Quantitative Trading - Analysis of Momentum strategies in the UK

Introduction

The objective of this project was to analyse momentum as a factor for investment strategy in the UK over a historical period of 1992-2017. Momentum investing as a strategy aims to capitalise on existing trends in the market, which simply means that once a trend is established, whether it is an up or down trend it is more likely to continue in that direction. At a high level, this strategy is a bet on past returns predicting the cross section of future returns, typically implemented by buying past winners and selling past losers. Academic literature has shown the success of momentum strategies across multiple time periods and in numerous assets globally (Daniel & Moskowitz, 2015). However, these strong returns of momentum strategies are exposed to the large drawdowns during market turmoil, which for a risk adverse investor, can be problematic.

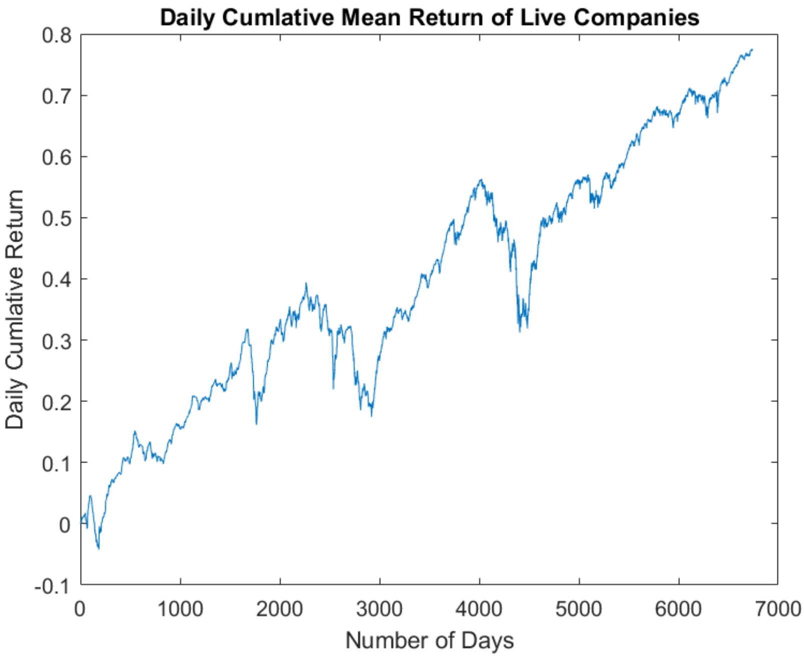

Figure 1: Evidence of sharp draw-downs in UK returns

Figure 1 above shows the sharp drawdowns that can take place in the market and this is when the momentum strategy does not perform well. From an investor point of view, to gain a large return from this strategy it must hold its investment in these sharp downturns during market crashes for example, which of course is not for the faint hearted.

Much literature exists in this area of investment research, particularly in various extensions of the momentum strategy. These extensions focus on using a momentum strategy with some form of risk management to try and reduce the exposure of the strategy to these drawdowns. This report follows closely the work of Barroso & Santa-Clara (2015), which aims at risk adjusting the momentum through volatility scaling. Barroso & Santa-Clara simply scaled the volatility to some predefined target level, shown as,

$$\gamma_t^* = \frac{\sigma_{\text{target}}}{\hat{\sigma}_t}\gamma_t, \tag{1}$$

where γ_t^* is the risk-adjusted momentum returns, γ_t is the raw momentum returns, σ_{target} is a constant volatility target for the strategy, and $\hat{\sigma}_t$ is the standard deviation of returns over a specified time period. They discovered that managing this risk eliminates crashes and nearly doubled the Sharpe ratio of their momentum strategy, which is obviously very attractive for investors as it will in turn improve the negative skewness and lower a high kurtosis. In Figure 2 below, we can see the periods high volatility within the UK data and Equation 1 above attempts to take these variations of volatility towards a constant target.

The analysis also replicates some of the work carried out by Ledoit & Wolf (2003), whom suggest using a covariance matrix obtained from the sample covariance matrix through a transformation called shrinkage. This tends to pull the most extreme coefficients towards more central values, thereby systematically reducing estimation error where it matters most. Statistically, the challenge is to know the optimal shrinkage intensity, and this report will use the same MATLAB code around calculating this optimal shrinkage factor, provided by Ledoit's publicly available research and code library.

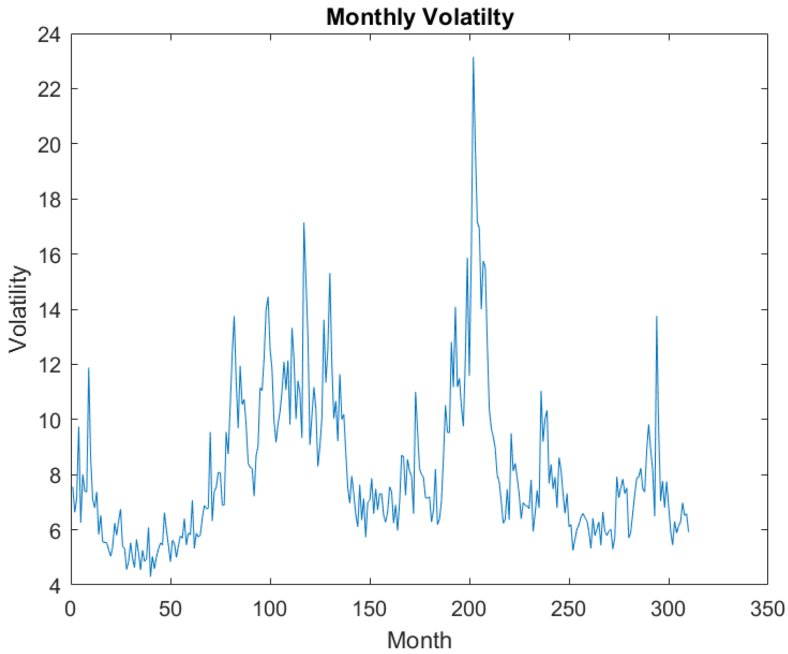

Figure 2: Monthly volatility of UK data.

The goal of this report is to replicate some of the main findings of the above mentioned authors, but on UK equities data.

Data Description

The data sample is constructed as follows:

- 1605 UK stocks daily returns in GBP from January 1992 to November 2017.
- These stock returns include a 'live' matrix, made up of a dummy variable to distinguish which companies are live on any given particular trading day.
- A momentum exposure matrix which was formed by taking the twelve month returns for each company, excluding the most recent month to take into account for the short term reversal effect.
- The market value of each company (millions) on any given particular trading day.

Of course, considering the sheer volume of data, the data sets are not clean and contain many missing values. In carrying out any analysis of the data it was important to clean the data through each regression. Although this can be computationally inefficient, its vital to ensure no performance bias or any statistical inference from results. In dealing with this challenge, it can be seen from the code that in order to clean the data, the 'validIndex' is computed. This basically means for each trading day only indexes that are valid are used, which entails the following:

- The company is live on that trading day.
- There is both a reported momentum exposure and market value for that company.

- There is a reported return value for that company on the following day, as in the regressions the aim is to regress on one day ahead returns. Although it can be argued that company market value is not needed for all the regressions, to ensure consistency throughout the analysis this check was kept in for all the analysis.

A simple procedure of obtaining the 'validIndex' is provided in Table 1 below. To interpret this simple example, pretend that each row is the data on the same day for live, mom, market value and the returns are one day ahead. Creating the code in this form allows firstly to ensure only clean data is used and secondly in the regressions you know that the vectors will be of the same size with the correct data.

Index	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Live	1	1	1	1	1	0	1	1	1	0
Mom	0.2	0.4	NaN	0.5	0.3	NaN	0.7	0.3	0.1	NaN
MV	300	200	100	500	700	NaN	300	150	400	200
Return	0.2	NaN	0.2	0.1	0.12	NaN	0.2	0.3	0.1	NaN
ValidIndex	1			4	5		7	8	9	

Table 1: Simple example of how the validIndex is computed for a given trading day.

Methodology

The focus in terms of methodology was conducting different combinations of Fama-Mac Beth regressions which allows the construction of factor portfolios. The first was an OLS regression which was carried out using the equation,

$$\gamma_{ols} = (X_t' X_t)^{-1} X_t' Y_{t+1}\,, \tag{2}$$

where γ is the coefficient or factor return, X_t is the momentum exposure and Y_{t+1} is the returns one period ahead. An intercept was also included within the regressions, simply by adding a vector of ones in MATLAB along with X_t. This regression implies that the equation is regressing, for each single trading day, todays stock returns on one-day lagged momentum exposures. Appendix A contains an outlay of what code to run in order to obtain each regression.

The second regression uses a GLS regression instead of OLS, with an equation,

$$\gamma_{gls} = (X_t' MV X_t)^{-1} X_t' MV Y_{t+1}\,, \tag{3}$$

where it is the same as the OLS except the MV is the market weights of each company for that day, where the weights on any day will sum to one. This matrix is a diagonal matrix, where the main diagonal contains the market weight for each valid company, and the off-diagonal elements are zero.

The third regression is of the same form as the GLS, but the MV is replaced with the covariance matrix, such that

$$\gamma_{shrink} = (X_t' \Sigma^{-1} X_t)^{-1} X_t' \Sigma^{-1} Y_{t+1}\,. \tag{4}$$

The covariance matrix is formulated from the work of Ledoit and Wolf. As briefly mentioned in the Introduction, Ledoit ad Wolf show that there is a lot of estimation error in just using a covariance matrix of past returns, therefore adopting their approach allows a reduction in this error. Consider the sample covariance matrix S and a highly structured estimator, denoted by F. Ledoit and Wolf found a compromise between the two by constructing a convex linear combination with,

$$\delta F + (1-\delta) S\,, \tag{5}$$

where δ is a number between 0 and 1. This technique is called shrinkage, since the sample covariance matrix is 'shrunk towards the structured estimator where δ is referred to as the shrinkage intensity constant. This approach is wonderfully described by Ledoit as, "Most people would be prefer the 'compromise' of one bottle of Bordeaux and one steak to either 'extreme' of two bottles of Bordeaux (and no steak) or two steaks (and no Bordeaux)". The structured estimator can be formed in many ways as seen in Disatnik and Benninga (2007), but this report uses the average correlation model which has been a common approach within academic literature and an approach also used by Ledoit and Wolf. To compute the average correlation constant ρ,average either side of the correlation matrix, then the structured estimator can be formed by,

$$F_{ii} = S_{ii} \quad and \quad F_{ij} = \rho\sqrt{S_{ii}S_{jj}}\,. \tag{6}$$

An important rule in obtaining the sample covariance matrix on each trading day, meant only stocks that had valid returns over the previous 500 days were included in each regression, an extra step in obtaining the 'validIndexe'. As these results would only start after a window of 500 days has elapsed, to ensure consistency and comparable results, all regressions have this feature. Although 500 days is substantial, given the size of the dataset (nearly 7000 days) and given the low volatility at the start of the data period, this does not have any performance bias on the main statistical parameters.

Once each regression was run they were risk adjusted using the method of volatility scaling by Barossa & Santa-Clara, outlined briefly in the Introduction. Barossa & Santa-Clara showed that using the realized volatility of daily returns allows for the estimation of momentum risk. The realized volatility was calculated using,

$$\hat{\sigma_t} = \sqrt{\sum_{j=0}^{250} \gamma_t^2/250.} \tag{7}$$

To get this figure to an annual figure, then multiply by $\sqrt{250}$. Although not significant as such, the target volatility for the report was decided upon each regressions standard deviation to allow a more realistic target volatility.

Results

To interrupt the results, the gammaStats_Grp7 function presents a table of results for comparison of the main statistical parameters of each regression against its respective risk adjusted factor returns.

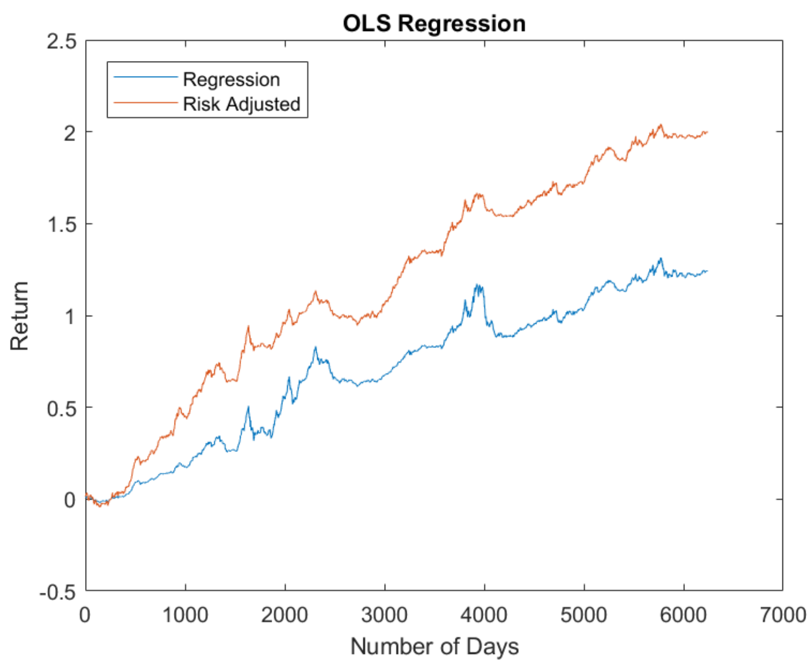

Figure 3: Comparison of OLS and risk adjusted OLS factor returns

GammaType	t-stat	Mean(%)	StdDev(%)	Skewness	Kurtosis	MD(%)	Sharpe
OLS	5.17	4.9	4.8	-0.776	10.75	29.2	1.71
RiskAdjOLS	8.48	8	4.7	-0.272	6.46	19.1	1.90

Table 2: Statistical parameters (annualised) from the OLS and OLS risk adjusted factor returns

It can be seen how the risk adjusted factor returns have improved the statistical properties of the factor returns, whereby improving the higher order moments of skewness and kurtosis that investors desire. As the initial regression has a standard deviation of nearly 5%, a target volatility of 4.5% was selected. This number can be easily changed in the matlab code by simply adjusting the target input into the volScaling_Grp7 function.

Moving on to the second regression for GLS yields,

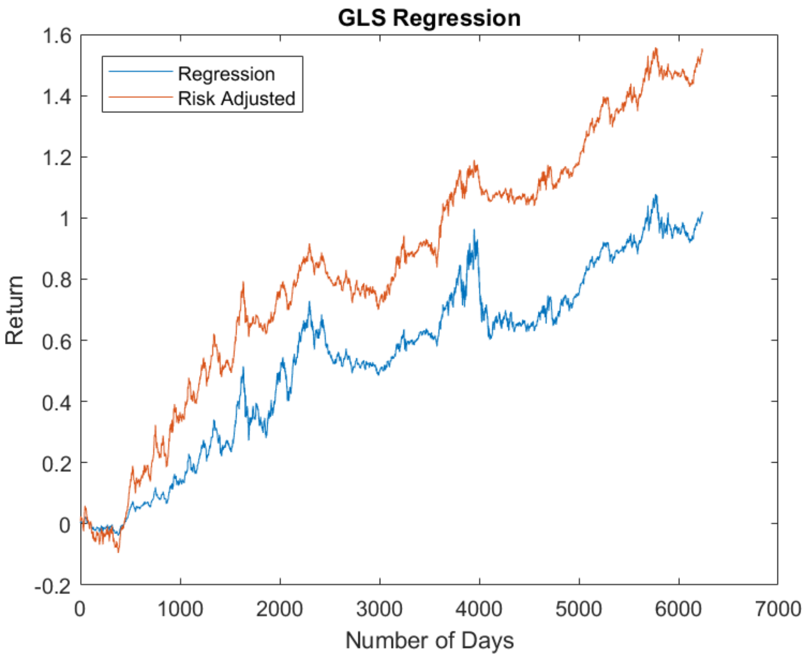

Figure 4: Comparison of GLS and risk adjusted GLS factor returns

GammaType	t-stat	Mean(%)	StdDev(%)	Skewness	Kurtosis	MD(%)	Sharpe
GLS	2.23	4.08	9.18	-0.362	10.276	36.15	1.72
RiskAdjGLS	3.60	6.20	8.42	-0.139	5.48	21.69	1.90

Table 3: GLS Statistical Results

It is evident how the standard deviation has increased within this regression. This may be due to the fact in the OLS each company had an equal weighting, whereas in the GLS, the regression is using market weights. Therefore the company with greater weights have more momentum exposure to that of the OLS. it can be seen from the tabular of results and particularly from a visual inspection of the graph how the drawdowns decrease within the risk adjusted strategy.

Lastly we can see the results of the third regression, whereby the covariance matrix is being updated daily.

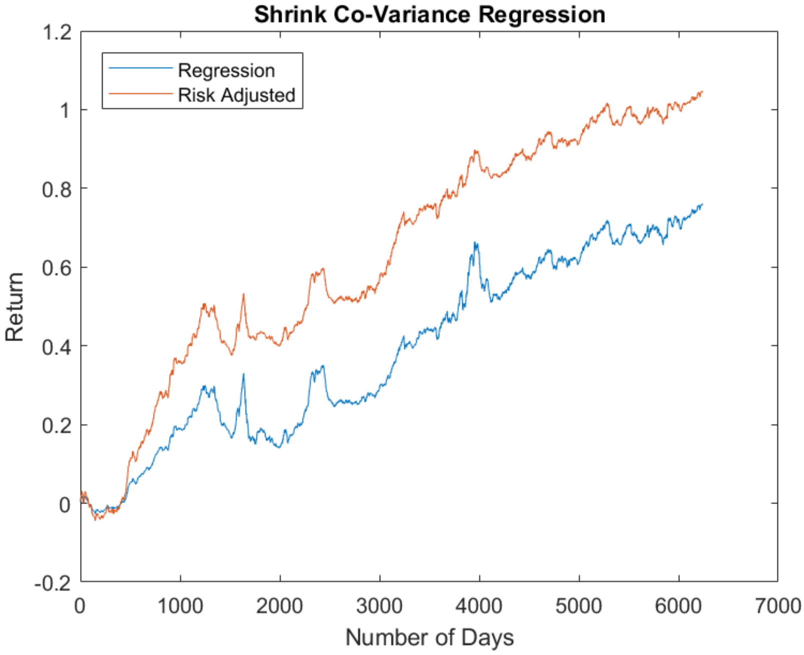

Figure 5: Comparison of Shrink V risk adjusted Shrink factor returns

GammaType	t-stat	Mean(%)	StdDev(%)	Skewness	Kurtosis	MD(%)	Sharpe
ShrinkReg	4.59	3.04	3.30	-0.323	9.1	19.03	1.55
RiskAdjShrink	6.78	4.19	3.0	-0.113	5.68	13.43	1.96

Table 4: Shrink regression results

It can be seen how the standard deviation has decreased compared to that of the GLS, which is somewhat obvious with the incorporation of the covariance matrix. This covariance matrix means factor portfolios have exposure of one to a given factor and zero exposure to all other factors along with the minimum risk possible which is an attractive combination. As in both previous regressions, all the statistical parameters show improvement.

Figure 6 below shows how the estimate of the optimal shrinkage intensity evolves through each day in the data sample. It is always between zero and one, which is expected. There is quite a wide range of variation within the shrinkage number, showing how the estimation error in the sample covariance matrix varies overtime. A high optimal shrinkage intensity concludes there is more estimation error in the sample covariance matrix and visa-versa. This variation in shrinkage intensity is quite different to that of Zivot's findings in the US, where it was shown how the intensity was near constant throughout the 23 year sample, with the intensity ranging from 0.7-0.8, albeit using the single index model compared to the average correlation model in this analysis (Zivot, 2015).

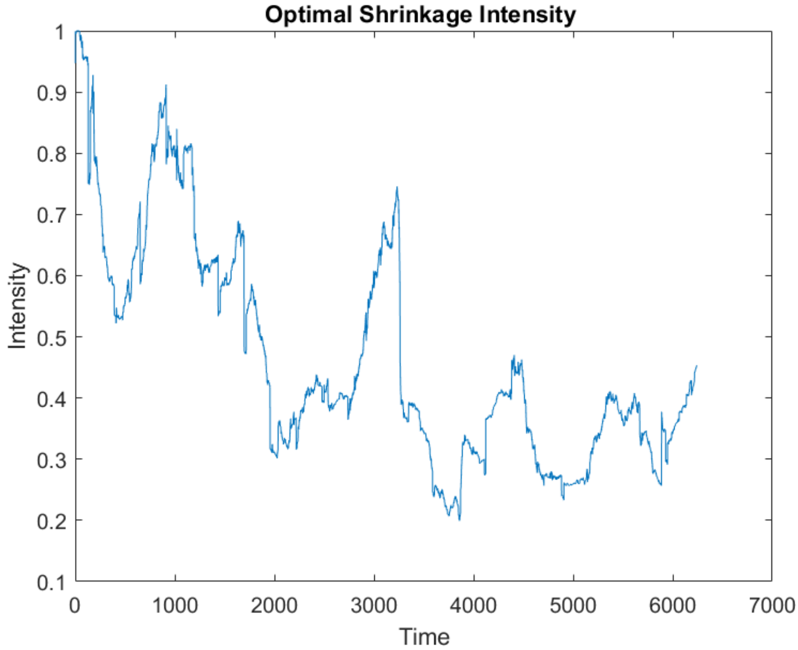

Figure 6: Optimal daily shrinkage intensity (δ) throughout life of data.

One of the most important gains of the risk management strategy is the improvement in higher order moments. Figure 7 below is the distribution of the GLS regression of both plain and risk adjusted momentum, but these results are matched in all the regressions. It can be seen how the plain momentum has a more leptokurtic distribution. Once risk adjusted, the results are evident on the improvement of kurtosis for example.

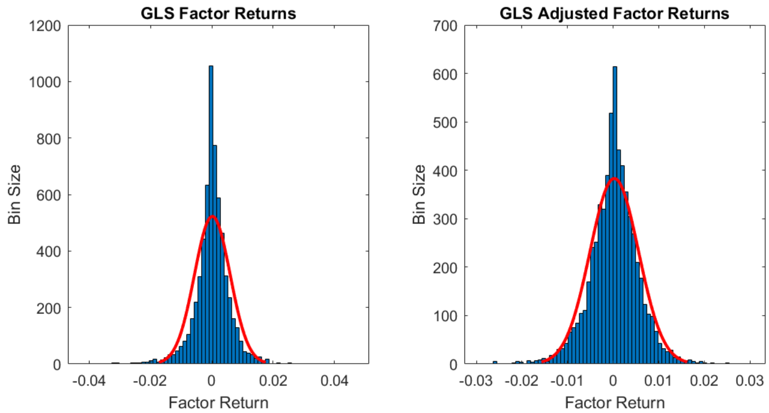

Figure 7: Example of how the risk adjustment method lowers the kurtosis and skew.

Conclusion

As Barrosa & Santa-Clara also reported, the results show that the unconditional momentum has a distribution that is far from normal, with large crash risk potential. Throughout all the regressions, it can be seen due to the high predictability of momentum, that it is possible to manage this risk. The results show improvements in all statistical parameters reported and agree with international evidence. Improvements of this project may include running various different shrinkage estimators that Benninga and Disatnik applied. A very important issue that was not considered in this analysis were transaction costs, which provides a very wide debate on how it affects risk management momentum strategies.

References

Barroso, P. and Santa-Clara. (2014). Momentum has its moments. *Forthcoming in the Journal of Financial Economics.*

Ledoit, O. and Wolf, M. (2003). Honey, I Shrunk the Sample Covariance Matrix. *UPF Economics and Business Working Paper.*

Benninga, S. and Disatnik, D.J. (2007) Shrinking the Covariance Matrix Simpler is better. *Journal of Portfolio Management*, 33(4), pg55-63.

Zivot, E. (2015). Modeling Financial Time Series with R. 2nd ed. Washington, chapter 6.

APPENDIX A

This is a simple flow outlay diagram of the MATLAB code to help the readers understanding. The analysis revolves around three scripts, one for each regression and functions running off them. The longest scripts to run is the shrink regression as it updates the covariance matrix daily (100-200 seconds depending on computer size).

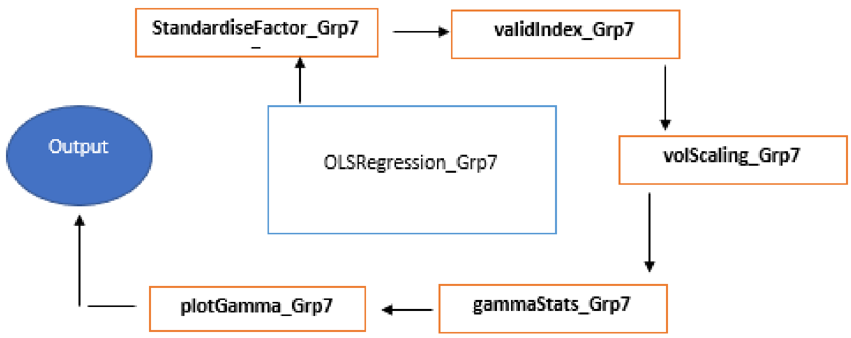

Figure 8: OLS regression code outlay and functions

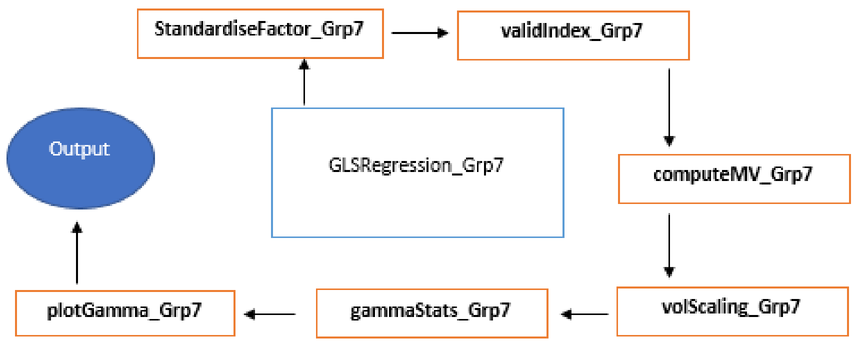

Figure 9: GLS regression code outlay and functions

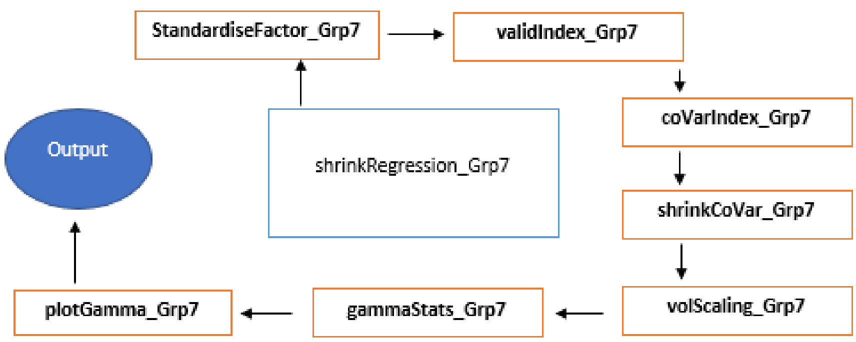

Figure 10: Shrink regression code outlay and function