Jonathan Vincent Tandaw, Mujahid Rahman, Simon Zhang, Catherine Jin.

Dr. Dror Kennet

Financial Econometrics

5/12/2023

Insights on the US and UK Stock Markets

**Executive Summary**

We analyze the US S&P500 and the British FTSE along with respective volatility indices, the VIX and the VFTSE respectively, to see if these series can predict a financial crisis. We chose historical data prior and during the 2008 Global Financial Crisis to see if modeling data prior to 2008 can predict the oncoming crisis. We find that GARCH and ARIMA modeling are inadequate for the task

**Methodology**

We use the quantmod package found in R to download historical financial daily data from FRED and Yahoo Finance (Ryan et al). We analyze the US S&P500 and the British FTSE along with their respective volatility indices, the VIX (Chicago Board Options Exchange) and the VFTSE (Yahoo Finance) respectively. All series are split the same way, with January 1st, 2007 to September 14th, 2008 as the training data and September 14th, 2008 to January 1st, 2009 as our testing data. Then, we ensure our data is tidy and clean by doing some preprocessing, which includes removing invalid NaNs and ensuring that the number of observations are consistent throughout all series, as well as accounting for non-trading days and holidays.

Once data is cleaned, we employ the use of the auto.arima() feature found within the “forecast” package in R to assist in model selection (Ryan et al). This function is powered by the Hyndman-Khandakar algorithm (2008), which combines unit root tests, minimization of information criteria and maximum likelihood estimations, to find the best-suiting model (“How Does Auto.Arima() Work?”). Of course, we will also employ some visual analysis by referring to the autocorrelation function and partial autocorrelation functions of the residuals of our models in addition the auto.arima() for “common-sense” or intuitive model building purposes. These series undergo the same stationarity and validity checks, where applicable. For example, the S&P500 and the FTSE are indices with open and closing prices, and so we take the difference between these prices to generate daily returns, which effectively stationarizes these series.

Verifying the credibility of our models is quite important, and so we employ several model checking steps. We eliminate variables statistically nonsignificant at the 5% level i.e. with t-ratio less than 1.645 and rerun models with the remaining variables. For GARCH and ARMA+GARCH models discussed, we employ the Box-Ljung test on the standardized residuals and squared standardized residuals of these models. These standardized residuals are the residual at time *t* divided by the square root of the conditional variance at time *t*, which in other words, is the estimate for the innovations that appear in these models. We will also consult visual tests like QQ-plots against Gaussian and Studentized distributions for a final peace-of-mind visual check.

As for forecasting, we employ the root mean square error as our forecast accuracy metric. We forecast using selected models fed with training data and compare those prediction results with the testing data that has been left out. Namely, we forecast using the model on training data and compare how close the values are to the unseen test data using root mean square error. This will be our final-say metric to determine which model fits the series best, for it is the most tangible metric we can use.

Lastly, to compare the relationship between the S&P500 and the FTSE to their respective volatility indices, we employ covariance matrix analyses and impulse response function analyses. The former tells us how the values of one series correlate with the other, namely the direction of a linear relationship between the two time series. Meanwhile, the latter tells us what would happen to one of the series and its lags when a standard deviation innovation of the other series shocks the given series. Lastly, we compare the trends found in each country together to compare between

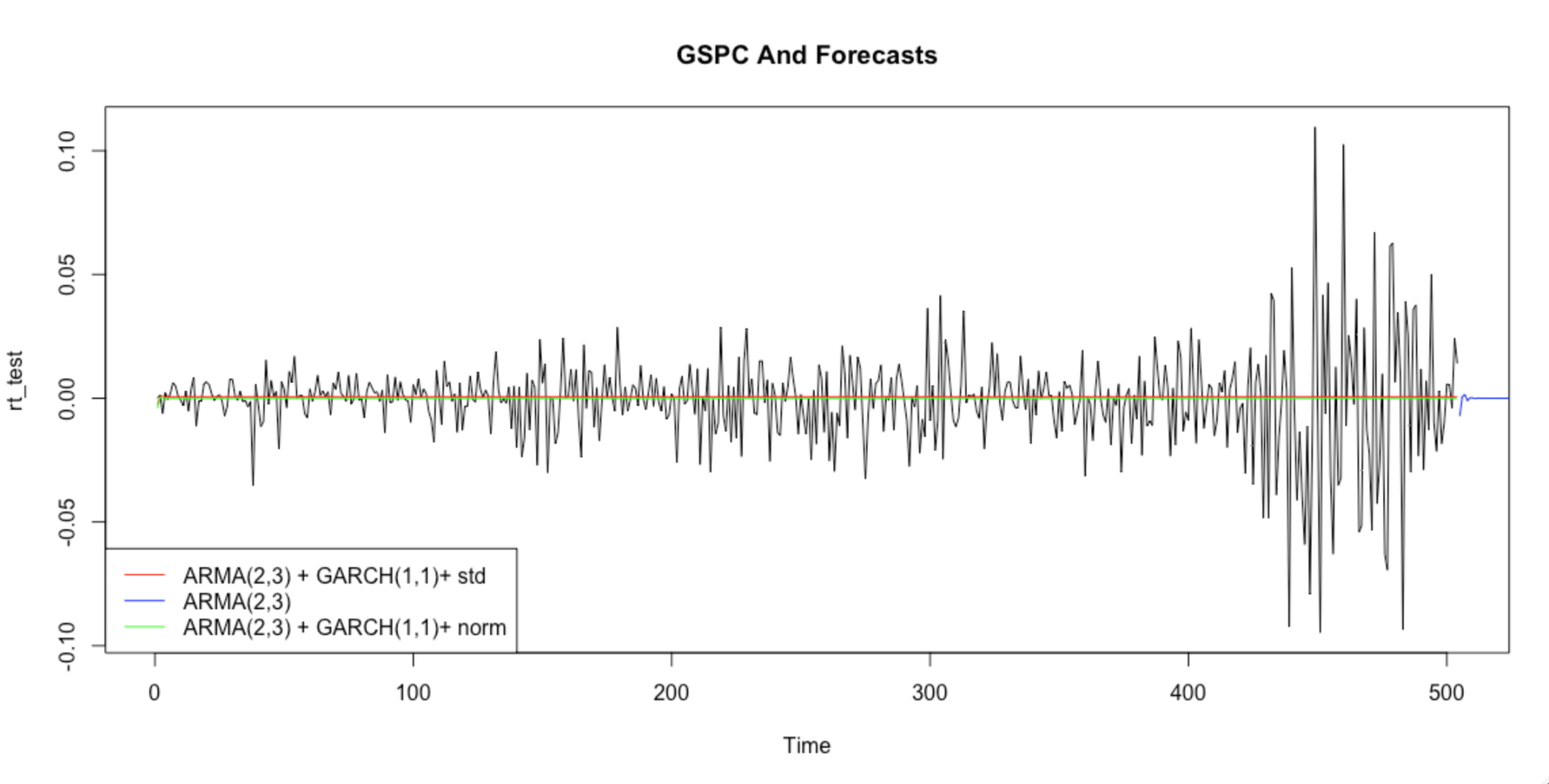
**Results**:

We find that an ARIMA(2,0,1) is the best model for the S&P500, using the auto.arima() function described previously. None of the autoregressive or moving average lags were nonsignificant, either. As a result, we will incorporate all of the lags in this model. The AIC of this model is -2558.5 and the RMSE of this model is 0.01965689. However, the ACF and PACF plots still reveal unaccounted seasonality in the residuals of the model. Perhaps some time trend or data transformation could assuage this concern. Yet, we are more concerned with the possibility of time-varying volatility that plagues many financial time series. Thus, we explore GARCH models instead.

In testing for ARCH effects using Engle’s ARCH test, we find that the p-values for the residuals and squared residuals confidently reject the null of homoskedasticity. Subsequently, we fitted an ARMA(2,1)+GARCH(1,1) model selected for the lowest information criteria available for this model, assuming normal innovations. The model output revealed relatively low information criteria, with AIC=-5.686098 and BIC of -5.627451. We note that ARIMA and ARMA+GARCH models can be compared with AICs because specifying a GARCH volatility component over an ARMA model - that is, doing a joint estimation - does not change the dependent variable, and their loglikelihood remains comparable. Upon inspecting the QQ-plot, we observed a reasonably straight line, although with some deviation in the tails, suggesting that the model may not be perfect but still reasonable. The Q-test results showed p-values of 0.3725 for the standardized residuals and 0.0578 for the squared standardized residuals (with ten lags). However, we note that the p-value for the Box-test on the squared standardized residuals are very close to nonsignificance at the five percent level, meaning more testing and model comparisons are necessary before placing our faith in this model.

Consequently, we reran the previous ARMA(2,1)+GARCH(1,1) model but relaxed the assumption of normal innovations and opted for a Studentized innovation model instead. After all, rarely will we find true Gaussian distributions in the real world. This new model provided higher p-values for the Q-test at the 5% level, with p-values of 0.1776 for the standardized residuals and 0.4893 for the squared standardized residuals, thus firmly rejecting any more residuals still unaccounted by the model. Additionally, the student-t distribution yielded more significant coefficients. The student-t innovation fit produced an AIC of -5.7414 and BIC of -5.6743, which importantly decreased from the previous model with normal innovations, providing more evidence that this model is a better fit. When backtested on test data, the root mean square error of this model is surprisingly similar to that of the ARMA model, at 0.01965. Thus, for our final comparisons further in the paper, we will use this ARMA(2,1)+GARCH(1,1) model with Studentized innovations as the model for the S&P500.

We put the forecast for the S&P500 below:

 We will now move on the modeling of CBOE VIX to see if this series can help predict financial crises.The VIX is an index created by the Chicago Board Options Exchange to measure market expectations for S&P500 price changes in the near future, using prices of S&P500 index option pricing. We will first approach modeling this series with standard ARIMA (Box-Jenkins) methodology, and then attempt GARCH modeling. We will transform this data by taking the log difference of the series for stationarity purposes.

First, we will model using ARIMA methods. The ACF and PACF plots feature large spikes at the tenth lag on both ACF and PACF plots, so we will try an ARIMA(10,1,10) model, however clunky and unparsimonious the model may be. But, as Hyndman and Athanasopoulos mentions, it is hard to use correlograms if a model exhibits ARIMA characteristics where p and q are both non-zero.5 Thus, in contrast, we will also supply results for an ARIMA(3,1,2) model found by the auto.arima() function within R.

The ARIMA(10,1,10) model has an AIC -1022.964 of and root mean square error of 0.10837 when backtested against a test set portion of the data. We note no lingering seasonality left to be modeled in the ACF and PACF plots of this model’s residuals. However, upon employing model checking methods, we find a problem. This model has characteristic roots close to unity, which may have contributed to its parameter instability when non-significant AR and MA lags were taken out. Thus, this linear algebra roadblock prevents successful model checking on this unparsimonious model. In fact, out of all models in this section, only this model failed this model checking step. Regardless, for the sake of comparison, we will continue to use this model as a comparison to other models, especially because the model provides a low root mean square error.

By contrast, the ARIMA(3,1,2) model has an AIC of -1023.797 and an RMSE of 0.10686. Thus, not only is this model more parsimonious, but it uses slightly more information and is less erroneous when backtested using the same test set as the ARIMA(10,1,10) model. Furthermore, none of its lags were nonsignificant at the 5% level. Unfortunately, this smaller model does not capture all the seasonality within the residuals, as the spikes at the tenth lags in both the ACF and PACF plots remain. This is its main weakness, especially when compared to the larger ARIMA(10,1,10) model. However, the spike is relatively small and at a much higher order, so could still stand a chance at forecasting future crises.

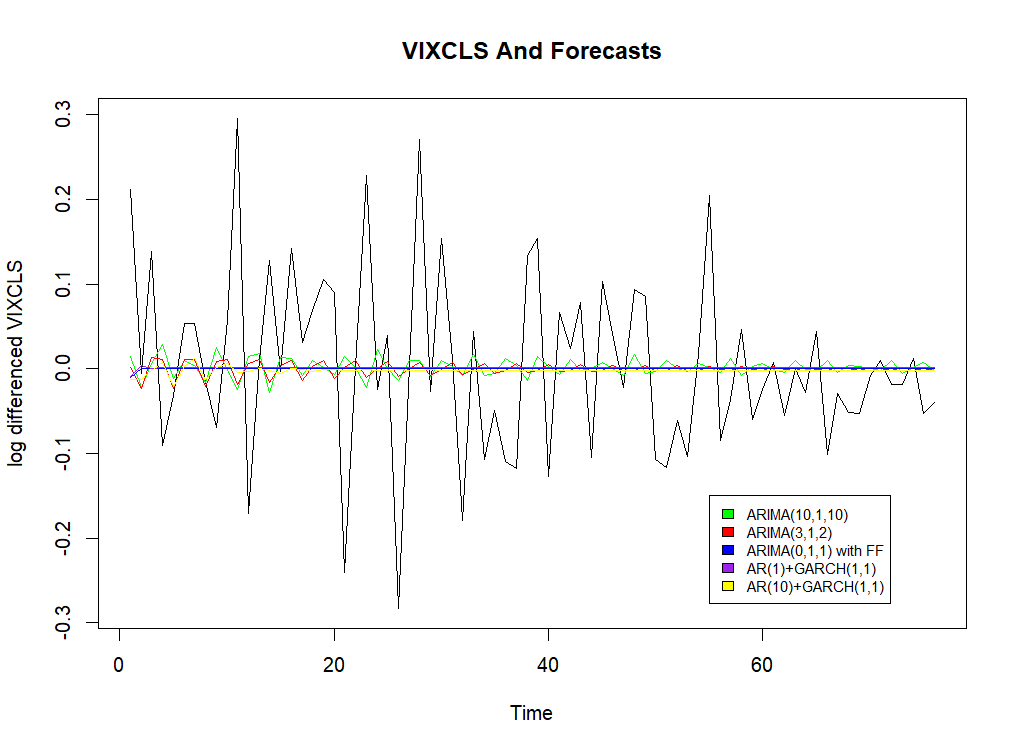
And yet, perhaps introducing some exogenous variables could help remove these spikes. Therefore, we refit the auto.arima() function, to find a better model when we include the effective federal funds rate or money supply (M2) level in the economy, or both. After all, these exogenous variables, or covariates, are quite important in the financial world without being necessarily causally related to the VIXCLS: The federal funds effective rate governs the interest rate at which banks can borrow and trade their cash reserves at the Fed amongst one another; while M2 governs the amount of cash outside banks are available for people to use in their transactions. Since these series are reported monthly, we disaggregate the data down to the higher-frequency daily data format.

After testing ARIMA models with money supply and federal funds effective rate, as well as both simultaneously, we find that the best performing model using these exogenous variables is the very parsimonious ARIMA(0,1,1) with only the federal funds rate as the exogenous variable. This model has an AIC of -1018.668 with an RMSE of 0.10678. Unfortunately, even this model cannot assuage the spike at the tenth lag.

As with before, we will also explore GARCH modeling for the VIX series, because financial time series often do have time-varying volatility. We will first present two models here which are good starting points. From the PACF of the VIX series, an AR(1)+GARCH(1,1) as well as an AR(10)+GARCH(1,1) makes intuitive sense, as they represent large spikes in excess of the 95% bands. However, we must note that we were unable to reject the null that there are no ARCH effects in the sample using Engle’s ARCH test. However, as the p-value sits at around 0.08, some researchers may be willing to entertain the idea of using these models. In any case, we will present them here for completeness.

Overall, both AR(1)+GARCH(1,1) and AR(10)+GARCH(1,1) models with studentized innovations exhibit lower AIC scores than their pure ARIMA counterparts. We note that these two families of models can be compared using AIC as specifying a GARCH volatility component over an ARMA model - that is, doing a joint estimation - does not change the dependent variable, and their loglikelihood remains comparable. These models have AICs of -2.387 and -2.410 respectively, which are much lower than that of the ARIMA models. However, the root mean square error of these GARCH models when backtested on the same test set as the ARIMA models are lower than their ARIMA counterparts, at 0.1068 and 0.1069 respectively. This shows that acknowledging real-world disturbances within our models can provide better predictive power. Next, we analyze their standardized residuals.

The AR(10)+GARCH(1,1) model, while clunkier, eliminates the tenth-lag spike. The very slight breach of the Bartlett bands for the PACF at lag 21 also is unlikely to cause problems given that it barely exceeds the 95% bands, and is at a very high lag. The same success cannot be said of the AR(1)+GARCH(1,1) model, which exhibits substantial seasonality yet to be modeled.

Further tests on the residuals of both models were done with Box-Ljung Q Tests of both the series and the square of the residuals. The non-squared residuals of the AR(1)+GARCH(1,1) residuals provided a p-value below 0.05, indicating that the residuals of this series still showed nonzero autocorrelation and thus is suboptimal. Its squared residuals, meanwhile, provided a p-value of 0.77, providing evidence that there remains no more unaccounted autocorrelation. The level and squared residuals of the AR(10)+GARCH(1,1) model has p-values nearing 1, similarly failing to reject the null of no autocorrelation remaining. We enclose the forecasts here:

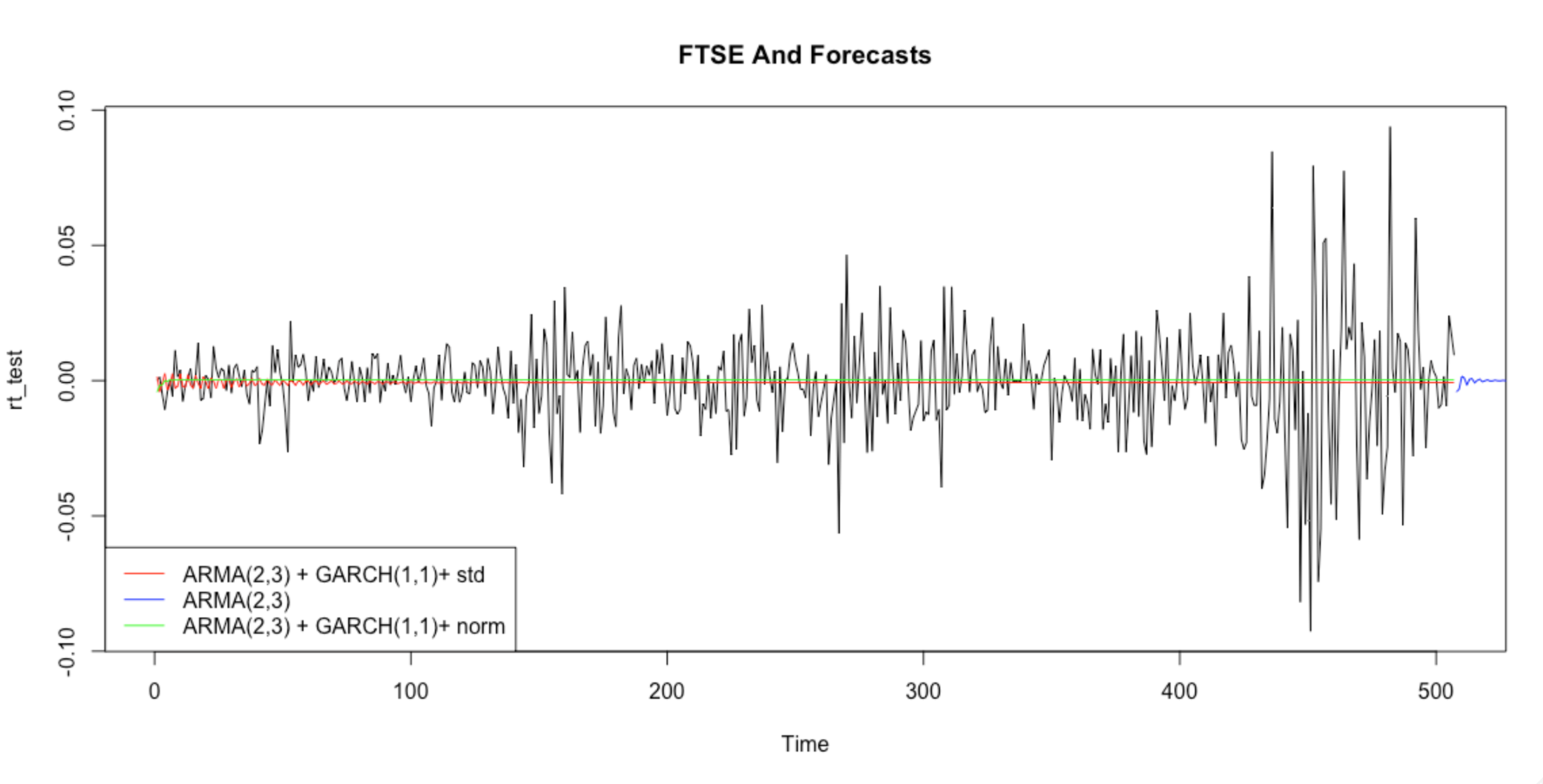
Next, we will now take a look at the UK FTSE series. We found an ARIMA(2,0,3) that best models the logged daily returns of this series. Note that daily returns are inherently and widely considered to be a first-differenced series, as returns are constructed as a difference between opening and closing prices, and any residual seasonality would have been exploited by the market for asset return gains. Checking our model for nonsignificant lags, we find that this specification does not exhibit any lags with t-ratio greater than 1.645, and thus we will use all the

lags of this model. The residuals of this specification do not exhibit further seasonality. We find that this model has a very low AIC which is -2671, and its forecast root mean square error when compared to the excluded test data is 0.01840, indicating good performance.

Comparing to US S&P500, we choose ARIMA(2,0,3) as the best-fit ARMA model for our dataset, with t-ratios of all variables greater than 1.645. As a result, we do not need to remove any insignificant coefficient estimates based on the t-ratio with an absolute value of 1.645. After performing model checking, we conclude that the model is adequate. Next, we attempt to determine if the series exhibits GARCH effects. The p-value of the series is practically 0, indicating that the time series exhibits serial correlation. The same can be said of this model’s standardized residuals and squared standardized residuals, suggesting that the time series is dependent and has a strong GARCH effect.

We fit an ARMA(2,3)+GARCH(1,1) model with low information criteria (AIC and BIC), assuming normal innovation. When inspecting the QQ-plot, we observe a reasonably straight line with some deviation in the tails, suggesting that the model may not be perfect, but still acceptable. Additionally, the p-values of Q-test results are 0.3725 for the standardized residuals and 0.0578 for the squared standardized residuals, indicating that the model is adequate.

Then we consider a Student-t innovation model, ARMA(2,1)+GARCH(1,1), which shows significant coefficients except for the intercept term for the volatility equation of the ARMA+GARCH specification, and also has low information criteria. The QQ-plot displays a reasonably straight line with some deviation in the tails, suggesting that the model may not be perfect, but still reasonable. Moreover, the p-values of Q-test results are 0.1776 for the standardized residuals and 0.5178 for the squared standardized residuals, indicating that the model is also adequate. After comparing the two models, we determine that the Student-t innovation model, ARMA(2,1)+GARCH(1,1), is the best fit due to its significant coefficients and similar AIC and BIC values compared to the ARMA(2,3)+GARCH(1,1) model. The root mean square error of this ARMA+GARCH model when backtested on test data is surprisingly very similar to just the ARMA model at 0.01839.

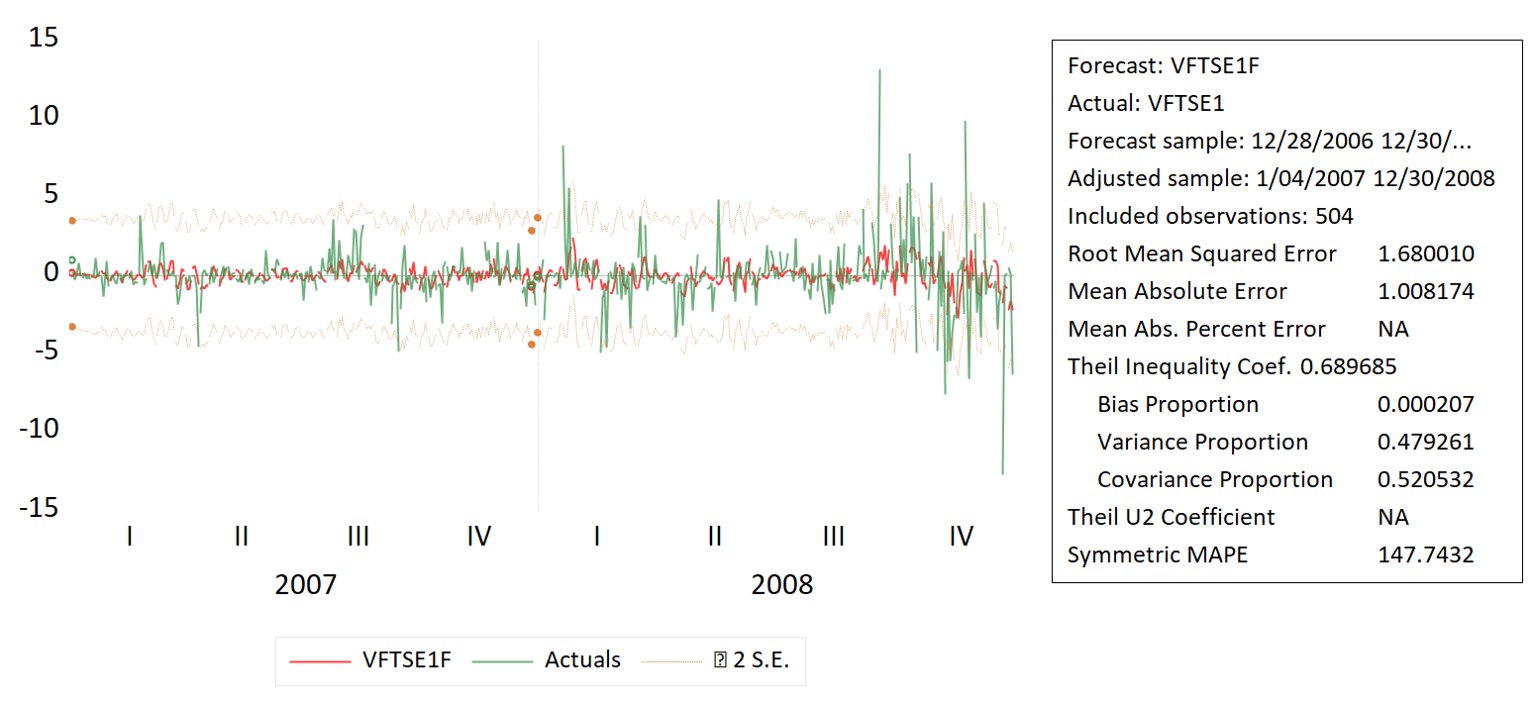
We include the forecasts below: 

For determining the UK market volatility, we choose FTSE 100 Volatility Index “VFTSE” as analysis sets. VFTSE is a real-time index that measures the expected volatility of the FTSE 100 index over the next 30 days, calculated based on the traded prices of options contracts on the FTSE 100.

Based on the results of the auto.arima() function, the identified model for VFTSE is ARMA(4, 1, 5), noting that we take first-differences to stationarize the series. We find that none of the autoregressive or moving average lags were nonsignificant at the 5% level. Examining the ACF and PACF plots of this series indicates that there exist no significant amount of unaccounted seasonality in the residuals of this model. The AIC of this model is 1986.55 and the BIC is 1986.99.

Running Engle’s Arch Test using the archTest() function in R, we test for the presence of ARCH effects in a time series data, and find that there is strong evidence for ARCH effects in the data, with a p-value of nearly zero. The above ARCH effect shows that the conditional variance is dependent on past values. Therefore, the GARCH model can be considered appropriate to estimate the parameters of the variance equation.

The combination of ARMA(4,5), which comes from auto.arima() optimization, and Garch(1,2) has the smallest AIC and BIC among the models estimated. By plotting the QQ Plot, we can visually tell the model has a not perfect but very reasonable estimate on the first difference of VFTSE index. Furthermore, the Box-Ljung Test of the standardized residuals of this model has a p-value of 0.8978 and the squared standardized residuals of this model has a p-value of 0.7269. The AIC of this model is 2.9551 and the BIC is 3.7324. Overall, the GARCH model has a better fit to the first difference of VFTSE. We include the forecasts of this series below:



An analysis of the covariance matrix between the S&P500 returns and the VIXCLS explains how these two time series are related to one another. The covariance between the S&P500 and its own first and second lag is moderately negative, but becomes moderate positive between its third lag, suggesting some kind of lag or rebound as some kind of market pushback. The covariance between VIXCLS and its own lags, on the other hand, show a linearly decreasing negative trend the greater is the lag. Moreover, the covariance between the S&P500 and the VIXCLS is increasingly negative until its third lag, suggesting that a larger daily return on the S&P500 corresponds to lower and lower perceived volatility in the market. A similar story appears when employing impulse response functions analysis of and between these two series. A standard deviation innovation of RT unto itself yields a negative covariance until the third lag, after which it will instead rebound to a moderately positive value, meanwhile a standard deviation innovation of VIX unto itself yields a constant but slight downward trend in VIX values. And lastly, VIX innovations only negligibly affect S&P500 returns while S&P500 innovations create a negative but slightly increasing values of VIX in response

To analyze the relationship between the return of FTSE and VFTSE, we conducted a covariance matrix analysis. The covariance matrix describes the covariance relationship between a vector variable, in which the diagonal elements represent the variance of each variable, and the non-diagonal elements represent the covariance between two variables.

Because the larger the value of covariance, the higher the degree of correlation between the two variables. A positive value means the trend of variables is similar, while a negative value means the trend of variables is opposite. From the covariance matrix, we can conclude that basically returns will have a similar trend (positive covariance) in response to the lagged VFTSE. Furthermore, most of the lagged returns have opposite trend (negative covariance) against concurrent and preceding VFTSE. Such results can be interpreted in plain language as the VFTSE tends to rise for a few days after stock market return drops, which echoes with what usually happens in reality.

Impulse response analysis reflects the dynamic influence of other variables when one variable is affected by "exogenous shock". We plot the impulse response based on the dynamic changes of VIX and FTSE index return over a period of time after this shock. When there is a one standard deviation shock on index return, the index return itself will absorb the majority at period 1 with a sharp rise followed by two consecutive mild declines. Finally the impact stabilizes at around period 8. However, the index return shock on VIX is consistent and negative. On the other hand, when there is a one standard deviation shock on VIX, the index return’s reaction starts at period 2 and vanishes at period 5 with inconspicuous variation. And VIX itself has a mild but continuous positive shock-absorbing effect that lasts at least 10 periods.

Now we switch from comparing index returns and its volatility within both countries and compare the US trends with the UK trends. Comparing the covariance matrix of U.S. stock index vs. VIX and that of U.K. stock index vs. VFTSE, we find that the pattern observed in the U.K. case does not apply to the U.S. More specifically, in the U.S. case, the covariances between the index returns and VIX values are mostly negative, demonstrating a strong inverse movement pattern between the two. While in the U.K. case, there is a clear boundary of plus or minus characteristic of covariance value at concurrent periods, and the pattern is more easily explained by market norm. In terms of impulse analysis, we observed quite similar reactions in both markets when there is an external shock in either stock market return or VIX index.

As for whether these models separately or in tandem were able to predict the 2008 Global Financial Crisis, we can strongly say that these series and the models we fit were unable to do so. These models and the approach is too simple to be able to detect such a strong shock against the economies of the world, and that of the United States, the effects of which are arguably still felt today. The data may also not be adequate enough to train our models in prediction. If we had included more shocks in the training data, then perhaps the model could stand a better change. For example, a training set with the Great Depression, the Dotcom bubble, the 2008 Global Financial Crisis, and the Great Recession could stand a better chance at predicting the economic downturn caused by the Covid-19 pandemic. Or, from a smaller scale, perhaps fitting our ARIMA models with deterministic trends such as time breaks (using a binary dummy variable) indicating when a country was in recession or not could have helped the model. We include the forecasts of all the models presented here in the appendix.

In closing, we must include a small discussion on other methods besides traditional (financial) econometrics that could be a better fit for this purpose. For example, advanced machine-learning methods such as neural networks, however maligned their “alchemical” hidden layers are, could provide better forecasts i.e. “out-of-sample predictions” than the traditional models presented here. Another approach, which has been increasingly popular in the financial and econometric world is network analysis, which could be a better tool to not only predict incoming financial downturn by identifying “weak links” within the intertwined global financial community, but also assess to what degree a given network of banks, institutions and organizations are resistant towards economic disaster. Barigozzi and Hallin (2017) analyzed the interconnectivity between stocks in the S&P100 and large firms in the years 2000-2013, which includes the 2008 Global Financial Crisis, a setup with great similarities to this paper. Kennet and Havlin (2015) explained how network analysis can be used by economists to model financial crises.

Works Cited

Barigozzi, Matteo and Marc Hallin. 2016.“ A Network Analysis of the Volatility of High-Dimensional Financial Series.” London School of Economics and Political Science. https://eprints.lse.ac.uk/67456/7/Barigozzi\_Network%20analysis.pdf

Dror Kenett & Shlomo Havlin, 2015. "[Network science: a useful tool in economics and finance](https://ideas.repec.org/a/spr/minsoc/v14y2015i2p155-167.html)," [Mind & Society: Cognitive Studies in Economics and Social Sciences](https://ideas.repec.org/s/spr/minsoc.html), Springer; Fondazione Rosselli, vol. 14(2), pages 155-167, November.

“FTSE100 VOLAT.INDX (VFTSE.AS).” Yahoo Finance. https://finance.yahoo.com/quote/VFTSE.AS/

Hyndman, Rob J. and George Athanasopoulos. 2018. “How does auto.arima() work?” Online textbook. https://otexts.com/fpp2/arima-r.html

Hyndman, Rob J. and George Athanasopoulos. 2018. “Non-seasonal ARIMA models.” Online textbook. https://otexts.com/fpp2/non-seasonal-arima.html

“Making Sense of the VIX Index: An Indicator of Expected Market Volatility.” Chicago Board Options Exchange. https://www.cboe.com/tradable\_products/vix/

Ryan, et al. “Package Quantmod”. Comprehensive R Archive Network. https://cran.r-project.org/web/packages/quantmod/quantmod.pdf

Appendix

