1 Non-normality of Market Returns

The assumption that stock returns are normally distributed is widely used, implicitly or explicitly, in theoretical finance. However, the evidence against this assumption has been mounting since the pioneering article by Mandelbrot (1963), which argued that price changes can be characterized by a stable Paretian distribution. Commonly, the non-normality of returns has been empirically assessed in two different ways:

1) Fat Tails (Leptokurtosis): The first form of non-normality relates to observing extreme returns in greater magnitude and with a higher probability than implied by the normal distribution. In particular, normal distribution is not suitable for data with higher frequencies, such as daily and intraday data, because of the extreme events. Figure 1, 2 shows the distribution of the S&P 500 returns with 5 minutes and daily frequencies. The blue line (empirical) is taller at its apex and shows a higher density at the extreme (i.e. leptokurtosis). In particular, the higher density at the left tail indicates a higher probability of extreme negative events. After applying the Jarque-Bera test to the empirical distribution, it is rejected the null hypothesis of normality in the distribution of 5-minute returns and 1-day returns (p-value = 0).

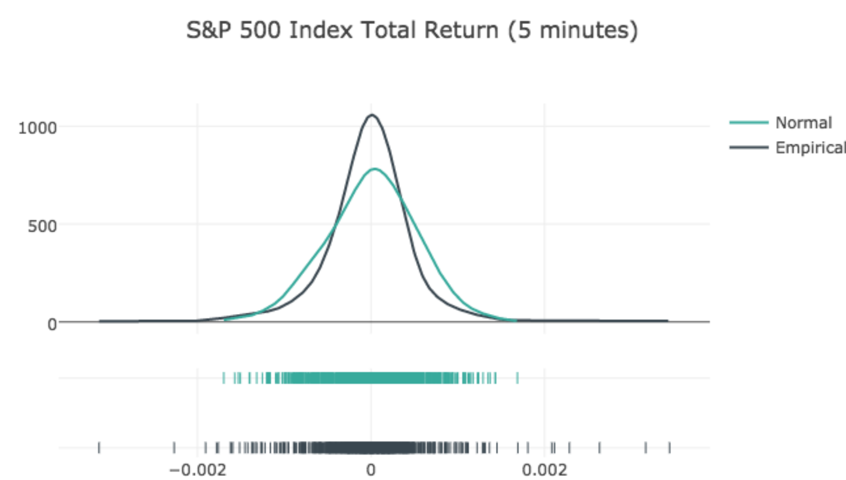

Figure 1: S&P 500 Distribution 5 minutes-return, 10 days sample

2) Serial correlation in asset returns: this occurs when one periods return is correlated to the previous periods return, inducing dependence over time.

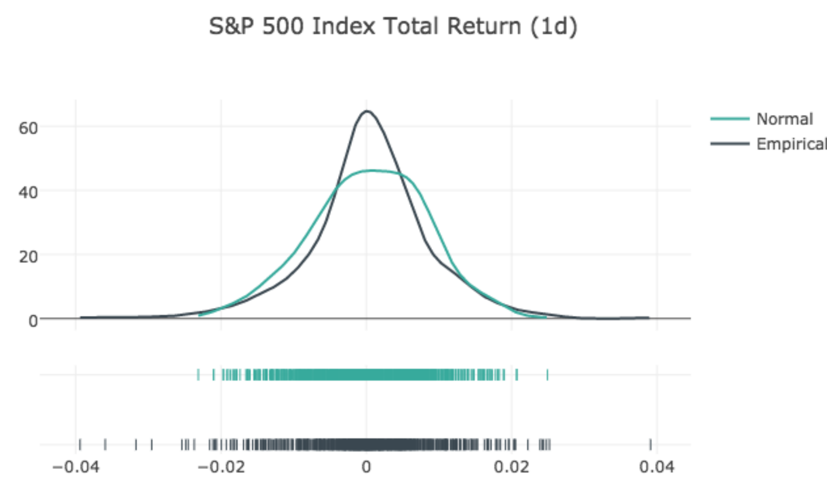

Figure 2: S&P 500 Distribution 1day-return, 10 years sample

1.1 Python Files

The python file to get the prices from Yahoo Finance and compute the distributions is found in:

-Distribution-
1) 01_code/asset_returns.py

Distribution and plots -Plotly package needed-
2) 01_code/asset_returns_figure.py

Note: No inputs/parameters are nedded

2 Binomial Option Pricing Model

2.1 Binomial Model

The binomial option pricing model is based on a formulation for the underlying's price process in which the asset, in any time period, can move to one of two possible prices: 1) price goes up by a factor of u or 2) goes down by a factor of $d = 1/u$ (figure 3). These factors are computed with the assumption that daily continuous growth rates for the underlying stock are normally distributed around zero with some variance σ^2. The first step in pricing options using a binomial model is to create a tree of potential future prices of the underlying asset(s) (figure 3).

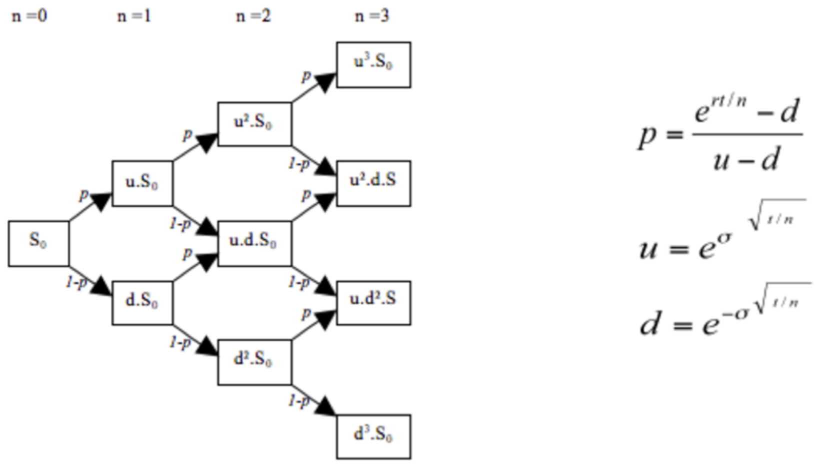

Figure 3: Binomial Options Pricing Model tree, with variance σ^2

The goal of the binomial options pricing model is to compute the price of the option at each node in this tree (figure 3) to finally get the value at the root of the tree. Having said that, the next step is to compute the exercise value at each node of the tree, using the simulated prices:

$$EV_{call} = max(0, S_n - K) \tag{1}$$

$$EV_{put} = max(0, K - S_n) \tag{2}$$

where K is the strike price of the option and S_n the underlying price at each node.

Subsequently, the expected value of the option price in a timestep is given by

European

$$E_{call/put}[P] = pEV_{up} + (1 - p)EV_{down} \tag{3}$$

American

$$E_{call/put}[P] = max\left(pEV_{up} + (1 - p)EV_{down}, EV_{current}\right) \tag{4}$$

where p is the probability of the stock price going up (1-p going down).

$$p = \frac{e^{r\Delta t} - d}{u - d} \tag{5}$$

To get the current price, it is still needed the backwards-discounted price $e^{-r\Delta t}E[P]$, with r as the risk free discount rate. Finally, the price of the option is given by the current price of the root of the three.

In the case of an European Option, the generalization is given by:

$$X = (B(,))^N \sum_{j=0}^{N} \left(\binom{N}{j} p^j (1-p)^{N-j} EV(S_0 \cdot u^j d^{N-j}) \right) \tag{6}$$

2.2 Python Files

The python file to price an option using the binomial model can be found:

1) binomial_option_model

Inputs: Run python file using the following inputs:
02_binomial_option_model.py -S 10 -K 10 -N 5 -s 0.2 -r 0.1 -d 0.2 -f call -o True

-S: Current price of the underlying asset
-K: Strike of the option
-N: Number of timesteps until expiration
-s: Volatility
-r Risk-free discount rate
-d Timestep
-f Type of option (put/call)
-o Style of option (European: True, American: False)

Output: Binomial Option Price

Note: No inputs (dataset) are nedded

2.3 Advantages of Binomial Model over Black-Scholes

The main feature of the Binomial model is the posibility to check the evolution of the option at every timestep. In this sense, the most important advantages are:

- Price American Options: In American Options, where an early exercise point is found it is assumed that the option holder would elect to exercise, and the option price can be adjusted to equal the Exercise Value (EV) at that point. Binomial Model allows to incorporate the adjusted Exercise Value at each timestep, see equation (4)
- Term-Structure of Rates and Variance: The binomial model uses different input parameters; among them the interest rate r and volatility σ. Commonly, these parameters are assumed to be constant. However in the

Binomial Model, it is possible to incorporate the term structure at each timestep, updating also the growth and decay factors.

3 Estimate Equity Portfolio Risk without using asset prices.

3.1 Problem Formulation

Standard practices estimates common risk factors using the joint realized return distributions of past stock prices and marketwide returns (stock returns beta). The traditional model, known formally as the capital asset pricing model (CAPM) uses only one variable to describe the returns of a portfolio or stock with the returns of the market as a whole:

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f) \tag{7}$$

The β value serves as an important measure of risk for individual assets (portfolios).

Fama and French (1996, 2006) introduced a Three Factor Model that added two factors to CAPM in order to reflect a portfolio exposure to (i) small caps and (ii) stocks with a high Book-to-market ratio.

$$r = R_f + \beta_3(K_m - R_f) + b_s \cdot SMB + b_v \cdot HML + \alpha \tag{8}$$

The three factor β_3 is analogous to the classical β but not equal to it, since there are now two additional factors to do some of the work.

If the asset prices are available, both models give insights about the risk of a stock (portfolio) depending on the level of confidence in the market index. Alternative approaches are built on the strong financial theory of the CAPM but also represent accounting-based estimates. **Accounting betas** work as risk measures that are founded on modern finance theory using financial statements information.

3.2 Model

The proposed approach estimates equity risk according to the following empirical model:

$$(ROE_{it} - Ref_t) = \alpha_i + \beta_{im}(ROE_{market,t} - Ref_t), \tag{9}$$

where ROE_{it} is firm i's quarterly reported accounting return on equity (ROE) at time t, Ref_t is the 10Y treasury bill rate at time t, and $ROE_{market,t}$ is the quarterly reported ROE estimated for the market. For this model, the ROE_{market} is estimated as the average of the reported ROE for the n firms with highest market capitalization that are in the same industry than firm i.

$$ROE_{market,t} = \sum_{i \in A} ROE_{i,t}/N_A \tag{10}$$

For simplicity, the model is applied to Technology firms using quarterly data from 2004. The parameters are stimated using a linear regression with intercept.

3.3 Data

In order to get the ROE by firm, a web scrapper was created. The program extracts the information from YCharts. It was decided this website because the data is available quarterly since 2004. The treasury yield is manually obtained from U.S Department of the treasury

3.4 Results

The model is tested with two firms: Apple Inc. (AAPL) and Alphabet Inc. (GOOGL). The initial market index is composed by the 50 firms with highest market capitalization that are in the Technology industry. Subsequently, it was decided to filter 18 firms that does not have information in all the quarters included in the analysis. Then, the final market index is composed by 32 firms.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.896
Model:                            OLS   Adj. R-squared:                  0.894
Method:                 Least Squares   F-statistic:                     421.4
Date:                Tue, 28 Mar 2017   Prob (F-statistic):           1.03e-25
Time:                        00:58:09   Log-Likelihood:                -153.59
No. Observations:                  50   AIC:                             309.2
Df Residuals:                      49   BIC:                             311.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.5069      0.025     20.527      0.000       0.457       0.557
==============================================================================
Omnibus:                        0.577   Durbin-Watson:                   0.130
Prob(Omnibus):                  0.749   Jarque-Bera (JB):                0.710
Skew:                           0.185   Prob(JB):                        0.701
Kurtosis:                       2.548   Cond. No.                         1.00
==============================================================================
```

Figure 4: Apple Inc. Regression Results

After running the linear regression using the formula (9), the β is close to 0.50 for both firms, which suggest lower equity risk than the benchmark (Technology Industry). Finally, It can be computed the $\beta_{portfolio}$ of the portfolio as the weighted average of each β_i. Assuming an equal weight portfolio, the $\beta_{portfolio} = 0.49$.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.809
Model:                            OLS   Adj. R-squared:                  0.805
Method:                 Least Squares   F-statistic:                     208.0
Date:                Tue, 28 Mar 2017   Prob (F-statistic):           2.93e-19
Time:                        00:58:09   Log-Likelihood:                -168.71
No. Observations:                  50   AIC:                             339.4
Df Residuals:                      49   BIC:                             341.3
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.4799      0.033     14.421      0.000       0.413       0.547
==============================================================================
Omnibus:                        5.414   Durbin-Watson:                   0.069
Prob(Omnibus):                  0.067   Jarque-Bera (JB):                2.132
Skew:                          -0.066   Prob(JB):                        0.344
Kurtosis:                       1.997   Cond. No.                         1.00
==============================================================================
```

Figure 5: Alphabet Inc. Regression Results

3.5 Python Files

The python file to run the model (included the scrapper) is found in:

01_code/03_equity_risk.py

Inputs:
Data_1: 01_code/dataset/treasury_q.csv
Data_2: 01_code/dataset/companylist.csv

Note: No parameters are nedded

4 Explaining GDP with Amazon trends

4.1 Problem Formulation

Measuring GDP accurately on a regular basis helps policy makers, economists, and business leaders determine appropriate policies, research direction, and financial strategies. While headline GDP figures are typically reported each quarter, it is helpful for government and business organizations to refer to updated numbers in the weeks between.

The four components of gross domestic product are personal consumption, business investment, government spending and net exports. In 2016, the U.S. GDP was 69 percent personal consumption, 16 percent business investment, 18 percent government spending and negative 3 percent net exports. [1]

[1]https://www.thebalance.com/components-of-gdp-explanation-formula-and-chart-

In the last years, one of the most important drivers of consumption growth is e-commerce. In this sense, e-commerce indicators could provide important information to explain/forecast macroeconomic indicators, in particular consumption and GDP.

4.2 Model

The original hypothesis was that GDP can be explained by Amazon Sales. As historical information of Amazon sales in not freely available, It was decided to replace this variable with google trends. Figure 6 describes the evolution of the quarterly Google Index and the GDP rate from Q1 2008 until Q4 2016. The GDP and the Google Index seem to behave in a similar manner.

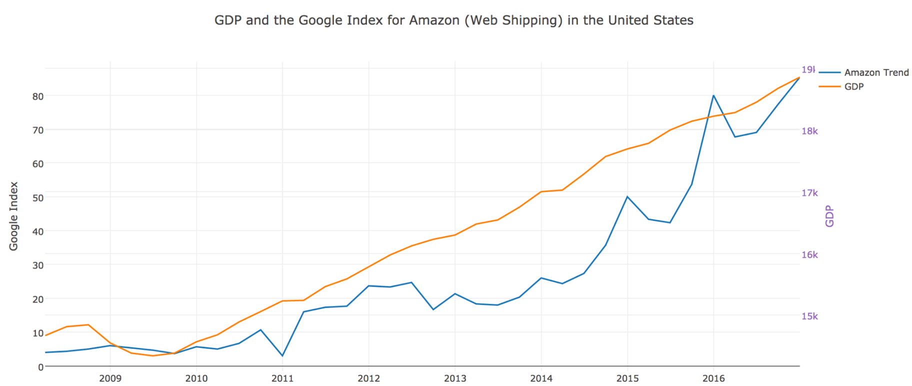

Figure 6: GDP vs Google Index

4.3 Data

The primary data sources are the Google Trends website for the google index and FRED for the GDP.

In order to get the Amazon trend, the input search term is **Amazon**, filtering by U.S Country and Google Shopping. The google index information is reported weekly. Then, the average for each quarter is computed in order to match the GDP data frequency.

3306015

4.4 Results

The results of the regresion are shown in Table 7. It can be seen that $R^2 = 0.85$ which is an indicator that GDP can be explained by Amazon Trends. Given that the p-value is equal to 0, the $\beta = 55$ is statistically significant.

As further research, It is proposed to incorporate Amazon trends to forecast GDP. The suggested approach would use ARIMA models; first using GDP, and then incorporating the exogenous variable Amazon trends. The expected output would show predictive performance improvements incorporating the exogenous variable.

OLS Regression Results

Dep. Variable:	GDP	R-squared:	0.849
Model:	OLS	Adj. R-squared:	0.844
Method:	Least Squares	F-statistic:	190.9
Date:	Tue, 28 Mar 2017	Prob (F-statistic):	1.65e-15
Time:	12:16:27	Log-Likelihood:	-278.47
No. Observations:	36	AIC:	560.9
Df Residuals:	34	BIC:	564.1
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.48e+04	142.978	103.487	0.000	1.45e+04	1.51e+04
Amazon	55.2023	3.995	13.818	0.000	47.083	63.321

Omnibus:	6.959	Durbin-Watson:	0.473
Prob(Omnibus):	0.031	Jarque-Bera (JB):	2.692
Skew:	0.336	Prob(JB):	0.260
Kurtosis:	1.841	Cond. No.	53.9

Figure 7: GDP vs Amazon Trend Linear Regression

4.5 Python Files

The python file to run the model (included the scrapper) is found in:

01_code/04_GDP_amazon.py

Inputs:
Data_1: 01_code/dataset/input_amazon.csv

Note: No parameters are nedded

5 Smart Beta implementation

5.1 Smart Beta

Fundamentally, smart beta has its roots in factor investing, itself the subject of long-standing academic research. Its roots go back as far as the 1960s, when William F. Sharpe identified risk factors as the primary drivers of equity returns.

Broadly, it can be split the equity smart beta universe into two high-level categories:

- Thematic approaches.- Preference for simple, transparent strategies with a strong and intuitive economic or investment rationale.
- Systematic approaches.- Opportunities for systematic smart beta approaches arise from a range of factors; for example, investor heterogeneity and systematic mispricing, and some structural issues associated with market capitalisation approaches

    Equally weighted approach

    Economically weighted approach

    Risk weighted approaches

5.2 Model

In order to implement a Smart Beta Strategy, It will be applied the Markowitz Mean-Variance Portfolio Theory with a subset of n U.S. equity securities that compose the S&P 500 index.

The goal is then to choose the portfolio weighting factors optimally. In the context of the Markowitz theory an optimal set of weights is one in which the portfolio achieves an acceptable baseline expected rate of return with minimal volatility.

The output weights obtained from Markowitz are used to build a portafolio, which will be tested using the market prices of the following year. The return, volatility [2] and performance (Sharpe ratio) of the portfolio is compared with the S&P 500. Additionally, It is assumed that the portfolio is not rebalanced during the tested year. In order to measure diversification, it will be tested portfolios with $n \in [10, 50, 100]$ equity securities

5.3 Data

The daily prices equity securities are obtained from Yahoo Finance. In order to compute the optimal weights, 5 years of information are used (2011-03-20

[2]The volatility is computed taking the standar deviation of daily returns and then annualized by a factor of $\sqrt{260}$

to 2016-03-20). The portafolio with the optimal weights is tested with the information from 2016-03-20 to 2017-03-20.

5.4 Python Files

The python file to run the model is found in:

01_code/05_smart_beta.py

Inputs: List of companies that compose SP500
Data_1: 01_code/dataset/SP500_companies

Note: Depending of the n value, The program can take from 1 to 15 minutes.
Note 2: Required package portfolioopt

5.5 Results

The results are shown in figure 8. Based on the sharpe ratio, the best performance is achieved by the benchmark (S&P 500). As expected, including more assets in the analysis decreased the risk of the portfolio (volatility). However, diversification impacts negatively the performance of the strategy. In this sense, the conclussion is that the advantage of this Smart Beta strategy is oriented to reduce the risk.

	Benchmark	Portfolio (10)	Portfolio (50)	Portfolio (100)
Return	16	14	8	6
Volatility	10	10	9	9
Sharpe	1.54	1.39	0.93	0.65

Figure 8: Smart Beta results with $n \in [10, 50, 100]$

6 FX market

6.1 Dataset

I propose two datasets.

- Intraday evolution of FX (Ticks, Volume, Order books): Some Datasets are freely available. However, most of them does not have the best quality. In order to do experiments, a good data source seems to be Free-forexdata. The dataset includes 2 days of intraday information of different currency pairs. Other resources as Bloomberg, allows to download Tick information of prices, volume and order books.

- Non-farm payroll (NFP) news.- The NFP report causes one of the consistently largest rate movements of any news announcement in the forex market. In this sense, incorporate the NFP news, with sentiment analysis, to predict trends in FX market could be very valuable

Although the analysis of NFP news seems very interesting, It represents more time to build a model that makes sense. For this reason, I will focus my analysis using the Intraday data. In particular, I will use the intraday information of USD/EUR. This dataset can be found in:

01_code/dataset/EURUSD

6.2 Relevant analytics from this data source

Financial markets witness high levels of activity at certain times, but remain calm at others. This makes the flow of physical time discontinuous. Therefore using physical time scales for studying financial time series, runs the risk of missing important activities. An alternative approach is the use of an event-based time that captures periodic activities in the market.

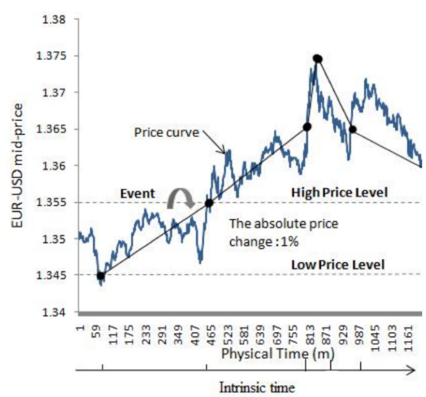

Figure 9: : The graph shows a 24-hour EUR/USD mid-price sample and sequential events (solid lines) defined by a threshold $\lambda = 1\%$. Intrinsic time triggers only at periodic events whereas physical time ticks equally across different patterns of different magnitudes in the price curve

Directional-Change events is an alternative way to model financial time series. A directional change event is characterized by a fixed threshold λ and is defined as the absolute price change between two local extremal values exceeding a given threshold λ. Figure 9 shows EUR/USD price activities on the 7, January 2009 sample onto a reduced set of four sequential events defined by a threshold $\lambda = 1$. [3]

[3]Edward Tsang, A Directional-Change Events Approach for Studying Financial Time Se-

6.3 Analytics

Using the dataset EUR/USD, the number of directional changes detected were the following:

threshold	Number of Directional Changes
0.05%	320
0.10%	33
0.15%	15
0.20%	3

Figure 10: Number of directional change events in two days of intraday data

Given a threshold, a higher number of directional change events represents a more active asset and more opportunities to trade.

The python file to get the number of directional changes can be found:

01_code/06_direcional_change.py

6.4 Co-integrated with currency pairs

The analysis is done using a dataset for each currency pair.

ries