Available online at http://scik.org
Math. Finance Lett. 2022, 2022:1
https://doi.org/10.28919/mfl/7643
ISSN: 2051-2929

DOES FORECASTING IMPROVE PORTFOLIO MANAGEMENT RETURNS? EMPIRICAL RESULTS FROM STRATEGY BACKTESTING

CHUNYEUNG KWOK*

Department Westminster Business School, University of Westminster, London NW1 5LS, UK

Abstract: This paper aims to determine whether forecasts from two popular macroeconometric models are useful to improve portfolio returns. The paper begins by estimating two large macro models namely Global Vector Autoregressive (GVAR) and Factor-Augmented Vector Autoregressive (FAVAR). The forecasts from these models are then used in a backtester, simulating a trading rule. In the first empirical test with a simple single position test, perfect forecast performed best but the highest return came from a strategy that uses GVAR forecast although it has a lower Sharpe ratio. The result from the second backtest with multiple positions is more in line with expectation as a strategy using the perfect forecast outperformed GVAR in all scenarios. The evidence from this paper shows how investment returns are driven by forecast accuracy but also heavily on portfolio management criteria.

Keywords: forecasting; modelling; portfolio management; GVAR.

2010 AMS Subject Classification: 62P05, 91G70, 91B64.

1. INTRODUCTION

Macroeconometric models have become the fundamental tool in the academia and central bank community. It is now almost impossible to have a rigorous understanding of the global economies without such tools. For example, there is literature devoted to comparing macroeconometric

---

*Corresponding author
E-mail address: w1654176@my.westminster.ac.uk
Received August 03, 2022

models and their applications, see Klein et al [13], Granger and Jeon [7], Welfe [22], Smith et al [21], Kwok [14] and Kwok [15].

This short paper attempts to answer the question of whether forecast improves investing and in which way it improves performance. The first part will be explaining how the forecasts from the macroeconometric model can be integrated into portfolio management practices, in light of the risk and reward preferences of the investor, risk management and related topics such as portfolio diversification and the end, and how we to put these together.

In the end, there are two empirical backtests covering a simple case of a single position long - short trading strategy and another on backtesting a strategy that includes multiple positions that incorporates portfolio management tools. The main contribution of this paper is in assessing the performance of investment returns with GVAR forecasts while backtesting two trading strategies. A further contribution is made in using portfolio criteria instead of simply assessing the forecast accuracy.

Several studies in the literature focus on picking the best models for forecasting prices in the stock markets but no assessment is made with regard to portfolio management which ignored the investor's preference for risk and reward. These portfolio criteria are particularly important in practice as they strongly affect returns, more so than the forecast accuracy. Therefore this is the biggest contribution from this paper as it not only assess GVAR forecasts but also incorporated portfolio management tools to derive the investment performance.

The first hypothesis of this paper: is whether a long short moving average cross-over strategy would have been profitable for trading the oil market. The second hypothesis: is whether the forecasts created from FAVAR and GVAR are contributory to the profits.

Two empirical tests were made to assess the performance. The first is a simple buy/sell single position trade. The second is a more realistic backtest which included portfolio management tools such as initial capital, volatility targeting and multiple positions. In the end, the performance is assessed by the return from the strategies and also the Sharpe ratio. The results from the empirical tests are mixed.

The bullet points below summarise the steps in this paper:

1 - Review backtesting methods and trading rules.

2- Produced forecasts from GVAR and FAVAR. Also obtained the actual data that a perfect forecast would produce.

3 - Formulate trading strategy into trading rules

4 - Conduct backtest 1 with forecasts from GVAR, FAVAR and actual data. The first backtest is for a single position only.

5 - Conduct backtest 2 with forecasts from GVAR, FAVAR and actual data. The second backtest is for multiple positions.

6. Evaluate the performance from both single and multiple positions backtests to determine whether forecasting helps investment.

2. ESTIMATING THE GVAR AND FAVAR MODELS AND FORECASTS

Estimating the GVAR model

In the case of GVAR, when modelling the world economy, each country is represented by its equation with a VARX* model which links the domestic country with the foreign countries and also a set of global variables such as oil and metal prices. In the individual model, the domestic model is linked to the foreign economies with their respective trade weights. Economically this is an intuitive approach as clearly, say policy shocks from India will have a lot higher impact on its neighbour such as Sri Lanka (which has many trades directly with India) than on Paraguay for example. Given the general nature of interdependencies in the world economy, see Pesaran et al [19], it is supposed that all country-specific variables and common global observed factors such as oil and commodity prices should be treated as endogenously (as part of the system i.e. a closed world economy). As the parameters to be estimated in the GVAR model are now restricted by the trade weights therefore this allows for the computation. Therefore in this sense, it is similar to the FAVAR approach which is by extracting ‘common factors ’ from relative trade weights rather than a statistical method.

The GVAR's objective of solving the curse of dimensionality is to impose a set of restrictions on the VAR model so that the model can be estimated practically while being consistent. The main restriction of the GVAR approach is by imposing the assumption of weak exogeneity of foreign country-specific and global variables. In other words, it assumes that the individual economy is relatively small in terms of the world economy except for the exception of the US, Dées et al [4]. The weak exogeneity is then tested empirically to see whether this assumption holds. Specifically, an individual country (or economy) is represented by a VARX* model (or in its error-correction form VECMX*) which links the domestic economy (defined by a range of domestic macroeconomic variables) to foreign economies (defined by corresponding foreign variables) which are treated as weakly exogenous. The domestic and foreign economies are then linked via weights matrices that match the international linkages in trade. The second step then stacks all individual country models together in a theoretically consistent manner that can generate forecasts for all world economic variables simultaneously.

The rest of the associated parameters are similar to those in a normal VAR, which are to be estimated to give context to economic interpretations of the model. It should be noted that x_{it}^{*} as a vector that captures the foreign-specific macroeconomic variables that are related to domestic ones are constructed via a weight matrix. The scheme of the weight matrix can be designed to reflect the trade and/or financial linkages. For example, the weight of Britain (domestic) is expected to have a large trade with the EU countries such as Germany (foreign), therefore it will have a larger weight than say, Malaysia.

As mentioned above, GVAR is a two-step process. The first was to estimate the VARX* model country by country and the second is to stack all VARX* models together and to be solved as a whole.

Country-specific VARX* models

The first step of the GVAR approach is the formulation of the individual VARX* (vector autoregressive with exogeneity) model for every country. In this section, we present the general methodology for advanced in Chudik and Pesaran [3] to model individual countries in the GVAR model applied to the model in this study. The approach assumes that there are N+1 countries in the global economy, indexed by i = 0, 1,. . . , N and the aim is to relate a set of country-specific variables e.g. GDP, inflation, interest rates etc. that are of interest to the study. The vector of interest is denoted as x_it collects the macroeconomic variables specific to the individual countries of interest indexed by I and over time, indexed by t = 0; 1, …,   T. Following the notation and definitions given in di Mauro and Pesaran [6] p.14-17, the general individual country model VARX* (2, 2) is represented as

$$x_{it} = a_{i0} + a_{i1}t + \varphi_{i1}x_{i,t-1} + \varphi_{i2}x_{i,t-2} + \Lambda_{i1}x_{it}^{*} + \Lambda_{i1}x_{it-1}^{*} + \Lambda_{i2}x_{it-2}^{*} + u_{it}$$

Where the definitions remain the same as defined before, we now introduce a few terms to solve the model as a whole. To form the GVAR model, we first introduce a new term zit define it as:

$$z_{it} = (x_{it}^{'}, x_{it}^{*'})'$$

Therefore we have:

$$A_{i0}W_i x_t = a_{i0} + a_{i1}t + A_{i1}W_i x_{t-1} + A_{i2}W_i x_{t-2} + u_{it}$$

Also recall that for i = 0, 1,. . . , N, which implies the equation above is individual country-specific and requires stacking to solve for x t which links all individual models together. We now introduce a few more terms to tidy up the model:

$$G_0 = \begin{pmatrix} A_{00}W_0 \\ A_{10}W_1 \\ \vdots \\ A_{N0}W_N \end{pmatrix}, \qquad G_1 = \begin{pmatrix} A_{01}W_0 \\ A_{11}W_1 \\ \vdots \\ A_{N1}W_N \end{pmatrix}, \qquad G_2 = \begin{pmatrix} A_{02}W_0 \\ A_{12}W_1 \\ \vdots \\ A_{N2}W_N \end{pmatrix},$$

$$a_0 = \begin{pmatrix} a_{00} \\ a_{10} \\ \vdots \\ a_{N0} \end{pmatrix}, \qquad a_1 = \begin{pmatrix} a_{01} \\ a_{11} \\ \vdots \\ a_{N1} \end{pmatrix}, \qquad u_1 = \begin{pmatrix} u_{0t} \\ u_{1t} \\ \vdots \\ u_{Nt} \end{pmatrix}$$

thus

$$G_0 x_t = a_0 + a_1 t + G_1 x_{t-1} + G_2 x_{t-2} + u_t$$

As the term G_0 is a known non-singular matrix (invertible matrix). G_0 is called non–singular if there exists an n × n matrix G_0^{-1} such that $G_0 G_0^{-1} = I_n = G_0^{-1} G_0$. Thus, by multiplying its inverse, the term disappears and we now obtain the GVAR (2) model with 2 lags where:

$$x_t \quad = b_0 + b_1 t + F_1 x_{t-1} + F_2 x_{t-2} + \varepsilon_t$$

Where the new terms collect the inverse of G0

$$F_1 = G_0^{-1} G_1, F_2 = G_0^{-1} G_2,$$

$$b_0 = G_0^{-1} a_0, b_1 = G_0^{-1} a_1 \; \epsilon_{it} = G_0^{-1} u_{it}$$

The GVAR model above can be solved recursively, see Pesaran [3]. To summarise, as shown above, the GVAR model allows the interactions among the domestic and foreign economies through three diverse channels. The first is the contemporaneous and lagged dependence of domestic variables x_{it} on foreign variables x_{it}^*. In addition, it also allows the effect and dependence of domestic variables x_{it} on global weakly exogenous variables such as oil and commodity prices. This can also be used as a simulation strategy that can reveal the contemporaneous effects of shocks from country i on j.

Data sources and variables

The current model contains 33 countries of which 8 eurozone countries are grouped into the Euro Area and treated as one country (in the sense of a separate VARX* model). This list of the countries in the model consists of the US, China, Japan, UK, Euro area (Germany, France, Italy, Spain, Netherlands, Belgium, Austria, Finland), Canada, Australia, New Zealand, Sweden, Switzerland, Norway, Korea, Indonesia, Thailand, Philippines, Malaysia, Singapore, India, South Africa, Turkey, Saudi Arabia, Brazil, Mexico, Argentina, Chile and Peru. As it stands, it contains the bulk of the world's output at around 90% [6] p.18.

In terms of variables, there are real output (quarterly in the natural log, seasonally adjusted, with 2015 indexed at 100 for all countries), inflation (constructed from local CPI index, quarterly in the natural log), real exchange rate (constructed from local currency against USD, where USD is set as 1, also in the quarter and natural log), real equity price index (from the local largest stock market index, quarterly and in the natural log), short term interest rate (constructed from the local central bank using interest rate, deposit rates, T-bill rates and money market rates, quarterly averages, in natural log, long term interest rate, constructed with interest rates, government securities and bonds, in quarterly averages and natural log. The datasets also include three global variables, namely oil price, raw material and metal price. The oil price is constructed with the Brent crude index, also quarterly and in log. Both raw material and metal prices are taken from primary commodity prices indices and also in the quarterly log.

Real output	33
Inflation	33
Equity price	26
Exchange rate	33
Short interest	32
Long interest	18
Oil Price	1
Material price	1
Metal price	1

Figure 1 -GVAR data series

Lag orders of individual VARX* models

The table above shows the lag orders selected by either AIC or SBC, whichever value is the highest.

Unit root test

Like many other papers in the literature, the Augmented Dickey-Fuller test is used instead of the older standard Dickey–Fuller test. The ADF test was carried out at 95%, implying if the test statistic for the variable is more negative than the critical values then it will be rejected as there is no unit root. The test was carried out on the level, differenced, twice differenced, with the trend and without trend on all variables namely real output (y), inflation (price level, p), equity price (eq), an exchange rate (ep), short-term interest rate (rs), long-term interest rate (lr).

Testing for Cointegrating relationships

Once the unit root had been tested, the corresponding cointegrating VARX* models are estimated as VECMX*. The next step is the identification of the cointegrating relationships within the individual models. The rank of cointegrating relationships for each model is then computed using Johansen's trace and maximal eigenvalue statistics Pesaran et al [17]. The summary of output from both tests is displayed above. The number of cointegrating relationships found is somewhat different to the result in Dees et al [4].

Figure 2 - VARX order

	p	q
ARGENTINA	2	1
AUSTRALIA	1	1
BRAZIL	2	1
CANADA	2	1
CHINA	2	1
CHILE	2	1
EURO	2	1
INDIA	2	1
INDONESIA	2	1
JAPAN	2	1
KOREA	2	1
MALAYSIA	1	1
MEXICO	1	1
NORWAY	2	1
NEW ZEALAND	2	1
PERU	2	1
PHILIPPINES	2	1
SOUTH AFRICA	2	1
SAUDI ARABIA	2	1
SINGAPORE	2	1
SWEDEN	2	1
SWITZERLAND	1	1
THAILAND	2	1
TURKEY	2	1
UNITED KINGDOM	1	1
USA	2	1

Testing for weak exogeneity

As mentioned before, the main assumption in the GVAR approach is the weak exogeneity of the foreign variables x_{it}^* concerning the respective VARX* model. As described in Pesaran et al [18], this assumption is compatible with a certain degree of weak dependence across u_{it} (the residuals). Following the work on weak exogeneity testing by Johansen [11] and Granger and Lin

[8], the weak exogeneity assumption implies no long-run feedback from x_{it} to x_{it}^*, suggesting that x_{it}^*error correction terms of the individual country VECMX* models do not enter in the marginal model of x_{it}^* Smith and Galesi [21]. This implies we can consistently estimate the VARX* models individually and later combine them together to form the GVAR. The test is a regression model described in Johansen [11] and Harbo et al [9]. The test employed by Dees et al [4] is as follows:

$$\Delta x_{it\ell}^* = a_{ij} + \sum_{j=1}^{r_i} \delta_{ij\ell} ECM_{ij,t-1} + \sum_{k=1}^{s_i} \Phi_{ikj}^{'} \Delta x_{i,t-k} + \sum_{m=1}^{n_i} \Psi_{im\ell ikj}^{'} \Delta \tilde{x}_{i,\ell-m} + \eta_{it\ell}$$

where ECMij ,t1, j = 1,2,. . . ,ri are the estimated error-correction terms corresponding to the cointegrating terms found as shown in previous section. It also should be noted that Δx_{itl}^* is the differenced vector collection of the foreign variables. This is a F-test for the significance of ij, = 0,j = 1,2,. . . ,ri above. While the lag orders of p and q were determined earlier via AIC.

The regression was run on the foreign variables in the VARX* models real output (y), inflation (price level, Dp), equity price (eq), short-term interest rate (rs), and long-term interest rate (lr). and also the global variables such as price of metal (pmetal), oil (poil) and raw material (pmat) with a 5% significance level.

Forecasting

Recall that the GVAR is constructed by stacking multiple VARX* models. In our case, we have estimated 33 individual VARX* (p,q) models with variable lags and stacked them together and became a GVAR (2) model. We now show that forecasts can be made from the generic GVAR (p) and applied the method to our study. Recall that the individual VARX* (2,2) i.e. two lags for both domestic and foreign variables:

$$x_{it} = a_{i0} + a_{i1}t + \varphi_{i1}x_{i,t-1} + \varphi_{i2}x_{i,t-2} + \Lambda_{i1}x_{it}^* + \Lambda_{i1}x_{it-1}^* + \Lambda_{i2}x_{it-2}^* + u_{it}$$

Where x_{it}– is a vector with a dimension of ki × 1 of domestic macroeconomic variables indexed by individual country i and time as t; x_{it}^*– is a vector with a dimension of ki × 1 of foreign macroeconomic variables indexed by individual country i and time as and uit – is a serially uncorrelated and cross-sectionally weakly dependent process. This can be re-written into:

$$A_i(L, P)W_i x_t = \phi_{it}$$

Where ϕit equals x_{it}, L as the lag operator; p as the domestic variable lag orders; W as weight matrix and x_t as the domestic variables denoted in t and i denote the country. In other words, it is simply a re-statement of the VARX* model as a function of domestic variables with lag orders multiplied by their corresponding weights. Also recall that, once the VARX* models have been estimated individually, the next step is to stack the models together to form the GVAR model.

Again, using the notations in Dees et al [4], by stacking the individual VARX* models (written as ϕit), we obtain the GVAR (p) model as

$$G(L, p)x_t = \phi_t$$

Where

$$G(L,p) = \begin{pmatrix} A_0(L,p)W_0 \\ A_1(L,p)W_1 \\ \vdots \\ A_N(L,p)W_N \end{pmatrix}, \varphi_t = \begin{pmatrix} \varphi_{0t} \\ \varphi_{1t} \\ \vdots \\ \varphi_{Nt} \end{pmatrix}$$

The GVAR ex-ante forecast model has now formed and can be solved via recursive method at any horizon N.

GVAR ex-ante forecasts

We now turn to the results produced by the estimated GVAR model. As mentioned before, there are 33 countries in total with 8 euro countries which will be estimated as one, therefore there are 26 country models. Each has its combination of lag orders up to a maximum of 2 as determined by AIC/ BIC. It should be noted not all VARX* models have equivalent lag orders nor the same set of domestic and foreign variables due to the specification tests of lag order and weak exogeneity in the last section. In the end, after removing the variables which did not meet the weak exogeneity assumption, we have 271 variables estimated placed in 26 VARX models and one auxiliary model for global variables such as oil price, metal and raw material price for 8 quarters i.e. 2 years. This means 2184 point estimates were created for all variables.

Estimating FAVAR model

FAVAR model

The next few sections are devoted to the exposition of the methodologies behind in the order of the motivation behind, the FAVAR framework, factors selection with principle components, the specification and estimation of the model and also structural identification for structural impulse responses.

Two equations approach

$$y_t = (F_t^{'}, z_t^{'})'$$

$$x_t = \Lambda^F F_t + \Lambda^z Z_t + e_t$$

Observation equation - where x_t is a T by N panel data matrix which contains large datasets of economic and relevant variables. Z_t is the variable of interest that we are trying to explain ('endogenous ' variable and used for impulse response analysis) is linked to the sum of x_t contemporaneously (i.e. large datasets of economic and relevant variables). F_t is a T by K matrix of unobserved factors which summarise the important information in x_t. Λ^f is the factor loadings.

$$B(L)\begin{bmatrix} F_t \\ z_t \end{bmatrix} = \omega_t$$

$$B_t = c + \sum_{p=1}^{P} \beta_p B_{t-p} + \omega_t$$

Essentially, the steps of estimating a FAVAR can be summarised in 5 steps below:

Step 1) Approximate F_t as K principal components of x_t where x_t is stationary and standardized

Step 2) Rotate the principal components to obtain $\widehat{F}_t$

Step 3) Estimate a FAVAR using $\widehat{F}_t$ and z_t and estimate the impulse response to a policy shock using Cholesky decomposition.

Step 4) Calculate impulse response of $\widehat{F}_t$ and z_t to the policy shock.

Step 5) Calculate the impulse response standard errors using a bootstrap method.

Principle component analysis

Following Stock and Watson [20], Bernanke et al [2] also used a two-stage procedure which used principal component analysis (PCA) to estimate the factors before estimating the VAR model. Principal component analysis (PCA) is a procedure that converts a set of observations x_t which are potentially correlated to Z_t into a set of factors (to be plugged into the FAVAR model) of linearly uncorrelated variables called principal components or PCs. When the transformation is completed, the first PC would have the largest possible variance and the PCs after would contain the remaining variance in the data, which the PCs must be orthogonal to the preceding PCs.

$$\arg\min_{\hat{F}_t, \hat{\Lambda}_i} \sum_{i=1}^{N}\sum_{t=1}^{T} (x_i - \Lambda_i F_t)^2$$

Essentially the PCA estimation above is aiming to find the minimum number of factors needed to explain the datasets, in which the distance between the factors is minimised and also orthogonal to each other so that they are distinct variations. The eigenvector associated with the largest eigenvalue indicates the direction in which the data has the most variance. Similarly, the second largest eigenvalue in the associated eigenvector is orthogonal to the largest eigenvector, in which the data has the largest remaining variance. As PCA is sensitive to the scale of the data, the common practice is to convert the data into stationary and also standardisation i.e :

$$Z_{ij} = \frac{X_{ij} - \bar{x_j}}{s_j}$$

Where X_{ij} from data for variable j in sample unit i , $\bar{x}_j$ for the sample mean for variable j and s_j for sample standard deviation for variable j.

The PCs are then rotated so that they are orthogonal. The factors remain uncorrelated and variances are preserved. In terms of the number of PCs to be used in the models, common methods include looking at the scree plots of the PCAs (if available) visually or using formal statistics such as the information criteria in Bai and Ng [1]. However often, it is simply an exercise depending on the output such as the shocks on variables for impulse responses. For example the authors increased the number of PCs k until there is no change in impulse responses. They found that the first three principal components capture the information in the dataset sufficiently and additional PCs did not contribute much.

Estimation of FAVAR and structural Identification

Now recall that the observation equation below, where $B_t = \widehat{F}_t, Z_t$, from the estimation of the factors by the PCA ,we should now have the factors so that the rest of the equation can be estimated.

$$B_t = c + \sum_{p=1}^{P} \beta_p B_{t-p} + \omega_t$$

In this case, ordinary least squares (OLS) can be used to estimate the equation above. In the example from Bernanke et al [2] for example, the authors used Cholesky decompilations with ordering $\widehat{F}_t$ based on Z_t. For example, the variable of interest was the federal fund rates and the authors were interested in separating the economic variables into slow and fast-moving variables. The recursive ordering here implies that certain series such as equity prices, and price index are likely to be affected first therefore they are ordered last and with slow variables ranking first such as GDP. Similarly, if we are interested in the sole effect of monetary policy only, to identify such effect, recursive ordering can be applied below such that:

$$\left(f_t^{s'}, z_t^{s'}, r_t, f_t^{f'}, z_t^{f'}\right)'$$

Variables are grouped into slow and fast-moving groups (s for slow and f for fast) which implies that the factors are needed to be extracted separately from those two groups. When the variables above e.g. $f_t^{s'}$ and $f_t^{f'}$ are ranked above or below the interest rate, therefore the fast-moving factors can be instantaneously affected in a lower triangular recursive identification scheme. In this case, this restriction implies that the central bank which has the control of r_t couldn't be affected by the fast-moving variables as they react to the change in r_t instantaneously (within the period *t*, so that cannot be observed). Further restrictions can also be applied for the identification schemes, such as sign restriction or imposing zero factor loadings so that the impulse responses would react accordingly (see chapter 16, Kilian and Lütkepohl [12]). Similarly, impulse responses can be obtained akin to other VAR-type models when after estimating the FAVAR model. Bootstrapping is often then used for approximating the distribution of the impulse responses, although there are no formal criteria for the draws required.

3. BACKTESTING METHODOLOGY - CONSTRUCTING THE BACKTEST

The importance of backtesting is noted in the industry as many quantitative funds heavily rely on them to achieve 'alpha' - excess return when adjusted for risk. The typical backtest would show the profit and loss of the strategy, volatility measures, ratios and risk-adjusted return, typically with the Sharpe ratio. Sharpe ratio is the most often used metric to evaluate the investment performance of a strategy. Although the Sharpe ratio is often discounted in practice and with good reason, the metric can still be fully used to assess a trading strategy that is based on forecasts.

The section below outlines how the backtester is set up and the trading rules selected for testing. The challenge of identifying working / profitable trading ideas has gathered a lot of attention from industry and academia and countless efforts have been made to identify market abnormalities.

Investor preferences - Risk and Reward

The first thing that needs to be determined is the size of the trading capital. This can be defined as the amount of capital that the investor is willing to risk, should all the investments have failed and result in a 100% loss. After setting the initial trading capital, say £50,000, the investor is then required to state the risk and reward that is expected from this capital. The risk and reward preference by the investor will then be used as a guide to set up the rest of the portfolio and trading positions. In this case, we will assume the investor has £50k free cash that is ready to take on the risk and potential rewards that it may provide. Based on the historical returns on the share market which is around 7% annually, the investor decided that it is not enough, therefore, he would like to trade into the oil futures market, which is much more volatile, thus increasing both the potential risk and reward. The statement above means that the investor is ready to set aside a bigger pile of capital to gain an extra return from the market. This function of stating the risk that an investor is willing to risk during trades, not the entire capital, is the goal of volatility targeting. This function will require the investor to answer how much can be lost from the portfolio.

How much risk can the investor cope with? For example, it would be very difficult to make an annual return of 15% if the investor is only investing in high-grade government bonds since the return is much lower around 1-2% annually. In this example here, the investor is willing to lose up to 20% annually of the initial £50k capital i.e. £10k. This figure was chosen as it is a relatively high figure that suits a higher volatility instrument such as commodity futures,

Setting a reasonably benign threshold would allow the trade to still be in the market before being automatically closed due to hitting the target by natural fluctuation in the price. Furthermore, this is also a good figure that would not allow trades entering into a permanent loss amount such as -50% where a 100% gain is required to break even. To find the daily volatility target, given the annual target will need to divide the annual figure of £10k by 16, this is due to there are normally 256 working days in a year and the daily figure is derived by taking the square root of time i.e. SQRT (256) = 16.

This is the maximum daily target volatility the investor is willing to meet. Next, we need to consider the actual price of the instrument and its related price volatility. Price volatility can vary greatly throughout the year. This is particularly true in early 2020 as the Covid 19 pandemic broke up, wiping out global oil demand. The volatility during normal is around +/- 5% per day but could become more than 20% during stressed periods. There are many ways to calculate and estimate future volatility, GARCH for example is one of them. In the example here, it is simply assumed +/- 5% per day. This figure can be changed by re-estimating such as finding the previous volatility but calculating the standard deviation from the open, high, low and closing price. The most distinguishing feature of this figure is that volatility is a penalty on the trading position. It is designed in this way to ensure that the trader would not exceed the daily volatility. The daily volatility was approaching 20% during early 2020 and this high figure would automatically reduce the tradable position for the trader.

At the beginning of 2017, the oil price was £53.14 per barrel and the minimum trade size is 5 contracts, therefore the minimum trade position would be £265.7. Assuming a normal 5% daily volatility translates to about £13.28 volatility per block of trade. Under this scenario, we would require to trade at least 47 contracts i.e. £12.487.90 of net position to fully achieve the 20% daily

Translating forecasts into tradable quantity

The example given here is set to 5% and kept constantly. While holding this throughout the period, the investor will be able to see how the final position is determined, based on other changing quantities such as forecasts and changing trading capital. If the investor has lost money and the trading capital is down to say £48,538 then the daily volatility target will also be lower at £607 as the portfolio has now become smaller but the 20% is capped at the same level. This automatically ensures that the investor is sticking to the volatility target set out initially so that no excessive losses will be made, despite lower or higher capital. The most important factor that determines the final trade position is the forecast.

In the example here, it has been constructed with the mind of standardising the final position according to the risk level. Given the constant 5% risk, a forecast of +1 will allow the investor to trade 47 contacts at £53.14 for the oil price. This is because 5% of these 47 contracts equal to £12,487.90 of which 5% is the daily volatility target of £625.00. In order words, this gives the investor a multiple of 47 to fully achieve this target. However, during the height of a market meltdown expected volatility of 30% is expected then, and a multiple of 7.84 is allowed, which is rounded up to 8 contracts only, given the exact +1 forecast. The examples in the final part will demonstrate how the forecasts were created from the macroeconometric models and how it was then applied to the backtesting tool.

4.CONSTRUCTING THE TRADING RULES AND BACKTESTING

Once the investor has set out the goal for the portfolio, it is now required to determine a trading rule or strategy. This part discusses some common strategies and further resources on how to develop them.

Trading rule

Long Short Moving Average Crossover

The daily price change of an instrument can appear to be random as indicated by the efficient market hypothesis (EMH). However, this is often not the case when we consider market trends in the long run. For example, the majority of equities show unrelenting growth, despite experiencing major global wars and macroeconomic crises. This gives the first idea that at the minimum some part of the trend can be forecasted and taken advantage of. Calculating a moving average say 5 days would capture the average price of an instrument of the previous 5-day averages while not being disturbed by daily movements. A moving average of 5 previous days or MA(5) is just a historical mirror looking at the trends as they were revealed. If however, we use another, slower MA that says 25 days, which represents a whole working month, then this line also captures trends and effects that have been developing but at a slower speed. This combination of a fast and a slow trend line is the basic principle behind many rules.

While events like currency movement would be reflected almost immediately by the financial market, some are less obvious. Examples include the interest rate change. Although interest rate change surprise often accompanies a sudden change in share price, for example, it also has a lingering effect that may last a longer period until other companies/industries price in this information. This is particularly obvious for events like a sudden event that has not been anticipated. Take the most recent coronavirus which was first reported in Jan 2020. Its effects were already known by the time of Feb 2020 but no significant drop until March which saw the majority of markets drop below 30% or more. Oil futures had also crashed to sub-zero prices for the first time in history. This is a case of showing that different information travel at different speed therefore there are at least more than one trend happening at any one moment. To identify the most recent, quicker changes/trends, a short, quick-moving average of 5 days would be sufficient.

The answer to the problem of identifying a slow and fast trend is often solved by plotting two MA lines on the chart. Whenever a fast-moving average is now above the slow-moving average then this signals a buy. The opposite of the fast MA dropping below the slow line would then signal a sell as it now enters a decreasing trend. If we are to express this particular rule into pseudo codes that can be inputted into a programme then it would be below:

"If the previous day the fast SMA was below the slow SMA and the current trade day there is a change where the fast SMA is now above the slow SMA, then signals Buy. However if, the previous

day the fast SMA was above the slow SMA and on the current trade day there is a change where the fast SMA is now below the slow SMA, then signals to Sell. If none of the conditions is met then, do nothing."

Backtesting

This trading rule would then determine whether the investor should buy or sell or do nothing at all. The price data depicts oil WTI futures from the beginning of 2017 until the most recent available data. By the end of this period, the trading rule would have triggered 40 orders with 20 buys and 20 sells. In the end, the user can then see the overall performance of the trading rule. Sharpe ratio, highest / lowest returns and a host of other performance metrics are also readily available to be used for assessing the performance. It should be noted that two different trading scenarios are demonstrated here.

Single position long short moving average cross over

The first is a single position trading only. This is a simplified version where when the trading rule has been triggered, it will buy or sell one contract only. It is assumed that one contract is the same price as the closing price. This is a simplified version of reality as the actual cost is likely to be different to the closing price if the trade was executed some other time. However, this does not distract the point the backtest is trying to show here, which is to test whether the trading rule is profitable.

Once the trading rule criteria have been hit, it will trigger an either buy or sell signal. This allows the trader to hold one position at any time, either buying or shorting. Therefore the trader's net position can either be short (-1), no holding (0) or long (+1). For example, the first trigger was a sell signal, therefore the investor will sell 1 contract, holding a negative one position -1. The next signal will then be a buy signal when the rule has been triggered. By buying back the 1 contract, the investor now has a null net position i.e. 0 contract thus realising either profit or loss from this trade. As a result, the total trades would be even-numbered in the end, with the alternation of one buy and one sell etc, flip-flopping until the test finishes. To be precise the trading rule has been inscribed in the command line below:

= IF (AND (SMA; 5t>SMA; 25t , SMA; 5t_1<SMA; 25t_{-1}), "BUY" ,IF (AND (SMA; 5t<SMA; 25t

, SMA; 5t_{-1}>SMA; 25t_{-1}), "SELL" ,""))

Say for example, on 07/03/2017 the trading rule has decided to execute a sell order, the trade would then purchase the oil futures at the closing price of 08/03/2017 thus, waiting until the end of the day. This is not ideal in reality, however, as the trader is most likely to have traded before the close due to liquidity or other concerns. Thus this no-forecast case requires the user to rely on perfect information i.e. when the end of the day has reached so that the investor has complete information to calculate the moving averages. The second and third case is to use forecasting models. The FAVAR and GVAR forecasts are then placed onto the backtest. The trading rules now rely on these values to determine whether a trade would be triggered. Compared to the no forecast scenario, this is more realistic as it reflects the way actual traders behave that they always forecast, albeit with different models or purely on gut feelings. It is expected that the no-forecast / perfect information case will provide the best return as it has full information. Therefore the closer the forecast result the better it will be.

Multiple positions long-short moving average cross over with portfolio Management

The second trading scenario is a more complex one as it allows more than one position to be transacted and held by the investor at each trade and at any time. This increase from a single position introduces a few challenges for the modeller. The first and perhaps most important is how should the investor allocate his preference of risk and reward. During the single scenario, it didn’t matter because the investor can hold only one position at any time, therefore, there is no need to introduce any trading capital requirement or risk target as the investor will be exposed to exactly one contract equivalent to risk and reward, no matter what the action is. The only way to avoid this risk and reward problem in holding a single position is to not trade altogether. Therefore, portfolio management principles or tools are not applicable for that scenario since there is only one position.

Now that we have introduced multiple positions, the investor can now purchase more than one contact at any given buy/sell triggered by the rule, depending on the investor’s conviction of the forecasts and the trading rule. From this, I have calculated a buy/sell strength indicator. The higher the indicator, the bigger size is the investor going to purchase or sell. This exact size of trade for transactions entirely depends on the forecasts. As mentioned earlier, the forecasts are also standardised by volatility before being translated into actual positions. This is to ensure that the trader adheres to the principles and volatility target set up in the beginning. Another assumption here which was explained previously is that the price volatility had been assumed to be 5% here. By holding the volatility to a fixed level, we can then determine whether other parts of the strategy can still be profitable. In practice, however, this assumption is likely to be invalidated and will need updating.

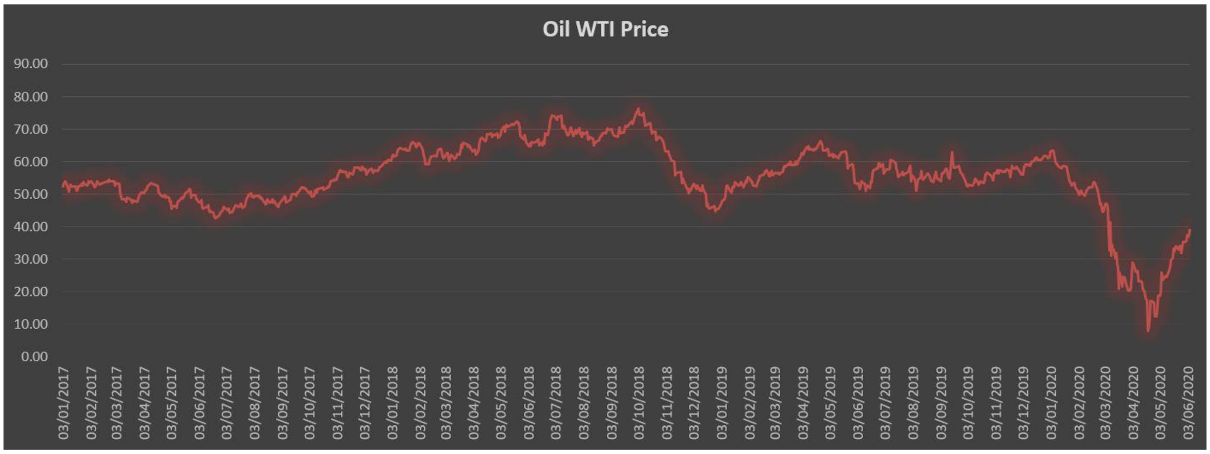
A very important implication here is that, when a forecast has been made correctly and the forecast strength has also been calculated correctly then, there is a much bigger room for profits and losses. This is due to the fact that the investor may no longer be restricted by one position only. Say that the forecast from GVAR is asking the trader to purchase 10 contracts while the no forecast case is only asking for 3, therefore whatever the profit and loss prompted by the forecast will be much bigger than that of the no forecast case. Another important implication of holding multiple positions here is that the returns and risks are much smoother now. This is simply due to the diversification of positions. In the single position scenario, the trader will make profit and loss solely on the correct direction of the trade, therefore if he wins, he wins big and vice versa. This, however, does not apply to the multiple holdings. This is because if the trader was prompted to sell say 5 contracts on the first trade, which results in a net position of -5 contracts, but the forecast has prompted a buy signal of +12 therefore, the investor would first buy back the first 5 contracts he had shorted at the new trade price and then buy a further of 7 contracts at the same price. Therefore by the end of this trading day, the investor now holds +7 in his net position. The mixing of clearing one's position from the previous trade plus adding new contracts according to the signal has resulted in a smoothing of the returns and also a risk. This is evidenced by the highest and lowest returns in the example. The no forecast case in the single position strategy has the highest return of 50.5% and the lowest of -12.3% for a single trade. While the multiple positions of the no forecast approach have the highest return of 16.8% and lowest of -2.2% only. This, of course, was also a result of setting the maximum annual volatility of 20%, therefore, the final return and risk would have also decreased as a result.

5. EMPIRICAL RESULT 1- SINGLE POSITION LONG-SHORT

Introduction

Having laid out the groundwork for the rationale and methods of the backtest in previous sections, this part proceeds to show how the to backtest was carried out. The first example here is for the single position with the trading rule of long-short moving average cross-over. The financial instrument chosen for this example was crude oil West Texas Intermediate (WTI). In theory, any financial instrument can be chosen for this backtest as long as there is historical data and the trading rule can be applied. The reason to choose oil price though is not entirely arbitrary. It is very common to see oil prices riding a very long continuous uptrend or downtrend.

Figure 3: Crude oil WTI price 2017-2020



The longest uptrend of oil price began from the end of 1998 at $12 per barrel to a peak of $140 in August 2008 before crashing down to $34 at the end of 2009. A recent coronavirus outbreak also saw the price reaching subzero first time in history. An investor who purchased a barrel of crude in 2017 at $50 would have ended up with an almost 100% loss as the price crashed in April 2020. However the actual trading is more complicated as the price itself is a benchmark, comprising a range of oil that was produced according to this specification. There is also a difference between the spot and futures market in which the buyer although may lose 100% in April 2020 due to an oil glut and virus outbreak, the investor can also sell the oil in the further months say June, July and so forth which the prices were still trading at a positive.

Oil prices and market

This forms the hypothesis that I am going to backtest: whether a long short moving average cross-over strategy would have been profitable for trading the oil market. The second hypothesis would then be whether the forecasts created from FAVAR and GVAR are contributory to the profits.

Three cases: No forecast, FAVAR and GVAR forecasts

No forecast

The first case is for no forecast or perfect information. For a single position trading, this is the best-case scenario in which the performance obtained here will be optimal. The success rates of the buy/sell and performance rely on the trading rule entirely. The oil price data has been populated on the table, below starting from 03/01/2017 to 05/06/2020. It should be noted that the data is recorded on business days only. The data was compiled and published by NYMEX for the Crude Oil WTI price which always uses the front-month futures for calculation. For example, if we are looking at the data on 15th May 2018, the data will then be showing the WTI futures contract price deliverable for June 2018, then for one month in advance. As soon as the cut-off for the June 2018 contract has passed i.e. on the 20th of May, the oil price will automatically collect the data of the July futures contract since July is now the nearest, yet unexpired month for a futures contract. The data here, therefore, is showing a rolling front month from the futures market price. The data is in daily format and also includes the open, high, low and close prices.

The trade signal of buying or selling or doing nothing is then generated in the next column, using the SMAs calculated. Per the explanation in the last section, the signals will be generated whenever it has met all conditions. The next column then records the transaction price which is the closing price one day after the signal has been generated. This will automatically roll over to the next transaction price when the next transaction has been triggered. It is prudent to point out that the return calculation of covering for a sold position is the opposite of selling a long position. The below example shows how this is calculated.

Figure 4: No forecast rule performance

On 07/03/2017, the trading rule detected its first pattern that meets all conditions and triggered a Sell signal. From this, the trader will purchase on 08/03/2017 at the closing price of $50.28. Thus the trader now has a net position of -1 contract. The next signal was triggered on 04/04/2017 and it was for a buy. Thus the trader bought one contract back on 05/05/2017 which cost $51.15 per contract. The total return from this trade would, therefore, be $50.28 - $51.15 =$-0.87 which meant the investor had lost -1.73% from this trade. This is because the trader sold at a cheaper price but bought back higher at $51.15. Take another example, the trader sold 1 share on 10/10/2017 at $50.92. Since the cost of buying it was only $47.48, the total return was therefore 7.25%, making a profit.

The above graph can be seen on the single position backtest tab. It pairs the cumulative return of holding an Oil WTI with the cumulative return of the trading strategy. The total returns were calculated in the second tab - performance comparison. It is simply an addition of all the positive and negative returns. From the graph above, we can then see that the total return for holding an oil contract has given the trader a considerable gain from 2017 to 2018. However, the long-term decline from the end of 2018 had led to a downward trend which erased all gains by the end of 2018. Although the oil price had picked up again throughout 2019, the massive crash, in the end, has again completely wiped out the profits. In the worst-case scenario, the contract would have been sold automatically to recoup some losses or per margin call that was issued by the broker. The point here though is that, a buy-and-hold strategy has failed quite dramatically. On the other hand, the strategy has proven to be quite versatile and accurate in its prediction. During much of the time, its performance has surpassed the buy and hold strategy. This is mostly due to the availability of shorting but also the accurate calls of the buy/sell signals.

Figure 5: Performance (no forecast)

No Forecast	
Performance Metrics	
Positve Returns	153.87%
Negative returns	-67.17%
Positive Trades	19
Negative Trades	20
Hit Ratio	49%
Average Returns	2.22%
Highest Returns	**50.5%**
Lowest Returns	**-12.3%**
Sharpe Ratio	**0.22**

The total, combined positive return for this strategy was over 150% while the negative return equalled -67%. Out of 39 trades that were triggered, 39 of them were positive while 20 were negative. The first trade was a sell order whilst the last is a buy order. Since buying the last contract, the rule was not triggered again therefore the last net position is 0. Although the hit ratio is just under 50% because of the big returns that were made during the shorts, particular in the 2018 year-end and 2020 April, this heavily skewed the total returns from positive trades. The Sharpe ratio was calculated at 0.22 which was taken by using the mean of returns and divided by the standard error of the returns. This Sharpe ratio will now be used as the benchmark comparing the other two forecasts.

Forecast with FAVAR

The next case is to test whether using forecasts produced from a FAVAR model can give any contribution. Given that there are always errors in forecasts and as such when compared to the perfect information case, it is expected that the strategy performance using price forecasts is going to be less accurate. The aim is to input as much relevant data as possible and forecast the oil prices

Figure 6: With FAVAR forecasts

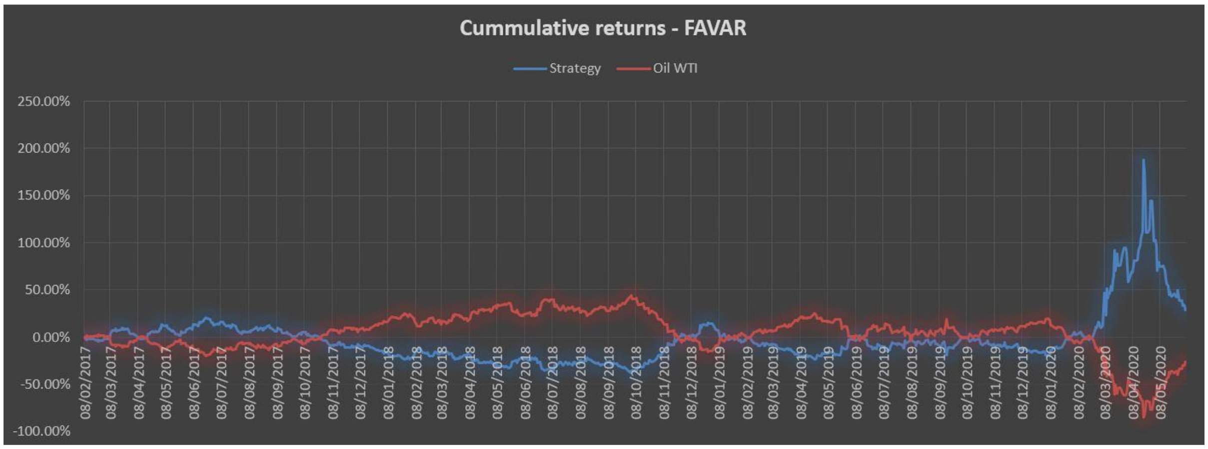


Figure 7: FAVAR Performance

FAVAR	
Performance Metrics	
Positve Returns	154.60%
Negative returns	-81.96%
Positive Trades	19
Negative Trades	20
Hit Ratio	49%
Average Returns	1.86%
Highest Returns	**43.6%**
Lowest Returns	**-12.2%**
Sharpe Ratio	**0.18**

would yield a satisfying performance, which should be as close to the perfect information as much as possible. The independent variables are sorted into fast and slow groups. This sorting is similar to the trend lines of 5 and 25 days to differentiate near and long-term trends. The trading rule now uses the forecast values from the FAVAR model. The trade signals are now triggered based on the SMA5 and SMA25 that were derived from the FAVAR forecasts. The FAVAR backtest was also run similarly to the one for the no forecast case. The metrics are showing surprisingly good performance with a positive return of 154% but a negative return of -82%. In general, the metrics are not too different from the no forecast scenario as it has the same hit ratio and similar returns pattern. However, when we look at the cumulative performance above, we can see that the strategy was only profitable for a short period in 2017 before losing ground to the buy and hold strategy. The year 2019 also saw a poor performance from the strategy where it triggered the buy and sell signals too early thus making a loss. However, the reprieve came during early 2020 when the strategy entered into a sell which logged a 50% return and several other trades that had the correct signals. This has heavily skewed the cumulative return to a peak of 180% before lowering to 50% in total. The overall Sharpe is similar to the no forecast case at 0.18.

Forecast with GVAR

The aim is to input as much relevant data as possible and the forecast of the oil prices would yield a satisfying performance, which should be as close to the perfect information as much as possible. The trading rule now uses the forecast values from the GVAR model. The trade signals are now triggered based on the SMA5 and SMA25 that were derived from the GVAR forecasts. The GVAR backtest was also run similarly to the one for the no forecast case. The GVAR has so far fared worse when compared to both FAVAR and perfect scenario cases. The strategy lost money most of the time during 2018 and 2019 but similarly gained a reprieve in 2020. The total negative return is the highest in all three tests at -90.22% while the gain is the lowest at 148.88%. The hit ratio is also at 44% with 17 positive trades only and 22 for negative. Therefore it is not surprising to see that the Sharpe ratio is also the lowest at 0.14.

Overall it has a similar pattern to the other two cases but delivered lower returns. Looking at the transaction times and we can see that some of them were triggered too early thus making a loss. However, it also performed very well during the last period of the crash.

As mentioned in the beginning, a priori belief is that no forecast / perfect information will perform the best. This was expected as it had the most information available and assumed that there is no forecast to be made by the investor at all. This, of course, is difficult to carry out in practice as most trades are not triggered at the end of the day due to liquidity reasons. The main disadvantage of a moving average cross-over is also observed during the live trading session. This is simply because it is essentially a backward-looking indicator and does not predict anything in itself unless, of course, we put lines together so that hoping to find out the slower trend lines and the reverse point. As a result, regardless of the methods used in reality, the trader will always engage in some kind of forecasting, either implicitly with gut feeling or experience or with a macroeconometric model.

Figure 8: With GVAR forecasts

Figure 9: Performance (GVAR)

GVAR	
Performance Metrics	
Positve Returns	148.88%
Negative returns	-90.22%
Positive Trades	17
Negative Trades	22
Hit Ratio	44%
Average Returns	1.50%
Highest Returns	**45.8%**
Lowest Returns	**-13.2%**
Sharpe Ratio	**0.14**

The most surprising result from above is perhaps that the FAVAR model did quite an impressive job of matching the perfect information hit ratio. This backtest, of course, similar to others is based on a set of assumptions and by changing those assumptions, the results would have appeared different. One of the biggest assumptions in this single position is that it does not require the trader to purchase/sell multiple positions. As a result, it ignores most portfolio management elements that are as important as the forecast itself.

6. EMPIRICAL RESULT 2 - MULTIPLE POSITIONS LONG-SHORT

The second trading scenario is more complex but more realistic. The main difference here compared to the single position backtest is the introduction of multiple positions and the risk and reward criteria. Now that we have introduced multiple positions, the investor can now purchase more than one contact at any given buy/sell triggered by the trading rule, depending on the investor's conviction of the forecasts and the trading rule. From this, I have calculated a buy/sell strength indicator. The higher the indicator, the bigger size is the investor going to purchase or sell. This exact size of trade for transactions entirely depends on the forecasts. The forecasts are also standardised by volatility before being translated into actual positions. This is to ensure that the trader adheres to the principles and volatility target set up in the beginning. Another assumption here is that the price volatility had been assumed to be 5% here. By holding the volatility to a fixed level, we can then determine whether other parts of the strategy can still be profitable.

A very important implication here is that, when a forecast has been made correctly and the forecast strength has also been calculated correctly then, there is a much bigger room for profits and losses. This is because the investor may no longer be restricted by one position only. Say that the forecast from GVAR is asking the trader to purchase 10 contracts while the no forecast case is only asking for 3, therefore whatever the profit and loss prompted by the forecast will be much bigger than that of the no forecast case. Another important implication of holding multiple positions here is that the returns and risks are much smoother now. This is simply due to the diversification of positions. In the single position scenario, the trader will make profit and loss solely on the correct direction of the trade, therefore if he wins, he wins big and vice versa. This, however, does not apply to the multiple holdings. This is because if the trader was prompted to sell say 5 contracts on the first trade, which results in a net position of -5 contracts, but the forecast has prompted a buy signal of +12 therefore, the investor would first buy back the first 5 contracts he had shorted at the new trade price and then buy a further of 7 contracts at the same price. Therefore by the end of this trading day, the investor now holds +7 in his net position. The mixing of clearing one's position from the previous trade plus adding new contracts according to the signal has resulted in a smoothing of the returns and also a risk. This is evidenced by the highest and lowest returns in the excel example.

The no forecast case in the single position strategy has the highest return of 50.5% and the lowest of -12.3% for a single trade. While the multiple positions of the no forecast approach have the highest return of 16.8% and lowest of -2.2% only. This, of course, was also a result of setting the maximum annual volatility of 20%, therefore, the final return and risk would have also decreased as a result.

Portfolio management criteria

Data and period

The data used here is the same as the one used in the single position test for the exact period as well. There is however a difference in terms of the dates recorded in the multiple positions tab. This is because only trade dates are recorded for the calculation of profit and loss and other metrics. For example, when a sell signal was generated on 06/03/2017 and the trade was placed on 07/03/2017 then the entry date of 07/03/2017 will be recorded. When the next signal for a buyback appeared on 03/04/2017 then the trade will be executed on 04/04/2017 and the date recorded. Therefore all trades were recorded in the same method for the succession of all 40 trades.

Risk and reward preference/metrics

The assumption here is that the investor is willing to risk up to 20% of the capital in total on an annual basis. The initial capital is £50,000 therefore the annual cash volatility target is £10,000.

From this, the daily volatility target can be found by divided 16 as this is the equivalent of the annualised daily target. The initial daily volatility target is, therefore, £625.00. It will change according to the capital size. Therefore if the capital is now £49,236 then the daily volatility target is now £620.65. The change in the capital is, of course, linked to the profit and loss when trades occurred. For example, a sell signal was generated on 06/03/2017 and the transaction happened on 07/03/2017. The buyback happened on 04/04/2017 and resulted in a loss of £211.00, therefore, the capital will now be £49,789. Any new trade to take place i.e. the additional positions that the trader needs to take to reach the recommended position will be determined by the daily volatility target calculated from the previous trade date. For example, the daily volatility target for trade to be taken on 25/04/2017 will use the value from the last trade date which is 04/04/2017. This is reflecting the order of actions that follows: 1) new trade date first clear the previous positions based on the new buy/sell strength. If there are additional trades that need to be purchased or sold then the trader will either buy or sell new additional positions with the new trade date's price.

The profit and loss at the end period are then simply calculated by combining the daily returns. There are 40 trades therefore 39 daily returns are available. The first trade regardless of whether it is a buy or sell would not show a return unless it was cleared therefore there are 39 returns available. Similarly, Sharpe ratio is calculated for each strategy for the whole period. Other metrics are also available to describe the details of the strategy and returns.

Three cases: No forecast, FAVAR and GVAR forecasts

No forecast

The first case here is for no forecast / perfect information. The assumption is similar to the ones in a single position. The biggest difference is in how the forecasts were converted into buy/sell signals which in the end translated into actual trading positions. Following the methods as set out and assumptions as described above, below graphs and tables show the produced metrics.

The graph shows the cumulative return of just over 3.5 years from the start of 2017 to June 2020. The strategy has returned 21.94% or £11,171 in total. Similar to the single position backtest, there are also 40 trades here. This is because the trading rule is the same therefore prompting the same trade dates. The only difference here is in the magnitude of the buy/sell strength and thus the final positions. The profit at the end period here is much less than the single position strategy as the overall return was over 80%.

Figure 10: Performance (no forecast)

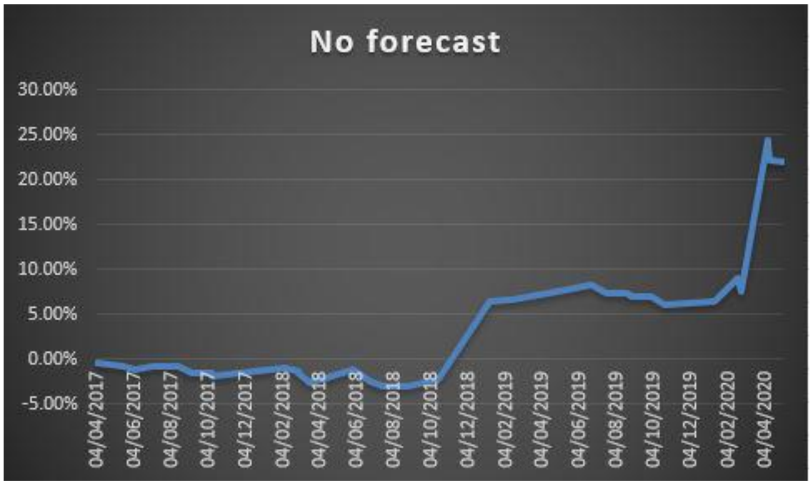
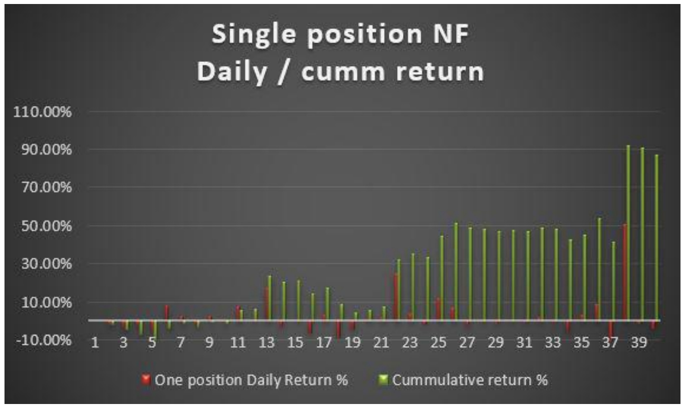


Figure 11: Performance table

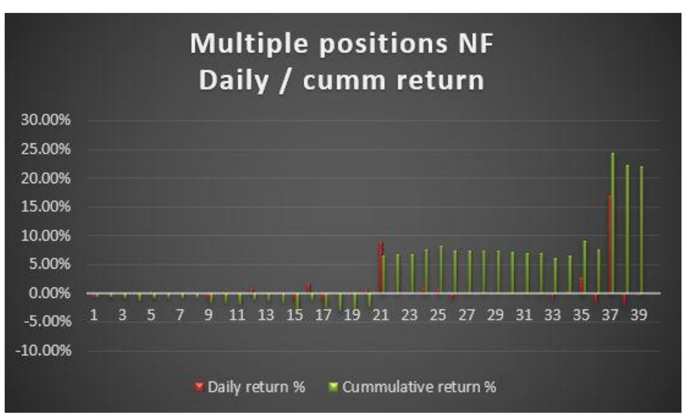
Capital	£50,000		
Annual percentage volatility target	20.0%		
Annual cash volatility target	£10,000	Average strength	0.05
P&L at end of period	£11,171	Average strength	0.05
Cummulative return %	21.94%	Higest Buy	2.82
Average return per trade%	0.6%	Lowest Sell	-1.10
Highest return %	16.8%	Total strength	2.05
Lowest return %	-2.2%	No. of Buy	20
Average position	0.05	No. of Sell	20
Maximum position	323.00	Sharpe ratio	0.18
Minimum position	-81.00		

Figure 12: Performance return



The shape ratio is also marginally smaller at 0.18 compared to 0.22 for a single position. It would, therefore, be reasonable to ask when should we still keep this strategy neither the profit nor the Sharpe is better than the single position? The main reason to use multiple over a single position is a matter of scale. For example, the maximum size allowed with one position is just one contract or the minimum that is requested by the broker. The obvious downside to this is the inability to upscale the strategy. It was only applicable to whatever the single contract was worth. Higher capital would not be applicable as it cannot add more positions. However other than this reason, there aren't many reasons why multiple positions should be taken as the primary. While the strategy is profitable, the majority of the profits were derived from trade 21 and 37 both of which were shorts. Looking at the metrics, the average return per trade is 0.6%. The biggest advantage of multiple strategies is that the lowest return is much smaller at only 2.2%. This is due to the volatility target that was set out in the beginning. The average buys and sells strength is near 0 which is expected as there are buys and sells of 50% each. The maximum position that was held by the trader is 323. The minimum size of a contract is assumed to the 5 barrels therefore 323 contracts would be equivalent to 1615 barrels. The biggest short position was -81 or -405 barrels.

Figure 13: Performance return



Forecast with FAVAR

Using the same data the trading rules are now generated from FAVAR forecasts. The cumulative return is now 19.3% or £10,300. This is marginally lower than the no forecast case. On the other hand, the highest return was just 4.3% and the lowest was -5.2%. The maximum long position at any one point was 546 contracts while the minimum position is only -15. This result reflects the fact that the strategy has a stronger preference for a long position compared to the no forecast case. This is shown in the total strength figure which is 3.6, much bigger than 2.05 in no forecast case. Overall FAVAR points to a stronger net-long than short position. Due to this overall long preference, there are more net long positions and when the 2020 oil price collapse happened, this strategy made a loss of 5% instead of gains like the no forecast scenario. The overall performance is still positive which is similar to the no forecast scenario as there are more lumps of positive gains scattered throughout the years rather than the two traders that happened with the no forecast case.

Forecast with GVAR

The final case was made with GVAR forecasts. The GVAR performed the worse during the single position backtest. However, the performance here is much stronger than no forecast and FAVAR as the final return was a staggering 70.1% or £45,961 in just 3.5 years. While the forecast accuracy is on par with the other two cases, the main difference is in the calculated buy/sell strength.

Figure 14: Performance FAVAR

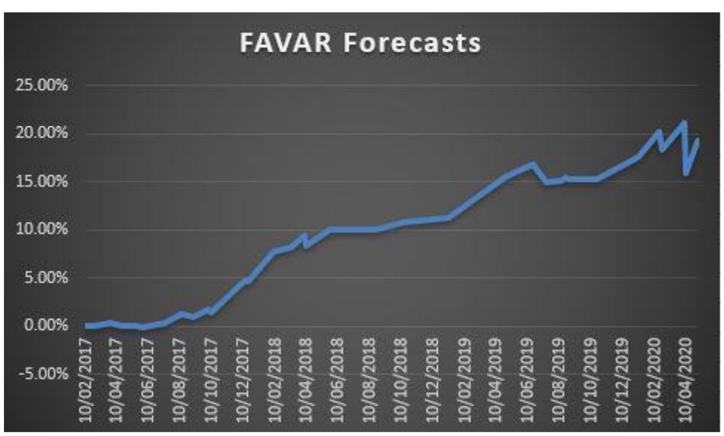


Figure 15: Performance table

Capital	£50,000		
Annual percentage volatility target	20.0%		
Annual cash volatility target	£10,000		
P&L at end of period	£10,300	Average	0.09
Cummulative return %	19.30%	Higest Buy	2.93
Average return per trade%	0.5%	Lowest Sell	-1.44
Highest return %	4.3%	Total strength	3.60
Lowest return %	-5.2%	No. of Buy	20
Average position	0.09	No. of Sell	20
Maximum position	546.00	Sharpe ratio	0.30
Minimum position	-15.00		

Figure 16: Performance return

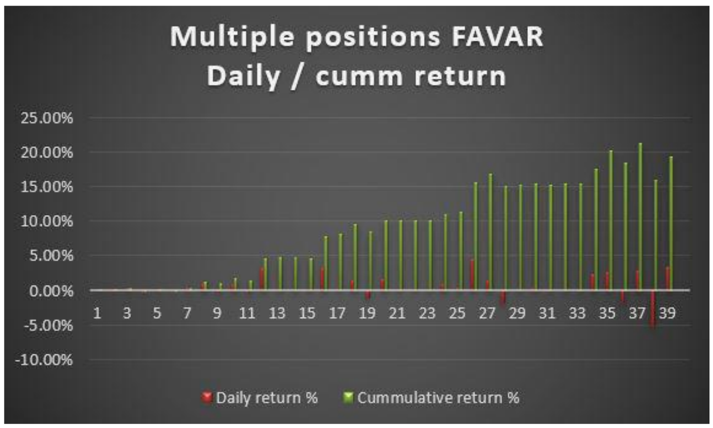
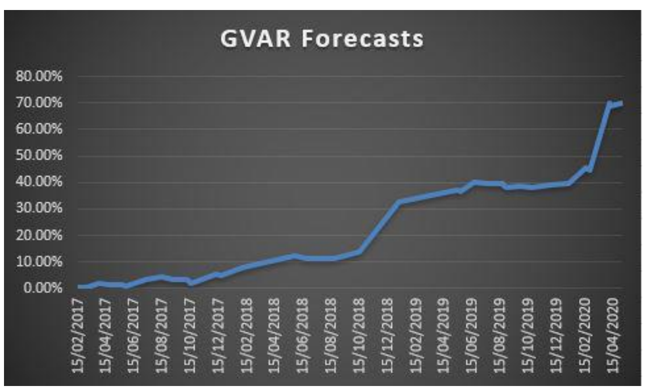


Figure 17: Performance GVAR



In the test period, the GVAR case already reached 25% for the cumulative return at trade 13 while no forecast was negative and FAVAR had 5% only. This early profit boosted the capital in the early period from the 13th trade which saw the capital grow much bigger over time. This early increase in trading capital has enormously increased the daily volatility which in turn magnified the position that it is allowed to take. This is reflected in the metrics where the biggest long position was over 580 contracts and the short was over -163 contracts which are much bigger than the other two cases. This early profit-taking has compounded the overall growth in the end. While the overall accuracy is on par with the other two but the increased trade size and therefore the increased profit has led this case to be much bigger in the long run. For example, it made the right calls and had very successful hit rates since 2018 calling in all correct directions. The highest return was 25% while the biggest loss was only -1.4%. As a result, the Sharpe ratio is a better figure of 0.35.

Performance Comparison and Conclusion

This exercise has shown that three different cases have been very different. The performance difference is staggering from a single position to backtest. From the comparison graph, we can see that the no forecast / perfect information has the worst performance. While the FAVAR had a lower result, in the end, it had been constantly above the no forecast case until the end when it was the only case that made a loss. While forecast accuracy was the most important element in the single position backtest, its relevance is less here. This shows that while it is important to have an accurate forecast model, it is also equally if not even more important to have set the risk and reward correctly in the beginning as it affects the outcome enormously. While it may seem paradoxical to see that no forecast / perfect information would be the best outcome, it is not an unexpected outcome. Single position trading heavily depends on the forecast outcome but multiple positions trading which is much closer to actual practice is itself an art as well as science. The elements from portfolio management have also been proven to be crucial for returns.

This paper has shown the mechanics of backtesting a strategy in two scenarios and with three different cases, therefore 6 different tests in the end. It has answered the questions that were asked at the beginning of how we can know whether a strategy works. And of course the usefulness of forecasts in the context of portfolio management. The single position test has been proven to be quite expected but not the case with the multiple positions as it is complicated by extra elements from portfolio management.

Figure 18 Performance - GVAR

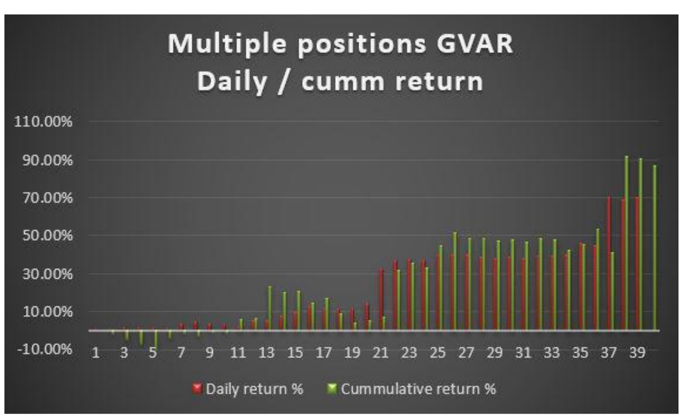
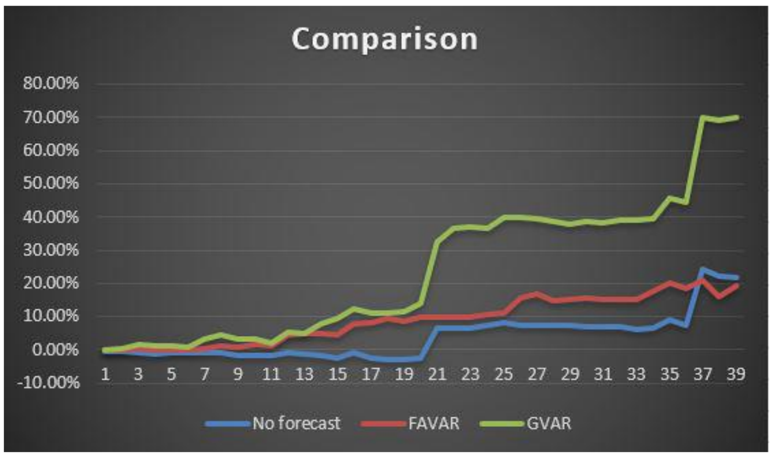


Figure 19 – Performance table

Capital	£50,000		
Annual percentage volatility target	20.0%		
Annual cash volatility target	£10,000		
P&L at end of period	£45,961	Average strength	0.01
Cummulative return %	70.10%	Higest Buy	2.66
Average return per trade%	1.8%	Lowest Sell	-1.26
Highest return %	25.4%	Total strength	0.47
Lowest return %	-1.4%	No. of Buy	20
Average position	0.01	No. of Sell	20
Maximum position	580.00	Sharpe ratio	0.35
Minimum position	-163.00		

Figure 20 Performance comparison



CONFLICT OF INTERESTS

The author declares that there is no conflict of interests.

REFERENCES

[1] J. Bai, S. Ng, Determining the number of factors in approximate factor models, Econometrica, 70 (2002), 191-221.

[2] B.S. Bernanke, J. Boivin, P. Eliasz, Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach, Quart. J. Econ. 120 (2005), 387-422.

[3] A. Chudik, M.H. Pesaran, Theory and practice of GVAR modelling, J. Econ. Surveys, 30 (2016), 165-197.

[4] S. Dees, F.D. Mauro, M.H. Pesaran, et al. Exploring the international linkages of the euro area: a global VAR analysis, J. Appl. Econ. 22 (2007), 1-38.

[5] S. Dees, M. Hashem Pesaran, L. Vanessa Smith, et al. Constructing multi-country rational expectations models, Oxford Bull. Econ. Stat. 76 (2014), 812-840.

[6] F. Di Mauro, M.H. Pesaran, eds., The GVAR handbook: Structure and applications of a macro model of the global economy for policy analysis. OUP Oxford, 2013.

[7] C.W. Granger, Y. Jeon, Evaluation of global models, Econ. Model. 24 (2007), 980-989.

[8] C.W. Granger, J.L. Lin, Causality in the long run, Econ. Theory, 11 (1995), 530-536.

[9] I. Harbo, S. Johansen, B. Nielsen, et al. Asymptotic inference on cointegrating rank in partial systems, J. Bus. Econ. Stat. 16 (1998), 388-399.

[10] C.R. Harvey, Y. Liu, Evaluating trading strategies, J. Portfolio Manage. 40 (2014), 108-118.

[11] S. Johansen, Statistical analysis of cointegration vectors, J. Econ. Dyn. Control. 12 (1988), 231-254.

[12] L. Kilian, H. Lütkepohl, Structural vector autoregressive analysis, Cambridge University Press, 2017.

[13] L.R. Klein, A. Welfe, W. Welfe, Principles of macroeconometric modeling (Vol. 36). North Holland, 1999.

[14] C. Kwok, Estimating structural shocks with the GVAR-DSGE model: Pre-and post-pandemic, Mathematics, 10 (2022), 1773.

[15] J. Kwok, Macroeconometric models for portfolio management, Vernon Press, 2021.

[16] H.H. Pesaran, Y. Shin, Generalized impulse response analysis in linear multivariate models, Econ. Lett. 58 (1998), 17-29.

[17] M.H. Pesaran, Y. Shin, R.J. Smith, Bounds testing approaches to the analysis of level relationships, J. Appl. Econ. 16 (2001), 289-326.

[18] M.H. Pesaran, T. Schuermann, S.M. Weiner, Modeling regional interdependencies using a global error-correcting macroeconometric model, J. Bus. Econ. Stat. 22 (2004), 129-162.

[19] M.H. Pesaran, T. Schuermann, B.J. Treutler, et al. Macroeconomic dynamics and credit risk: a global perspective, J. Money Credit Bank. 38 (2006), 1211-1261.

[20] J.H. Stock, M.W. Watson, Macroeconomic forecasting using diffusion indexes, J. Bus. Econ. Stat. 20 (2002), 147-162.

[21] V. Smith, A. Galesi, GVAR toolbox 2.0–User guide, https://sites.Google.com/site/gvarmodelling/home, 2014.

[22] W. Welfe, Macroeconometric models (Vol. 47). Springer Science & Business Media, 2013.