Analysis of a 56 stock portfolio//summary data//January 3rd 2005 to December 31st 2015.

I. Section I-A

Kingsley Oteng-Amoako June 1st, 2016.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
nobs	2770	2770	2770	2770	2770	2770	2770	2770	2770	2770
NAs	1794	1763	4	426	4	4	4	4	4	838
Minimum	-0.05	-0.15	-0.17	-0.13	-0.10	-0.09	-0.18	-0.08	-0.28	-0.13
Maximum	0.09	0.25	0.17	0.25	0.10	0.11	0.33	0.09	0.41	0.15
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Variance	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Stdev	0.01	0.02	0.02	0.02	0.01	0.01	0.02	0.01	0.02	0.02
Skewness	1.14	3.06	-0.06	0.98	0.19	0.16	1.98	-0.20	0.94	0.41
Kurtosis	9.36	66.25	11.13	18.76	5.11	7.13	53.50	6.34	43.98	12.19
1. Quartile	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
3. Quartile	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
nobs	2770	2770	2770	2770	2770	2770	2770	2770	2770	2770
NAs	4	889	4	4	4	651	4	342	4	4
Minimum	-0.13	-0.13	-0.11	-0.48	-0.12	-0.08	-0.11	-0.14	-0.09	-0.21
Maximum	0.11	0.13	0.09	0.72	0.19	0.20	0.08	0.22	0.09	0.20
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Variance	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Stdev	0.02	0.02	0.01	0.04	0.02	0.01	0.01	0.02	0.01	0.02
Skewness	-0.02	0.01	-0.33	2.50	0.30	2.24	-0.40	1.30	0.27	-0.41
Kurtosis	7.50	6.31	10.12	62.56	6.16	40.48	7.43	23.38	6.28	13.96
1. Quartile	-0.01	-0.01	-0.00	-0.01	-0.01	-0.00	-0.01	-0.01	-0.01	-0.01
3. Quartile	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01

Table 1: reports the summary results of returns for the stocks X10 through to X20 for the period January 2005 to December 2015, (a) number of days of observations, (b) number of missing data points, (c) minimum, (d) maximum, (e) mean, (f) median, (g) variance, (h) standard deviation, (i) skewness, (j) kurtosis, (k) mean of 1st quartile and (l) mean of 3rd quartile.

	X21	X22	X23	X24	X25	X26	X27	X28	X29	X30
nobs	2770	2770	2770	2770	2770	2770	2770	2770	2770	2770
NAs	4	4	4	2645	4	2760	2069	4	4	4
Minimum	-0.08	-0.36	-0.36	-0.05	-0.09	-0.02	-0.06	-0.10	-0.10	-0.15
Maximum	0.08	0.08	0.08	0.05	0.14	0.02	0.36	0.11	0.11	0.33
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sum	0.94	0.86	0.86	0.04	1.23	0.01	0.86	1.69	1.69	1.57
Stdev	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.02	0.02
Skewness	-0.35	-5.57	-5.57	0.21	0.77	0.18	11.94	0.25	0.25	1.57
Kurtosis	8.54	144.80	144.80	0.66	14.76	-1.47	239.39	4.87	4.87	29.46
1. Quartile	-0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
3. Quartile	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

	X31	X32	X33	X34	X35	X36	X37	X38	X39	X40
nobs	2770	2770	2770	2770	2770	2770	2770	2770	2770	2770
NAs	1009	4	1037	4	698	4	4	1476	1476	4
Minimum	-0.10	-0.36	-0.10	-0.12	-0.26	-0.26	-0.12	-0.19	-0.10	-0.12
Maximum	0.12	0.08	0.12	0.08	0.26	0.30	0.16	0.26	0.22	0.09
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Variance	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Stdev	0.02	0.01	0.02	0.01	0.03	0.03	0.01	0.02	0.02	0.01
Skewness	0.58	-5.57	0.22	-0.50	0.26	0.64	0.25	2.36	1.44	-0.16
Kurtosis	9.04	144.80	5.99	9.37	8.70	12.68	18.55	44.06	22.80	11.17
1. Quartile	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
3. Quartile	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01

Table 2: reports the summary results of returns for stocks X21 through to X40 for the period January 2005 to December 2015, (a) number of days of observations, (b) number of missing data points, (c) minimum, (d) maximum, (e) mean, (f) median, (g) variance, (h) standard deviation, (i) skewness, (j) kurtosis, (k) mean of 1st quartile and (l) mean of 3rd quartile.

	X41	X42	X43	X44	X45	X46	X47	X48	X49	X50
nobs	2770	2770	2770	2770	2770	2770	2770	2770	2770	2770
NAs	809	1782	741	4	4	574	4	886	4	4
Minimum	-0.09	-0.16	-0.21	-0.10	-0.25	-0.19	-0.10	-0.14	-0.17	-0.18
Maximum	0.13	0.18	0.26	0.08	0.26	0.34	0.15	0.13	0.37	0.14
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
Variance	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Stdev	0.01	0.02	0.02	0.01	0.02	0.03	0.01	0.02	0.02	0.02
Skewness	0.27	0.47	0.36	-0.15	0.41	1.48	0.53	-0.09	2.37	0.04
Kurtosis	9.43	11.49	38.95	5.80	20.01	21.41	14.62	10.53	51.60	13.80
1. Quartile	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
3. Quartile	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

	X51	X52	X53	X54	X55
nobs	2770	2770	2770	2770	2770
NAs	118	4	4	4	1972
Minimum	-0.18	-0.14	-0.24	-0.16	-0.07
Maximum	0.14	0.11	0.17	0.13	0.11
Mean	0.00	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00	0.00
Variance	0.00	0.00	0.00	0.00	0.00
Stdev	0.02	0.02	0.02	0.02	0.02
Skewness	-0.11	0.11	-0.52	-0.31	0.67
Kurtosis	11.49	7.77	11.93	8.06	3.29
1. Quartile	-0.01	-0.01	-0.01	-0.01	-0.01
3. Quartile	0.01	0.01	0.01	0.01	0.01

Table 3: reports the summary results of returns for the stocks X41 through to X56 for the period January 2005 to December 2015, (a) number of days of observations, (b) number of missing data points, (c) minimum, (d) maximum, (e) mean, (f) median, (g) variance, (h) standard deviation, (i) skewness, (j) kurtosis, (k) mean of 1st quartile and (l) mean of 3rd quartile.

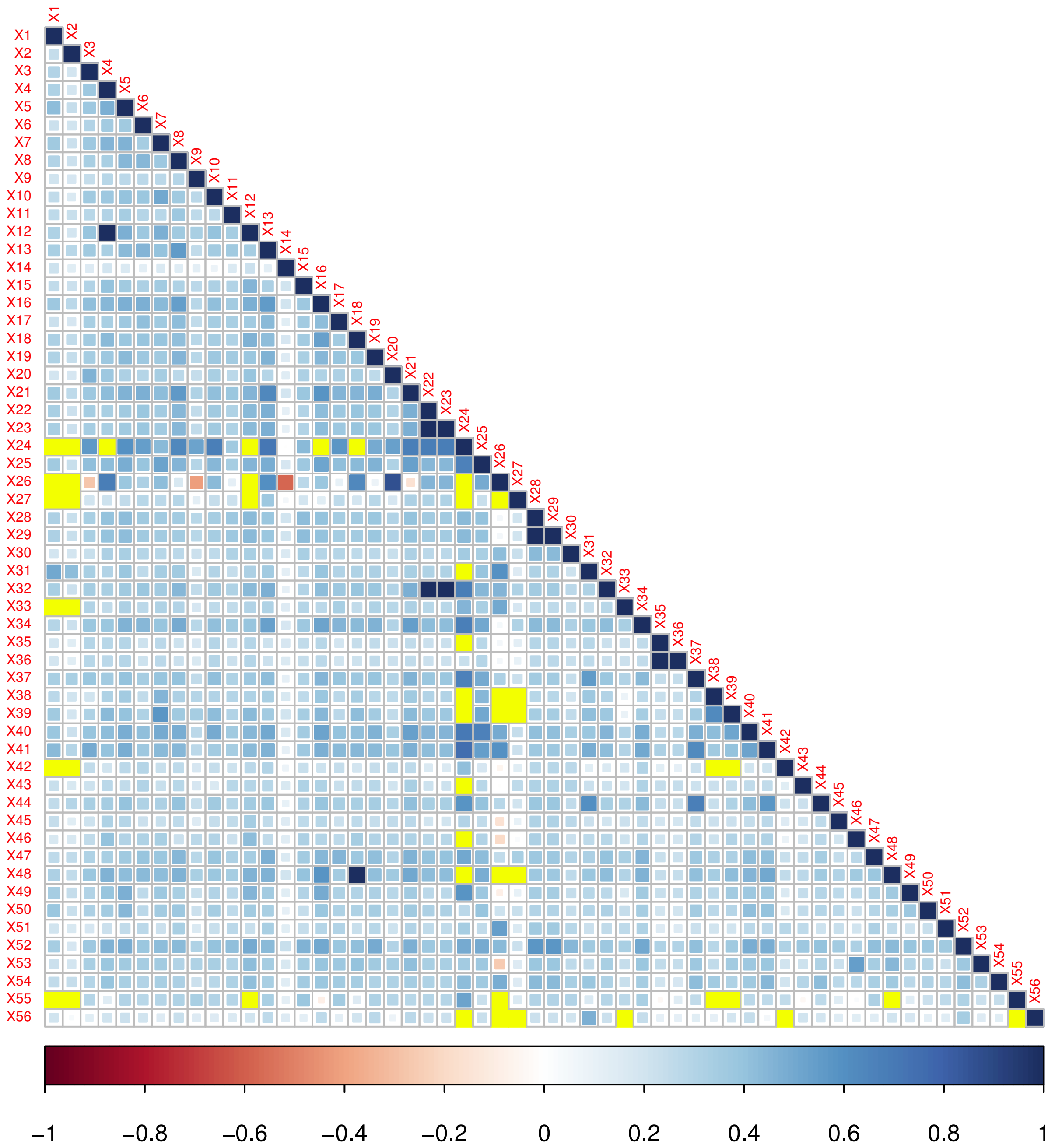

Fig. 1: reports the **cross-correlation of returns** for all stocks (X1 to X56) for the period January, 2005 through to December, 2015.

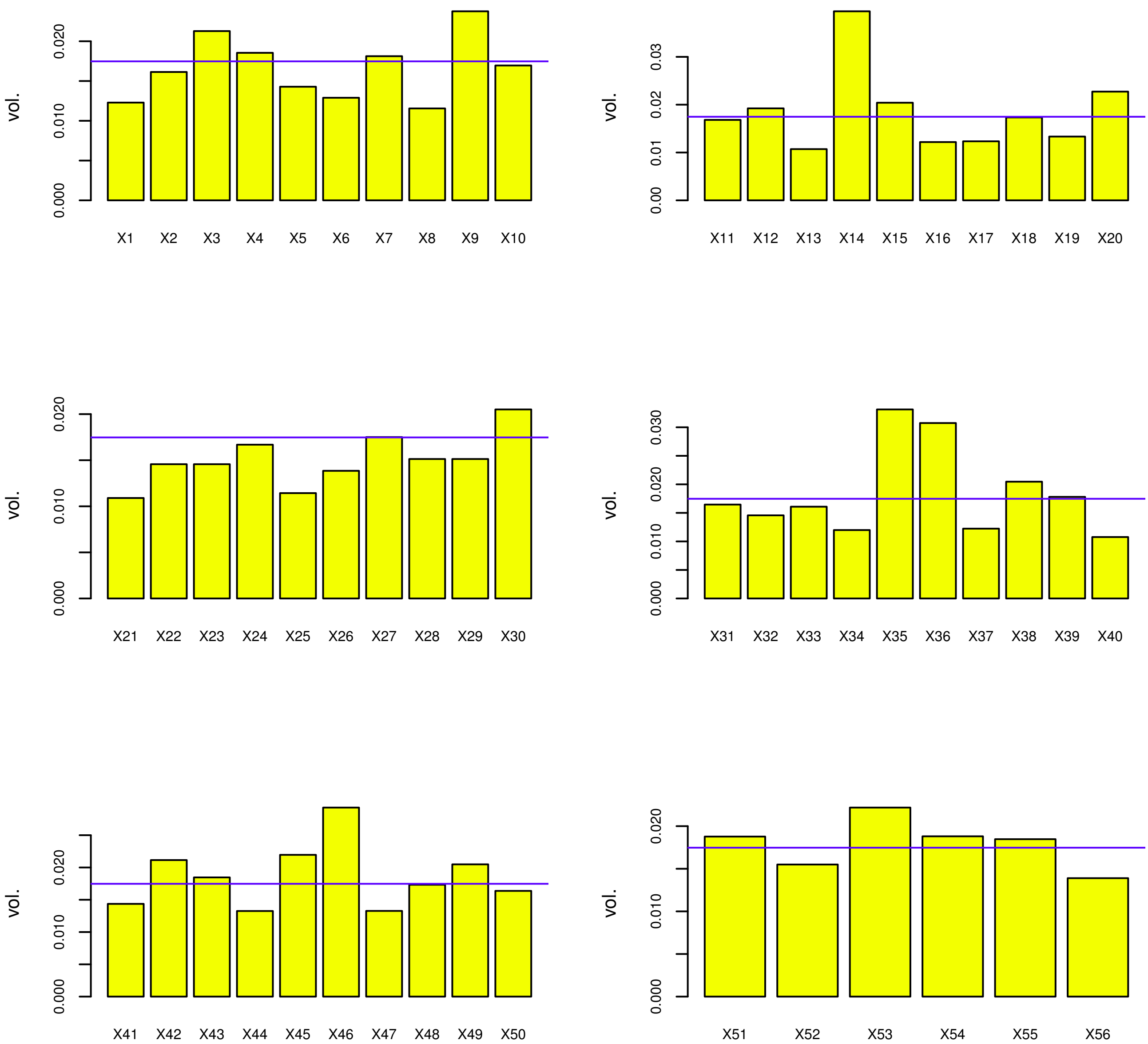

Fig. 2: reports the **volatility of returns** for all stocks (X1 to X56), for the period January 2005 through to December 2015.

II. Section II-A

A hierarchical regression is initially considered to test for a causal relationship between the forecast returns as the dependent variable and the various independent variables, i.e. the previous stock returns and other contributing factors. Consider a linear model where we exame the relationship between a variable X_k (termed the dependent variable) and a variable y_l (termed the independent variable) as,

$$y_l = X_k\beta + \varepsilon \tag{1}$$

Assume that the linear regression model has errors that have an expectation zero, uncorrelated error terms and have a constant variance. Subsequently the best linear unbiased estimator for

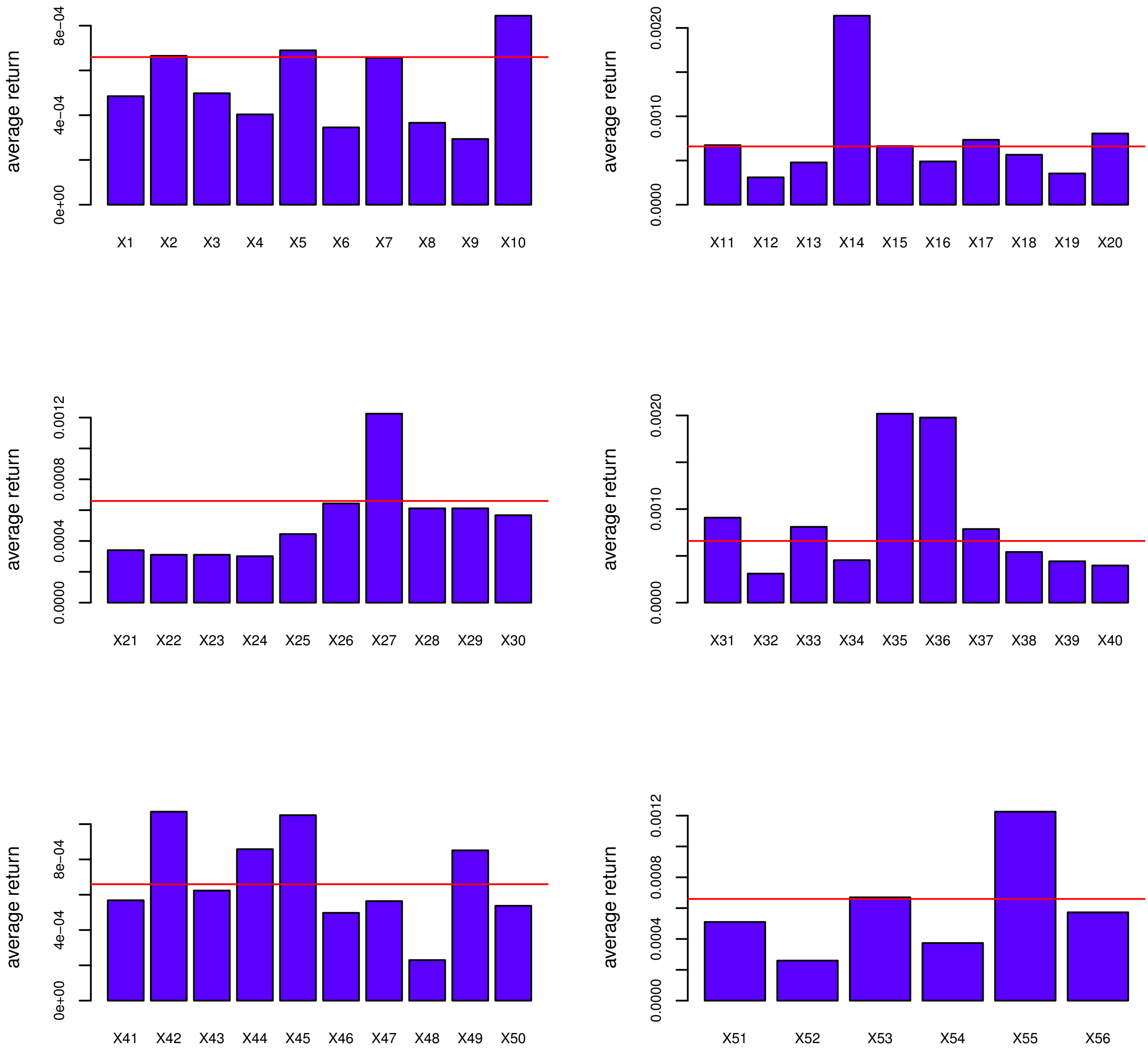

Fig. 3: reports the **mean return** for all stocks (X1 to X56) for the period January 2005 through to December 2015.

the coefficients of the model correspond to the Ordinary Least Squares (OLS) algorithm,

$$\beta^* = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{2}$$

Finally, we consider the empirical risk of the fit (i.e. error) of the multidimensional linear model as,

$$\begin{aligned} R(\beta) &= \frac{1}{2N}\left|\left|\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} 1 & x_1(1) & \cdots & x_1(D) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_N(1) & \cdots & x_N(D) \end{bmatrix}\right|\right|^2 \\ &= \frac{1}{2N}||\mathbf{y} - \mathbf{X}\beta||^2 \end{aligned} \tag{3}$$

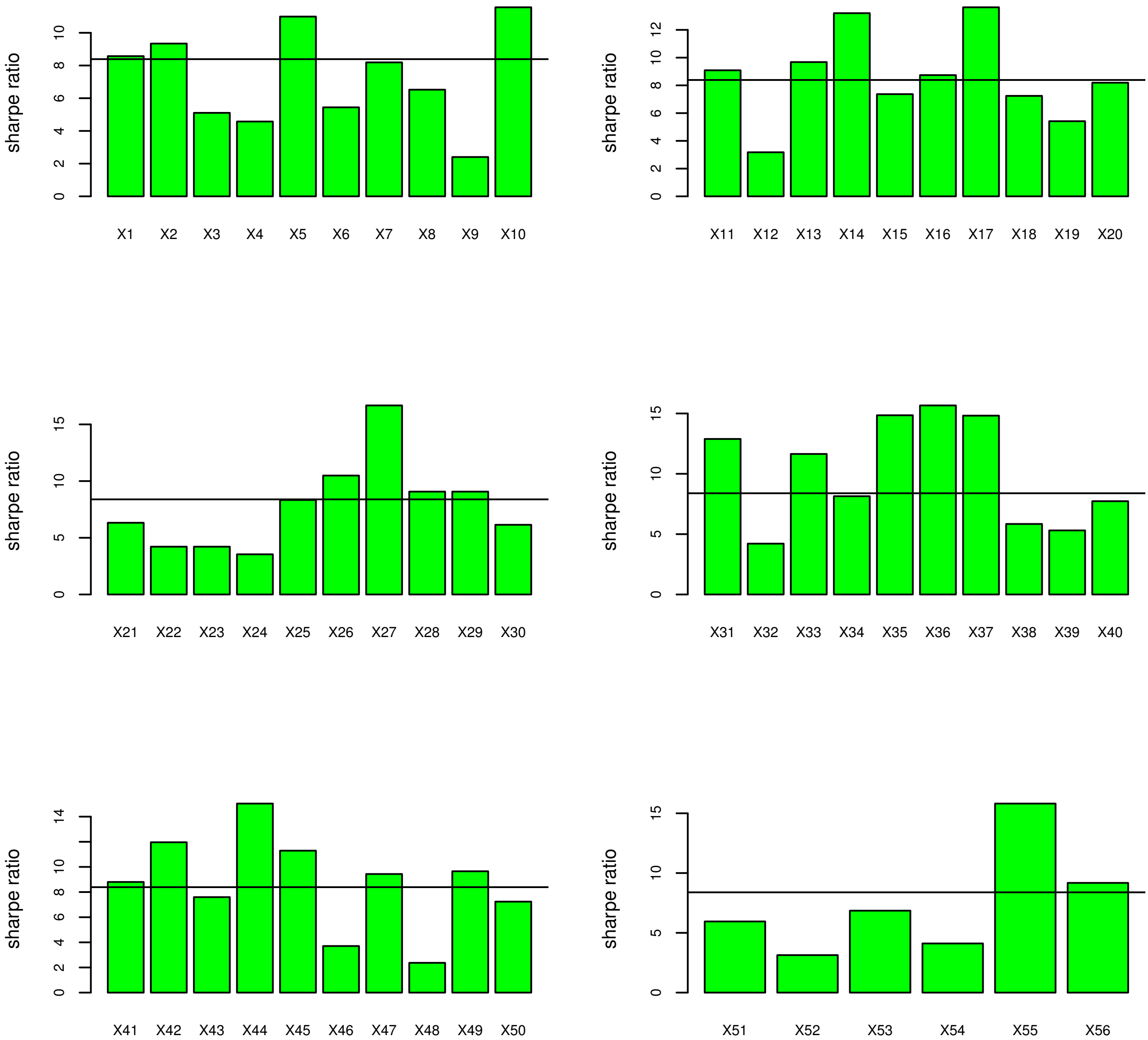

Fig. 4: reports the **Sharpe ratio** for all stocks (X1 to X56) for the period January 2005 through to December 2015. We consider the 2 YR bond yield 1.685 % (June 1st 2016) as the *risk-free* rate.

Forecasting using a single aggregate factor.

In forecasting returns we initially consider whether stock returns are dependent on aggregate returns of the basket (*market factor*), i.e. whether there is a causal relationship between current returns ($y = RET_{t-1}$) and previous three day lagged basket returns ($x_1 = MKT_{t-3}$). For example in figure 7 and 8, we consider the plot of stock returns and forecasted returns for stocks X28 and X14 respectively. For the linear model we obtain a standard error of 23.19% across all 56 stocks for the horizon January 2005 to December 2015.

Known limitations with this proposed approach are,

- Limited accuracy in the forecasts with an average error of 23.19% across all stocks when using a single market factor. This is however consistent with the market being 'efficient'.

- The model fails to adequately account for jump risk, i.e large negative or positive daily stock market returns. It is observed in figure 6 that although we obtain a model error 18.67% there is a large deviation between the long-term mean of stock returns and forecasted returns[1].

Forecasting using a single aggregate factor and prior stock return.

Prior returns of each stock are introduced into the model in addition to a market factor, i.e. we consider the causal relationship between current stocks returns ($y = RET_{t-1}$), previous three day lagged market returns ($x_1 = MKT_{t-3}$) and three day lagged stock returns ($x_2 = STCK_{t-3}$). For this linear model using two independent variables obtain a standard error of 23.36% across all 56 stocks for the horizon January 2005 to December 2015. This would suggest that use of lagged daily returns as a factor does not improve model quality.

Forecasting using a single aggregate factor and prior stock return and additional signals.

We also introduce two sets of signals into the hierarchical model, although the standard error fell there were no significant improvements to be made in the modeling process which could not be achieved using such techniques as independent components analysis or principal components analysis given historical data on the basket. I will however be happy to present additional modelling results, if requested.

III.

IV. Next steps:

In our models we have considered an Ordinary Least Squares estimator - this can be improved upon. Also, as suggested in the final analysis, there was some overlap in the factors considered in the model which can also be addressed through further modelling. We also suggest the use of a Kalman filtering technique or stochastic calculus to address the issue of jumps. Lastly in creating the market factor we considered all stocks in the basket however there appear to be outliers and thus further research would address this to create a market factor.

[1]For example, to address this issue a stochastic process with Poisson jumps may be applied to model a system with persistent jumps, i.e.; $dS/S = (\mu - \lambda\kappa)dt + \sigma ds + dq$, where dz is a Wiener process and dq is the Poisson jump process such that $dq = (\tilde{N} - 1)$ if a jump occurs and $dq = 0$ for no jump. In each time interval dt the stochastic process assumes there is a λdt probability that a jump will occur.

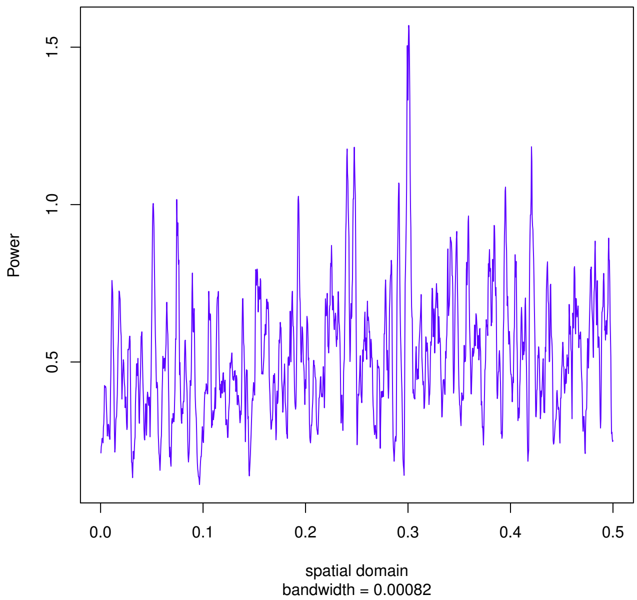

Fig. 5: Periodogram for stock X28 for the period January 2005 to December 2015; estimated model error of 18.67%.

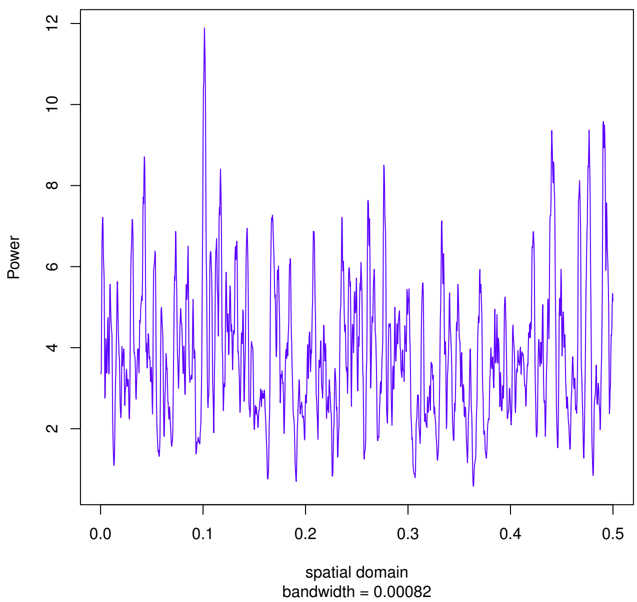

Fig. 6: Periodogram for stock X14 for the period January 2005 to December 2015; estimated model error of 36.40%.

Appendices

Discrete Fourier Transform (DFT): In order to evaluate a polynomial of degree-bound n at n different points of the form,

$$G_n = \sum_{j=0}^{n-1} g_j x^j$$

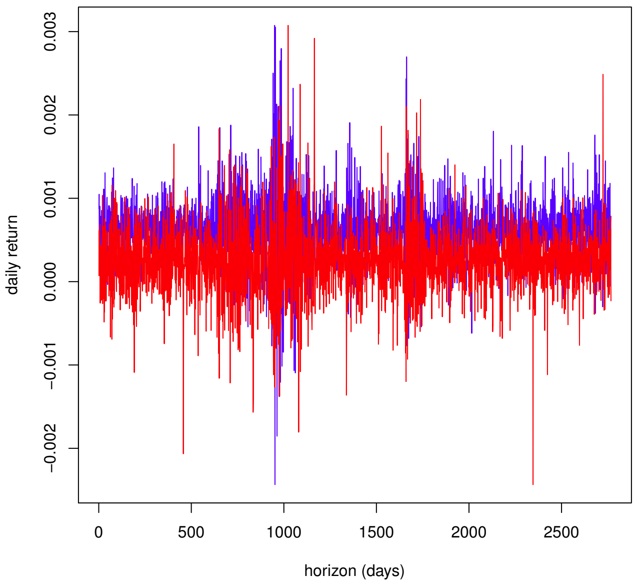

Fig. 7: Stock returns (blue) and forecasted returns (red) for stock X28 for the period January 2005 to December 2015.

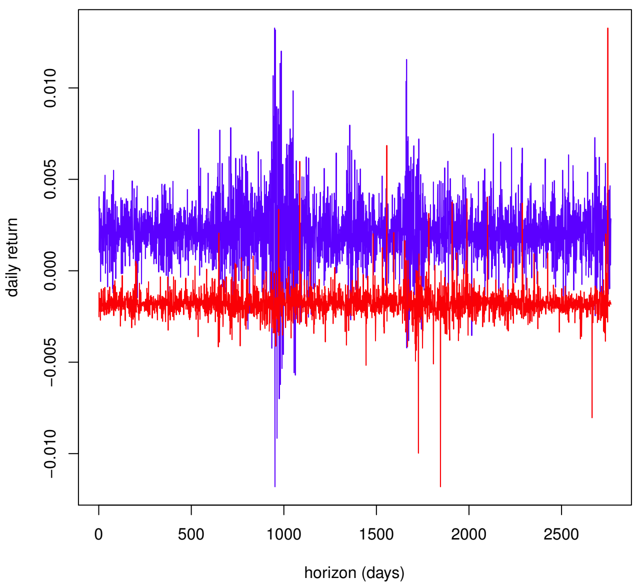

Fig. 8: Stock returns (blue) and forecasted returns (red) for stock X14 for the period January 2005 to December 2015.

We assume that n is of the power of tower; we may assume n a transform of n complex roots,

$$W \Rightarrow w_n^k = \{w_0, w_1, ..., w_{n-1}\} \tag{4}$$

where the nth root of unity in the spatial domain is given as,

$$w_n \equiv e^{-i\dfrac{2\pi}{n}} = \cos\left(\frac{2\pi}{n}\right) - i\sin\left(\frac{2\pi}{n}\right) \tag{5}$$

wherein the coefficient of the complext roots A is defined as,

$$A \Rightarrow a = (a_0, a_1, ...a_{n-1})$$

Thus the Discrete Fourier Transform (DFT) $y = DFT_n$ is defined in the spatial domain as,

$$\begin{aligned} y_k &= A.W \\ &= A(w_n^k) \\ &= \sum_{j=0}^{n-1} a_j w_n^{k_j} \\ &= \sum_{j=0}^{n-1} a_j \left(\cos\left(\frac{2\pi}{n}\right) - i \sin\left(\frac{2\pi}{n}\right)\right) \end{aligned} \tag{6}$$

The computation of the DFT given the coefficient A of degree-bound n and the coefficient W of degree-bound n is complexity $O(n^2)$ (n^2 complex multiplications and $n(n-1)$ complex additions).

Fast Fourier Transform (FFT): The Fast Fourier Transform utilizes the properties of the complex roots to reduce the computation time down from complexity $O(n^2)$ to $O(n \log n)$. Consider the coefficient A, where we represent the coefficient as two new polynomials $A'(x)$ and $A''(x)$ of degree-bound $m : m = n/2$,

$$A'(x) = a_0 + a_2 x + a_4 x + ... + a_{n-2} x^{m-1} = \frac{1}{M} \sum_{x=0}^{M-1} f(2x) w_{2M}^{(2x)u},$$

$$A"(x) = a_1 + a_3 x + a_5 x + ... + a_{n-2} x^{m-1} = \frac{1}{M} \sum_{x=0}^{M-1} f(2x+1) w_{2M}^{(2x+1)u},$$

Thus the coefficient A may be evaluated recursively in terms of the even-coefficients and odd-coefficients as,

$$\begin{aligned} A(x) &= A'(x^2) + xA''(x^2) \\ &= \frac{1}{2}\left(\frac{1}{M} \sum_{x=0}^{M-1} f(2x) w_{2M}^{(2x)u} + \frac{1}{M} \sum_{x=0}^{M-1} f(2x+1) w_{2M}^{(2x+1)u} \right) \end{aligned}$$

where $A'(x^2)$ and $A"(x^2)$ are each Discrete Fourier Transforms where we may evaluate each as m distinct values of the complex m th roots of unity as,

$$W = \left((w_{n/2}^0)^2, (w_{n/2}^1)^2, ..., (w_{n/2}^{n/2})^2 \right) \tag{7}$$