either move up or down.

To explain the idea let us consider a one-step only case in which the starting value of the stock is S_0, then in the next step (time T) it either goes up by a factor $u > 1$ or down by a factor $d < 1$, we will also think of a probability q of going up or $1 - q$ of going down, the precise value of q will appear magically in a few steps below; putting all those ideas in a diagram we have:



[1]A standard reference for this is [9]. My R implementation of the option price and tree construction follows [2] closely.

Also, we assume there is a risk free interest rate r, so that \$1 invested now will grow to e^{rT}. Now we'll consider a call option with strike K, the payoff is below:

$$
\begin{tikzcd}
 & (uS_0 - K)^+ \\
C_0? \arrow[ur, "q"] \arrow[dr, "1\text{-}q"'] & \\
 & (dS_0 - K)^+
\end{tikzcd}
$$

We want to calculate the price of the call at time 0 C_0, the idea is to consider a hypothetical portfolio that has the same payoff as our option and, in an arbitrage-free world, the cost of such portfolio and the cost of the option should be the same at time 0.

The hypothetical portfolio has a units of the stock and b dollars in the bank, the evolution of the portfolio is as follows:

$$
\begin{tikzcd}
 & auS_0 + be^{rT} \\
aS_0 + b \arrow[ur, "q"] \arrow[dr, "1\text{-}q"'] & \\
 & adS_0 + be^{rT}
\end{tikzcd}
$$

In order to determine the start up cost of the portfolio (which will be the cost of the call) we need to find a and b such that:

$$auS_0 + be^{rT} = (uS_0 - K)^+ \tag{1}$$

$$adS_0 + be^{rT} = (dS_0 - K)^+ \tag{2}$$

Solving for a and b in that system of equations, plugging them into $C_0 = aS_0 + b$ and rearranging terms we obtain the following formula for the call price:

$$
\begin{aligned}
C_0 &= e^{-rT}\left[\frac{e^{rT} - d}{u - d}(uS_0 - K)^+ + \frac{u - e^{rT}}{u - d}(dS_0 - K)^+\right] \\
&= e^{-rT}\left[q(uS_0 - K)^+ + (1 - q)(dS_0 - K)^+\right]
\end{aligned} \tag{3}
$$

If we look closely we can convince ourselves that the term inside the parentheses is the expected value of the option using a probability measure q, the term e^{-rT} discounts that expectation to present value.

The way to price options considering more time steps is very similar, we build a tree of stock prices with more branches and we also construct an option-price tree; then we start at the branches at the end (child nodes) and to get the value of the father node we discount the expected value using the child nodes and using the probability q; we work backwards all the way to the initial node, that's our option price at time 0.

Now with the R implementation we can calculate the price of a call and put option for a 2 step binomial tree; using the following parameters: $S_0 = 10$, $K = 10$, $r = 10\%$, $u = 1.236$, $d = 0.808$.

First we show the stock price tree:

$$\begin{tikzcd}
 & & 15.28 \\
 & 12.36 \arrow[ur] \arrow[dr] & \\
\$10 \arrow[ur] \arrow[dr] & & 10 \\
 & 8.08 \arrow[ur] \arrow[dr] & \\
 & & 6.54
\end{tikzcd}$$

Now the call price tree:

$$\begin{tikzcd}
 & & 5.285 \\
 & 2.85 \arrow[ur] \arrow[dr] & \\
\$1.538 \arrow[ur] \arrow[dr] & & 0 \\
 & 0 \arrow[ur] \arrow[dr] & \\
 & & 0
\end{tikzcd}$$

The price of the call is \$1.538, even if I do not show the put tree calculation here, the put price calculated in R is \$0.586. One way to check if the implementation is correct is by checking a famous no arbitrage relationship called put-call parity shown below, plugging the numbers in we verify everything is ok:

$$call - put = S - Ke^{-r} \tag{4}$$

I would use binomial trees instead of Black-Scholes to price some path-dependent options like Americans, for those there are no closed solution for the differential equation involved under the Black-Scholes pricing framework. I would also price options that have dividend payments in specific time periods, binomial trees would help to be more accurate.

What makes path-dependent options more difficult to price is the fact that could be exercised depending on the possible stock price paths, hence the optimal exercising strategy also needs to be inferred to preclude the possibility of arbitrage; as opposed to plain vanilla puts and calls in which we only need to know the end price of the stock, for plain vanilla puts and calls I like the Black-Scholes formula since it's fast.

As we can easily infer, binomial trees are somewhat cumbersome to implement and to estimate the price of more complex path-dependent options like Asian ones for a long time period can be hard; it is known that simulation methods are preferred in those cases.

An interesting result that is worth mentioning is that the Black-Scholes formula can be derived by taking the limit when the time steps in the binomial tree become infinitely small, moving from a discrete time framework to continuous time; the limit of the binomial distribution in the tree converges to a lognormal distribution in the continous case.

3 Equity Portfolio Risk

Estimate Equity Portfolio Risk without using asset prices

A typical approach to estimate the risk of an equity portfolio is by estimating the variance of the portfolio return under the Capital Asset Pricing Model (CAPM) framework, good lecture notes for this here [10].

Let us consider a portfolio with two stocks, from different sectors for a better diversification: Pfizer Inc from the Pharmaceutical sector and Cisco Systems Inc. from the Technology and Telecommunications Equipment Sector.

Let w_P denote the proportion of our wealth invested in Pfizer stock and w_C the proportion allocated to Cisco, then the variance of the return of the portfolio $R(w)$ is calculated as:

$$Var[R(w)] = w^T * \Sigma * w \tag{5}$$

where Σ denotes the covariance matrix and the vector w has the allocation weights.

Since I can not use the stock prices for this exercise I will use information on the profitability of those companies since it is likely to contain information on the future performance of the firms and their inherent risk.

In particular, I will use public profitability ratios time series for both companies to construct a profitability index by means of Principal Component Analysis. The data I used is quarterly published by both companies, from 2007Q1 to 2017Q1, and I used the Macrobond data aggregator to extract the data.

The profitability ratios considered for both Pfizer and Cisco were Effective Tax Rate, Gross Profit Margin, Net Margin, Net Sales to Total Assets, Pre-tax Profit Margin and Sales to Working Capital Ratio. I threw all them into a PCA and called the first principal component profitability index, getting one for Cisco and one for Pfizer.

PCA results are shown in table 2. We observe that for Pfizer, the first component preserves 45% of the variance while for Cisco it is 49%. Even if those are relatively low values the profitability index constructed in this way has information on all the ratios mentioned above, as opposed to using just one or a subset of them.

The covariance matrix estimated from the profitability indices is shown below, we notice that since both sectors are relatively unrelated the covariance is low:

	Pfizer					
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Standard deviation	1.65	1.07	1.00	0.78	0.68	0.26
Proportion of Variance	45%	19%	17%	10%	8%	1%
Cumulative Proportion	45%	64%	81%	91%	99%	100%
	Cisco					
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Standard deviation	1.72	1.29	0.94	0.68	0.17	0.04
Proportion of Variance	49%	28%	15%	8%	0.5%	0.02%
Cumulative Proportion	49%	77%	92%	99%	100%	100%

Table 2: PCA for a profitability index construction

Pfizer weight	Cisco weight	Risk
0%	100%	3
10%	90%	2.4
20%	80%	2
30%	70%	1.7
40%	60%	1.5
50%	50%	1.4
60%	40%	1.4
70%	30%	1.6
80%	20%	1.9
90%	10%	2.2
100%	0%	2.7

Table 3: Risk of different allocations for a Cisco and Pfizer portfolio

$$\Sigma = \begin{bmatrix} 2.73 & -0.02 \\ -0.02 & 2.96 \end{bmatrix}$$

The risk measurement results for different weights are presented in table 3. We observe that the most conservative portfolio is one that has equal weights for both stocks, hence diversifying the portfolio equally among the Pharmaceutical and Technology sector decreases risk. Another thing we can note is that Pfizer is less risky than Cisco.

4 Predict Macro Indicators

Use freely available data from the web to predict/explain macroeconomic indicators; Financial/Economic/Fundamentals data are not allowed

Macroeconomic data is usually available with a lag and time series are often restated. Google Trends provides a wealth of data that could be useful to enhance predictive models involving macroeconomic variables. That information has already been used to predict car sales in [4] and also to predict Unemployment Rates in [1].

Model	AIC	RMSE	Adj.Rsq
A $hpi \sim hpi\_lag1 + google$	425.6	1.36	99.5%
B $hpi \sim hpi\_lag1 + google + google\_lag1$	427.6	1.36	99.5%
C $hpi \sim hpi\_lag1$	495.8	1.83	99.1%

Table 4: Candidate model comparison

For the empirical study described here I decided to predict the US House Price Index for its relevance during the recent financial crisis, where a drop in this index was also linked to a severe decrease in homeowner wealth, as well as increases in unemployment among other adverse effects.

The data used for the study was the seasonally adjusted House Price Index monthly time series from Case Shiller CoreLogic as well as data from Google Trends, the search term used was: zillow + cyberhomes + "home prices" + "home refinance" as suggested in [6].

Three linear regression models were proposed, results are summarized in table 4. The base model C is an $AR(1)$, on the other hand candidate models A and B include contemporaneous and lagged values of the Google trend time series.

Overall, the addition of the Google variable presents a decrease in Root Mean Squared Error (RMSE) even if the fit of the model measured by the Adjusted R squared did not change significantly; the Akaike Information Criterion (AIC) also shows that models including the google variables are superior.

Furthermore, both the lagged value of HPI and the contemporaneous Google variable were statistically significant whereas the lagged value of the Google variable is not (these results were not presented in this report for succinctness).

My results are also in line with results presented in [6] where the authors performed a Granger causality test; they find out that the Google variable "granger causes" the HPI while HPI does not "granger cause" Google.

Another clear extension for this work is to try to forecast HPI by substituting all its useful predictors , for example GDP, commercial real estate, interest rates, by their Google Trends doppelgängers. One way to find more bespoke predictors could be to use Google correlate or the Yewno knowledge graph.

5 Smart Beta Strategy

Implement one Smart Beta strategy and discuss pros and cons compared to a chosen benchmark

The Smart Beta Guide by Black Rock [3] definition reads that Beta strategies generally take long positions, aim to improve returns, reduce risks and enhance diversification; targeting factors such as macro, sector or region factors (among others) that drive returns.

Following that definition I decided to implement a typical quadratic optimization problem aiming at maximizing the returns, penalizing risk, constrained for long only positions seeking to identify factors that drive larger returns among a diverse pool. Specifically I considered the opportunity to invest in several Mid and Large Cap regional indices of the MSCI family focused in Emerging and

Frontier Markets for the following regions: Asia, Africa, Latin America and a global EM index.

The optimization problem is formalized below, following a consistent notation with problem 3: w denotes the allocation among the available assets, multiplied by the expected returns mu; the covariance matrix Σ in the second term has a negative sign to penalize risk; λ is the risk aversion parameter which I will fix to 1; while the constraints guarantee that the resulting portfolio takes only long positions and that the allocation adds up to 100%.

$$\begin{aligned} \max_{w} \quad & \mu \cdot w - \lambda \cdot w^T \cdot \Sigma \cdot w \\ \text{subject to} \quad & \\ & \mathbf{1} \cdot w = 1 \\ & w \geq 0 \end{aligned}$$

The historic data set was divided into a test set from June 2009 to Dec 2016 to calculate the parameters μ, Σ and find the optimal allocation w. The rest of the historic data set is for investing, the strategy is to buy the portfolio according to the optimal allocation in Jan 2017 and check the returns obtained by Aug 2017; the benchmark for comparison purposes is to allocate everything on the S&P 500 index during the same time period.

The portfolio allocation by the described procedure was:

- Latin America 4.5%
- Asia 95.5%
- Africa 0%
- EM 0%

Return obtained with optimal portfolio was **23.59%**! while the return investing only in S&P 500 was **8.13%**.

A well known caveat of using an optimization approach of this sort is that the procedure usually finds extreme portfolios corresponding to the corners of the solution space which, as we saw in the example above, can pick a portfolio not very well diversified. One way to adress this is to change the constraint in order to ask for strictly positive allocations for all assets, by requiring a min investment of 10% on each asset I got a better diversified portfolio that got a 21.9% return:

- Latin America 10%
- Asia 95.5%
- Africa 10%
- EM 10%

Another thing to keep in mind is that this kind of problems could be numerically unstable, in the sense that relatively small changes in the inputs can completely change the solutions. In particular I observed that using a much smaller value of λ changes to a 100% allocation to EM!

Besides the problems outlined above there is also the need to include tax, fees, turnover constraints, etc.

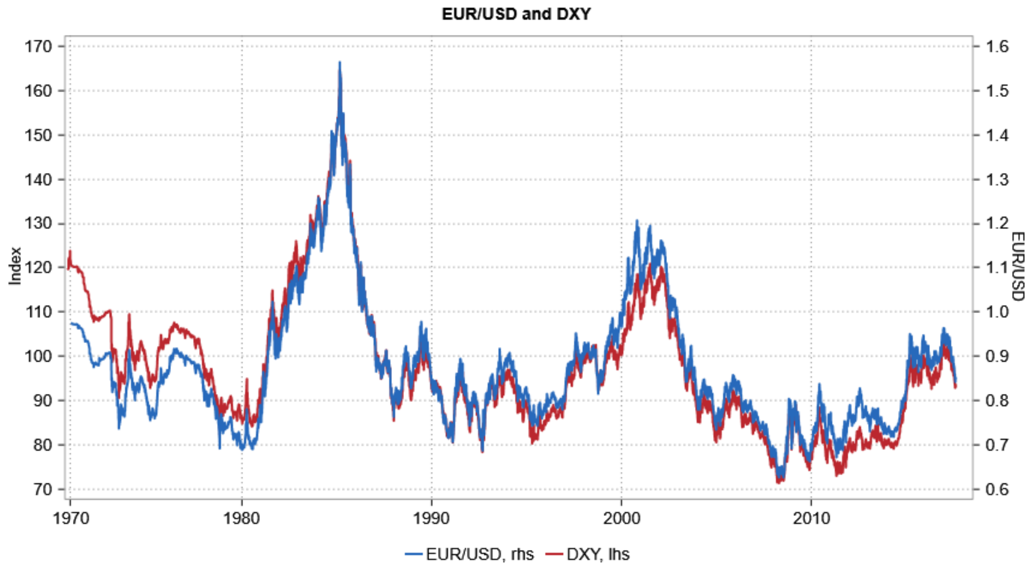


Figure 2: EUR/USD and DXY history

6 FX Pairs Trading

6.1 Data

Suggest one data source that might be useful to explain or predict the FX market

For this study I decided to use the US Dollar Index (DXY) from the Intercontinental Exchange (ICE) which represents the value of the US dollar with respect to a weighted basket of currencies, and the EUR/USD[2]. As we can see in figure 6.1 both time series seem to be tied to each other throughout its history.

Another idea would be to use the prices of oil which are well known to be a driver of FX rates particularly for economies with an important dependency to this commodity[3]. However, given that prices of oil heavily depend on political pressure by the main oil suppliers it is not very likely to be a good neutral strategy as the use of FX rates which moves more freely depending on trading relationships and other macro and financial conditions.

For this problem I also tried different data sets like the Citi Terms of Trade Indices for Commodities and Google Trends data. Despite the fact that some of them were cointegrated with a an FX rate it is not entirely clear to me how that information could be used for a trading strategy since it basically works by longing one asset and shorting another. There is no tradable asset linked to the terms of trade or Google Trends data to my knowledge.

[2]Literature on FX pairs usually mentions the use of EUR/USD and GBP/USD as cointegrated pairs, it would be interesting to see how this relationship changes as post-Brexit policies unfold.

[3]For example, I observed that the Norweigan Krona and oil prices are cointegrated.

Variable	p-value
Δdxy	0.01
$\Delta eurusd$	0.01

Table 5: Augmented Dickey Fuller test for the difference series

Augmented Dickey-Fuller Test	Phillips-Perron Unit Root Test
$D - F = -2.49, p - value = 0.37$	$D - F = -3.82, p - value = 0.018$

Table 6: Stationarity tests on the spread

6.2 Analytics

Derive and discuss relevant analytics from this data source

Going back to the DXY and EUR/USD case, I decided to follow the Engle-Granger approach as opposed to Johansen's since the former is more transparent. The first step in this process is to check that the first differences of the time series to use are stationary. I performed an Augmented Dickey Fuller Test on the first order differences of *eurusd* the EUR/USD spot rate and the *dxy* ; the p-values are shown in the table 5, which suggest to reject the null hypothesis, concluding thet the first order differences are stationary.

6.3 Co-Integration

Determine whether your proposed analytics are co-integrated with currency pairs

The candidate model with its corresponding estimated coefficient is shown below:

$$dxy \sim 108 * eurusd \tag{6}$$

And we define the following spread to be used for the trading strategy:

$$spread = dxy - 108 * eurusd \tag{7}$$

We now conduct a couple of stationarity tests on the residuals of the regression model used above, to check that the spread is stationary. Results are shown in table 6.3, even if the Dickey Fuller Test did not give evidence to reject the null hypothesis Phillips-Perron did. Either way, by construction the DXY is strongly bound together to the EUR/USD by a specific weight.

6.4 Pairs Trading Strategy

Describe and implement a pairs trading strategy exploiting your analytics

The spread time series is shown in figure 3, since it is a stationary time series the trading strategy will trigger an action whenever it deviates by an amount, say $\delta = 0.1$ from its mean value $\mu = 0.215$. The strategy is to buy the portfolio

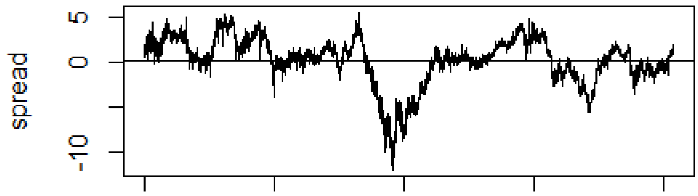


Figure 3: Stationary spread and its mean

when it's cheap (sell it if it gets expensive) and unwind the position when the spread gets back around its usual levels:

$$\text{If } spread < \mu + \delta \Longrightarrow \text{Buy } dxy \ \ \text{Sell } 108 * eurusd$$
$$\text{If } spread > \mu + \delta \Longrightarrow \text{Sell } dxy \ \ \text{Buy } 108 * eurusd$$

For more realistic and detailed trading strategies using cointegration the book [11] seems to be a standard good reference. Also in [5] the authors propose pairs trading strategies from a theoretical point of view by means of models like the Ornstein-Uhlenbeck process whci sounds interesting.

Discussion and further work

Even if Google Trends provides interesting and useful time series, those might actually be very short, compared with any other financial time series, this caveat should be taken into consideration when using Google Trends data to adjust models.

For the methodologies discussed here it would be interesting to see how the use of information extracted from the Yewno knowledge graphs would help to improve the models and prediction quality. In general extracting data from texts using natural language processing techniques seems to be the most promising venue to get valuable information from news, books, central bank statements, etc. to predict financial and macro variables.

References

[1] N. Askitas and K.F. Zimmermann. "Google Econometrics and Unemployment Forecasting". In: Discussion Papers (2009). URL: `https://books.google.com/books?id=2PXrkQEACAAJ`.

[2] *Binomial European Option Pricing in R.* `http://linanqiu.github.io/2015/10/26/Binomial-European-Option-Pricing-in-R/`. Accessed: 2017-08-12.

[3] *Black Rock Smart Beta Guide.* 2017.

[4] Hyunyoung Choi and Hal Varian. "Predicting the Present with Google Trends". In: *The Economic Record* 88.s1 (2012), pp. 2–9. URL: http://EconPapers.repec.org/RePEc:bla:ecorec:v:88:y:2012:i:s1:p:2-9.

[5] Binh Do, Robert Faff, and Kais Hamza. *A new approach to modeling and estimation for pairs trading.* 2006.

[6] Rajendra Kulkarni et al. "Forecasting Housing Prices with Google Econometrics." In: GMU School of Public Policy Research Paper (). DOI: http://dx.doi.org/10.2139/ssrn.1438286. URL: https://books.google.com/books?id=2PXrkQEACAAJ.

[7] H. Eugene Stanley Rosario N. Mantegna. *An introduction to econophysics: correlations and complexity in finance.* 0th ed. Cambridge University Press, 2000. ISBN: 9780521620086,0521620082.

[8] G.M. Viswanathan; U.L. Fulco; M.L. Lyra; M. Serva. "The origin of fat-tailed distributions in financial time series". In: *Physica A: Statistical Mechanics and its Applications* 329 (1-2 2003). DOI: 10.1016/s0378-4371(03)00608-3.

[9] Steven E. Shreve. *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model: Binomial Asset Pricing Model.* New York, NY: Springer-Verlag, 2003. ISBN: 0387401008 9780387401003.

[10] Karl Sigman. *Capital Asset Pricing Model (CAPM) Lecture Notes.* 2005.

[11] G. Vidyamurthy. *Pairs Trading: Quantitative Methods and Analysis.* Wiley Finance. Wiley, 2011. ISBN: 9781118045701. URL: https://books.google.com/books?id=SQLU014X7JwC.