UNIVERSITY OF AGDER

A comparison of volatility prediction between ARIMA-GARCH and VAR models

Yngve Aslaksen Gram & Ørjan Dovran Thomassen

Supervisor

Valeriy I. Zakamulin

University of Agder, 2015

School of Business and Law

Department of Economics and Finance

Abstract

In this thesis the authors use ARIMA-GARCH and VAR to predict future volatility of 6 macroeconomic variables from the US. The data is monthly and spans the period 1964-2014, where the last 20 years are used as the out-of-sample period. The univariate GARCH models are widely used in volatility estimations in the fields of macroeconomics and finance, and the authors feel that there are better ways of predicting volatility. We find that VAR outperform both GARCH and EGARCH when it comes to predicting future volatility out-of-sample, which is consistent with previous research. The naive model is found to outperform both GARCH and EGARCH, while there is no conclusive answer for which GARCH model is superior.

Contents

1 Introduction

Measuring and predicting volatility is an imperative concept in vast areas of fields spanning from finance and economics to medicine and politics. In the field of finance, volatility is used as a measure of market risk, and predicting volatility is an important element for empirical asset pricing, portfolio allocation and risk management. As for macroeconomics, volatility forecasting can be useful for central banks, in order to conduct proper monetary policy decisions. There is also a possibility to trade volatility products, such as the CBOE Volatility Index (VIX).

Empirical research shows that volatility can be predicted (Poon and Granger,2003), and research within the field of financial economics has tried to explain the causes of stock market volatility and macroeconomic volatility. Studies by Morelli (2002), Sadorsky (2003), Kearney and Daly (1998), Sollis (2005), Engle, Ghysels, and Sohn (2008), Mele (2009), Christiansen, Schmeling, and Schrimpf (2012) have examined the relationship between the stock market and different macroeconomic factors. The results from these papers show that macroeconomic volatility and stock market volatility can forecast each other. There have also been studies on the relationship between macroeconomic factors. Wheelock and Wohar (2009) conducted a survey of the literature concerning with the usefulness of the term spread for predicting changes in economic activity. They revealed that the term spread can forecast recession and output growth. Apergis and Miller (2005) analyzed whether money supply has an asymmetric effect on output volatility or not, implying that conditional variance of industrial production changes more because of positive changes than negative changes in money supply volatility.

Poon and Granger (2003) classified four different models used in volatility forecasting: (1) Historical volatility models, which are measured by standard deviation of historical returns; (2) ARCH class conditional volatility models, which are econometric time series models that estimate conditional variance of returns; (3) Stochastic volatility models; (4) Option-based volatility models that use volatilities, which are implied by option pricing models, like the Black-Scholes model.

Previous research by Christoffersen and Diebold (2000), Marquering and Verbeek (2004), Fleming, Kirby, and Ostdiek (2001), Andersen, Bollerslev, Christoffersen, and Diebold (2006) shows that predicting volatility can be valuable in risk management, portfolio optimization and timing strategies.

The main goals of this thesis are: (1) Create optimized univariate GARCH models based on an information criteria for each variable; (2) Test whether our optimized models are well specified in-sample or not; (3) Study whether the volatility of different macroeconomic variables can be used to predict each other in a VAR model or not; (4) Investigate if there is statistical evidence towards bivariate Granger causality out-of-sample; (5) Compare which of the forecasting models perform best in predicting volatility out-of-sample.

The macroeconomic variables used in this thesis are the term spread, industrial production index, personal consumption expenditures index, money supply, S&P500 and the civilian unemployment rate. The survey carried out by Wheelock and Wohar (2009), and the studies by Ewing and Thompson (2008), Judson and Orphanides (1996), Apergis and Miller (2005), Fallahi, Pourtaghi, and Rodríguez (2012) show that these variables can be vital for a general understanding of the economic state.

This thesis will compare the forecasting performance between: (1) Vector AutoRegressive model (VAR); (2) AutoRegressive Integrated Moving Average-Exponential Generalized AutoRegressive Conditional Heteroskedastic (ARIMA-EGARCH) with a Generalized Error Distribution (GED); (3) Standard GARCH (1,1) with a normal distribution; (4) Naive model. The forecasts generated by these models will be compared during the out-of-sample period.

The first thing to be done was to detrend the variables, as this was necessary to create models to predict volatility. After that, appropriate lag lengths were chosen for the ARIMA-EGARCH model and the GED was chosen to model the distribution. This model was optimized in-sample, during the period of 1964-1994, where the asymmetry term for unemployment rate was set equal to 0, due to our model failing to account for unemployment rate being multimodal. Various tests were conducted to find out how well the model fits the in-sample period.

Once the 6 different univariate models were created, the authors used them to predict out-of-sample volatility from 1994-2014, this was done using a recursive window where all previous information is taken into account by the model. The authors created a VAR model by choosing the lag length of the model, which minimized the Hannan-Quinn Information Criterion (HQIC) in-sample. The variables was then tested for Granger causality, where it was found out that all variables either helped predict another variable or were predicted by other variables. A similar forecasting approach as the one used in the univariate case was then carried out with the VAR model, where we used a recursive window to estimate volatility out-of-sample. A baseline was estimated as a ARIMA(0,0)-GARCH(1,1) model with which to compare the performance of the other models.

We found that the ARIMA-EGARCH models we created seemed well specified in the in-sample period, but performed inferior to the naive model out-of-sample. We found evidence of bivariate Granger causality for multiple variables, although the robustness of the Granger causality was questionable. The VAR model outperforms all of the other models, while the naive model outperformed GARCH and EGARCH. We found no conclusive evidence that either GARCH or EGARCH give useful volatility forecasting with monthly data.

This thesis is organized in the following way. Section 2 presents a review of research literature on the topic of forecasting volatility. Section 3 presents the data that are used in this thesis, and the motivation behind the variables. Section 4 gives a presentation of the theoretical framework associated with forecasting volatility. Section 5 reports the empirical results, and in Section 6 the discussion comparing the results with previous research is presented together with the directions for further research. The conclusion is presented in Section 7, and the appendix can be found in Section 8.

2 Literature Review

There has been conducted several empirical studies about the relationship and predictability of stock market volatility and macroeconomic volatility, and the importance of reliable estimation when predicting volatility. Brownlees, Engle, and Kelly (2011) emphasized that poor estimation of future risk can lead to investors being unnecessarily exposed to market fluctuations. There are different alternatives of predicting volatility, and it is found that the predictive power of standard deviation and conditional volatility of a variable can be improved by including other variables in a forecasting model.

This section will look into how and if the volatility of the different macroeconomic variables are interacting. It will also give the authors of this thesis a blueprint on how to carry out the goals, and will be used to compare our results with.

2.1 Volatility Relationship

Beltratti and Morana (2006) studied the relationship between macroeconomic and stock market volatility, using S&P500 data over the period 1970–2001. They examined the relationship between volatility of S&P500 and macroeconomic factors, such as money supply (M1), industrial production growth rate, inflation rate and Federal funds rate. They found evidence of a bidirectional link between macroeconomic and stock market volatility, but the Granger causality is stronger from macroeconomic to stock market volatility rather than the other way around. Structural breaks processes in stock return volatility were explained by volatilities of interest rates and money growth, which in turn could be explained by discrete changes in monetary policy.

Morelli (2002) investigated if the conditional stock market volatility can be explained by conditional macroeconomic volatility, based on monthly UK data. He calculated monthly returns for the FT All Share Index, and the following macroeconomic variables were used; industrial production, real retail sales, money supply, the German Deutsche mark/pound exchange rate and an inflation rate were used, covering the period from

January 1967 to December 1995. The conditional volatility was estimated using ARCH and GARCH models, and the relationship between the conditional volatilities was examined by a two variable VAR model. The study found evidence of a relationship between stock market and macroeconomic volatility in terms of VAR estimation. The volatility in the macro variables was found not helping to explain the volatility in the stock market.

Sadorsky (2003) studied the macroeconomic determinants of Pacific Stock Exchange Technology 100 Index conditional volatility, using daily and monthly data over the period of July 1986 to December 2000. The empirical results show that each of the conditional volatilities of oil prices, term premium and the consumer price index has an impact on the conditional volatility of technology stock prices. It also shows that daily stock data display more persistence than monthly data when calculating conditional volatility.

Kearney and Daly (1998) examined to which extent the conditional volatility of stock market return are related to the conditional volatility of financial and business cycles. They used a monthly dataset for Australian variables such as stock market returns, interest rates, inflation, money supply, industrial production and current account deficit over the period July 1972 to January 1994. To estimate the conditional volatility, they implemented an ARCH model. The results from this paper show that Australian stock market conditional volatility is directly associated with conditional volatility of inflation and interest rates. While industrial production, current account deficit and money supply are indirectly associated with stock market conditional volatility.

Schwert (1989) did not analyze the causes of stock price volatility, but rather associations between stock volatility and other variables. He used daily returns of the Standard and Poor's (S&P) Composite Index to estimate the standard deviation of monthly stock returns, monthly data for the money supply (monetary base), industrial production and the inflation rate (PPI, not seasonally adjusted), all during the period from late 19th century to 1987. He concluded with five significant empirical results, two of which are of great importance to this thesis. First, by using a VAR model and measuring F-statistics, Schwert found statistical evidence that macroeconomic volatility can be used to predict stock return volatility. On the other hand, there were found more statistically significant evidence for financial volatility helping to predict macroeconomic volatility. Secondly, he found evidence of stock market volatility being related to the economic state, meaning the volatility is higher during recessions than expansions. The explanation behind this is that leverage is increasing during recessions, which is causing an increase in the volatility of leveraged stocks. Among the findings is that in certain sub-periods money growth volatility helps to predict stock market volatility, and also that stock market volatility helps to predict money growth volatility. In the sub-period from 1953-1987, inflation volatility helped to predict stock volatility. The relationship between inflation, money growth volatility and stock volatility is statistically significant. The evidence for industrial production volatility helping to predict stock volatility are not statistically significant, but the evidence of stock market volatility helping to predict industrial production are statistically significant in certain sub-periods.

Sollis (2005) studied the forecasting of SP Composite Index returns and volatility using, inflation rate, monthly change in the three month Treasury bill rate, the monthly change in the 10-year bill rate, growth rate of industrial production, and the growth rate of money supply (M1). Sollis (2005) used tests of forecast encompassing to evaluate one-step-ahead forecasts of S&P500 returns and volatility. He implemented a ARMA(1,0)-GARCH(1,1) model with no macroeconomic variables as a benchmark for forecast comparison. For 1970's, the macroeconomic variables contain information for forecasting stock market return and volatility. As for 1990's, this is not the case. Thus, he concluded that the predictive content of macro variables used for forecasting SP Composite index returns depends on the general macroeconomic environment.

Engle et al. (2008) studied the relationship between stock market volatility and macroeconomic activity using historical data of aggregate stock market volatility as Schwert (1989) did. They studied how much volatility relates to the economy, and how much volatility anticipates the future. They used monthly inflation rate (producer price index) and industrial production growth as macroeconomic variables. They used a mean reverting unit daily GARCH process and a MIDAS (mixed data sampling) polynomial, which allows for extraction of two components of volatility. Due to concerns of structural breaks, the paper considered various sub-samples and formally tests for breaks: pre-WW1, the Great Depression era, post-WW1, and a split to examine the Great Moderation. Among the empirical results, they found that the macroeconomic variables help in terms of long horizon forecasting.

Mele (2009) investigated the predictive power of financial volatility for the economic activity in the US. The financial variables include stock price index and term spread (difference between 10 year government bond yield and 3 month Treasury bill yield), while the macroeconomic variables include seasonally adjusted industrial production index, consumer price index and unemployment rate. He defined volatility, as a moving average of past absolute returns, and found that financial volatility predicts 30% of post war economic activity. Aggregate stock market volatility explains 55% of real growth during the Great Moderation. Financial volatility and macroeconomic volatility explains about 50% of industrial production growth, and that the predicting power of stock market volatility increased during the last 25 years, which included the Great Moderation.

Christiansen et al. (2012) studied whether return volatility on S&P500 is predictable by macroeconomic variables and financial variables. Their study ran from December 1926 to December 2010, using monthly observations. Among their 38 predictive variables were inflation, industrial production, interest rates and different forms of spreads. Part of their study was to use model selection and forecast combinations to investigate whether macroeconomic variables are robust predictors of financial volatility. Their finding shows that macroeconomic variables help to predict financial volatility in an out-of-sample setting.

2.2 Economic Relevance

Practical application of forecasting financial and economic volatility can be justified by asset pricing models, portfolio optimization and diversification, and the use of risk management in business. The systematic risk factor (β) in a simple linear asset pricing model like the Capital Asset Pricing Model (CAPM) is expressed as a fraction of the covariance between the market and an asset, and the variance of the market. The CAPM can be used to compute a firm's cost of equity, and thus, forecasting volatility is important in the context of asset pricing. A more precise future estimate of a firms cost of equity can give better estimates of the value for future projects.

Christoffersen and Diebold (2000) studied how relevant volatility forecasting is for risk management. They implemented US 10 year Treasury Bonds, daily returns of S&P 500, German DAX, UK FTSE, Japanese TPX, and dollar rates for German Mark, British Pound, Japanese Yen and French Franc, with a sample period from January 1973 to May 1997. They developed a model-free procedure for measuring volatility predictability across horizons. The results show that volatility forecasts are helpful in risk management, if volatility varies in a predictable way, but the usefulness diminishes for longer forecasting horizons than 10 to 20 days.

Marquering and Verbeek (2004) analyzed the economic value of predicting index returns and volatility. They estimated out-of-sample forecasts for return and volatility, using monthly data over the period 1970-2001 with simple linear models, which were estimated repeatedly. The economic value is examined with different trading strategies, where the strategies that use volatility timing perform best. Volatility timing is found to increase the Sharpe ratio and utility levels of strategies that are only based on timing of returns, i.e. strategies with both volatility and return timing will exceed strategies with only return timing.

Fleming et al. (2001) analyzed the value of volatility timing for short-horizon asset allocation strategies. They used conditional mean-variance analysis, in order to deal with the research question. Funds were allocated using a mean-variance optimization rule with daily re-balancing across four asset classes: stocks, bonds, cash and gold futures. They evaluated the value of volatility timing by comparing volatility timing strategies with unconditional mean-variance efficient static strategies. Their findings show that the volatility timing strategies outperform the unconditionally efficient strategies that have the same target expected return and volatility.

Andersen et al. (2006) provides a survey of important insights and theoretical development that has emerged from the literature of volatility, and with a focus on forecasting applications. The survey lists a number of studies for application of volatility forecasts. Among them we have inflation uncertainty relationship with labour market variables, exchange rate volatility relationship with US monetary policy, simulation of gas and oil price paths using a GARCH model, and modelling electricity demand uncertainty with weather forecast uncertainty. The different applications have various benefits in areas stretching from politics to medicine. As for finance and economics, measuring and forecasting volatility is an important objective for asset pricing, portfolio allocation, risk management and monetary policy decision making.

3 Data

Data in this thesis are chosen based on the relationship between the variables, and on their economic relevance. A number of studies that examined the relationship between stock market volatility and macroeconomic volatility apply many of the same variables, such as money supply, inflation rate, real output, exchange rates, unemployment rate, term spread and oil prices. Among these, the variables, which usually show a significant relationship with the stock market, are term spread, money growth, inflation rate and industrial production rate. In light of previous studies, this thesis will examine the volatility relationship between the S&P500, term spread, industrial production index, personal consumption expenditures, money supply and civilian unemployment rate. The macroeconomic data in this thesis are obtained from Federal Reserve Economic

Data[1][2][3][4][5][6], while the stock market index, S&P500, is obtained from Yahoo! Finance[7]. This thesis will use monthly announced figures, since it is assumed that unexpected changes in variables lead to revisions in expectation when employing macroeconomic variables(Morelli,2002). With 611 monthly observations, this thesis covers the period from January 1964 to November 2014.

The variable, which represents stocks in this thesis is the S&P500, which is considered to be one of the leading indicators of the US economy. S&P500 is a market value-weighted index, and this thesis use the adjusted close price, thus the variable is adjusted for dividends.

The term spread (SPREAD), is the difference between the 10-year Treasury constant maturity rate and the 3-month Treasury bill rate, which is the secondary market rate. These variables are not seasonally adjusted. Wheelock and Wohar (2009) conducted a survey on the literature concerning the usefulness of the term spread for predicting changes in economic activity in the US and outside of US. The questions that are answered are; if term spread forecasts output growth, and if term spread forecasts recession. Based on the previous research, Wheelock and Wohar (2009) concluded with answering "yes" to these questions. The former research which this survey is based on, finds that the term spread is useful for forecasting at horizons of 6 to 12 months, and also if other variables are included in the forecasting model. Graphs showing the correlation between the term spread and recessions in the US show that every recession happened when the yield curve was inverted (i.e. higher yield on short-term rates than long-term), or when the decline in 10-year Treasury securities was relatively higher than the decline in 3-month Treasury securities.

The US industrial production index (IP) is an economic indicator that measures real

[1] http://research.stlouisfed.org/fred2/series/GS10#
[2] http://research.stlouisfed.org/fred2/series/TB3MS#
[3] http://research.stlouisfed.org/fred2/series/INDPRO#
[4] http://research.stlouisfed.org/fred2/series/UNRATE#
[5] http://research.stlouisfed.org/fred2/series/PCE#
[6] http://research.stlouisfed.org/fred2/series/M2SL#
[7] http://finance.yahoo.com/q/hp?s=GSPC&a=00&b=3&c=1964&d=11&e=1&f=2014&g=m

output, and it is seasonally adjusted. Ewing and Thompson (2008) estimated volatility for industrial production for the US using GARCH and EGARCH models. They found that industrial production volatility is predictable, and overestimates of production lead to a greater increase in volatility than underestimates do. This knowledge of future uncertainty in production can help supply chain managers to improve performance and increase the value of their company.

The US Personal Consumption Index (PCE) is the primary measure of consumer expenditures on goods and services. PCE is a measure of consumer price inflation and it is Federal Reserves (FED) preferable inflation measure and is seasonally adjusted. Judson and Orphanides (1996) presents empirical evidence on benefits of the sensibility of price stability as a monetary policy goal. They measured annual inflation volatility and found that inflation volatility is negatively correlated with income growth, and volatile inflation is associated with low income growth. Thus, forecasting inflation volatility can be a useful tool in monetary policy decision making.

The Philips curve shows the relationship between unemployment rate and inflation rate in an economy. It is an inverse relationship, which states that decreased unemployment rate will commonly correlate with increased rates of inflation. Mullineaux (1980) tested Milton Friedman's hypothesis about the relationship between uncertainty in inflation and unemployment, and tested if increased uncertainty of inflation causes a growth in the unemployment rate and a reduction in industrial production level. In the short-run, this hypothesis was not rejected, and that minimizing uncertainty about future inflation should be implemented by monetary policy makers.

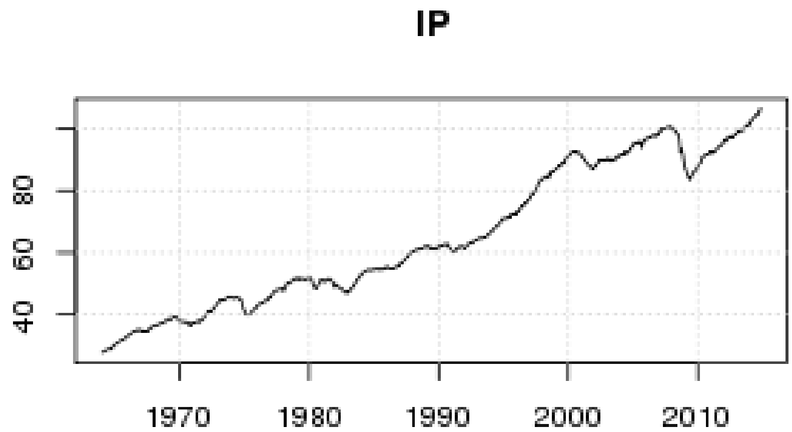
The variable that represents the money supply is M2, and is seasonally adjusted. M2 is defined as the sum of currency held by the public and transaction deposits at depository institutions (M1) plus saving deposits, small-denomination time deposits (issued in amounts of less than 100,000), and retail money market mutual fund shares. Apergis and Miller (2005) used a multivariate GARCH model in order to analyze whether money supply (M1) volatility has an asymmetric effect on output (industrial production) volatility in the US or not. Results imply that the conditional variance of industrial production changes more because of positive changes than negative changes in money supply volatility. Consequently, forecasting money supply volatility has an important effect on the real economy.

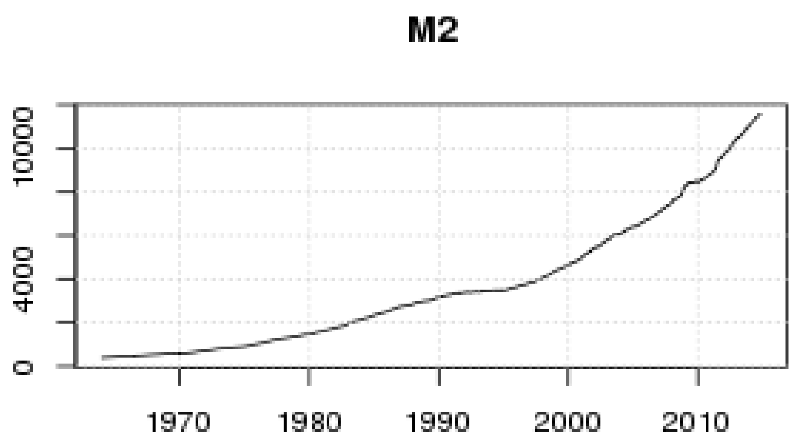
The unemployment rate is given by the US civilian unemployment rate (UEMP), which represents the number of unemployed as a percentage of the labor force, and it is seasonally adjusted. Fallahi et al. (2012) studied the effect of the unemployment rate and its volatility on crime in US. The volatility was estimated with quarterly data using an EGARCH model. Their findings show that the unemployment rate volatility has a negative effect on auto theft, and in the short run, a positive effect on burglary. This shows that forecasting volatility is useful in other disciplines as well.

Figure 1 and 2 displays the raw data for each time series used in this thesis. IP, M2, PCE and SP500 exhibit signs of long term trends, while UEMP and SPREAD display a more cyclical short term behavior. After the dot-com bubble in the early 2000's, SP500 also show short term cyclical behaviour. M2 and PCE exhibit a smoother trend throughout the sample period, with a relatively small decline around the financial crisis of 2008. The overall change in M2 and PCE throughout the sample period is relatively larger than the other variables, and PCE appears to be least affected by economic trends.

Figure 3 and 4 displays the detrended variables, where the long term trends and short term trends of the variables are no longer visible, though some of the variables seem to concentrate around a non-zero constant. Volatility clustering seems to be apparent for the variables, i.e. large changes are followed by large changes. IP and SP500 exhibit signs of negative behavior at times when the different financial crisis and breaks that has occurred, while M2 shows the opposite trend. Figure 5 and 6, which displays the detrended standard deviations of the variables, shows no clear signs of trending behaviour. Even though it is clear that these variables are affected by financial crisis and the general health of the economy.

Figure 1: Time series of raw data





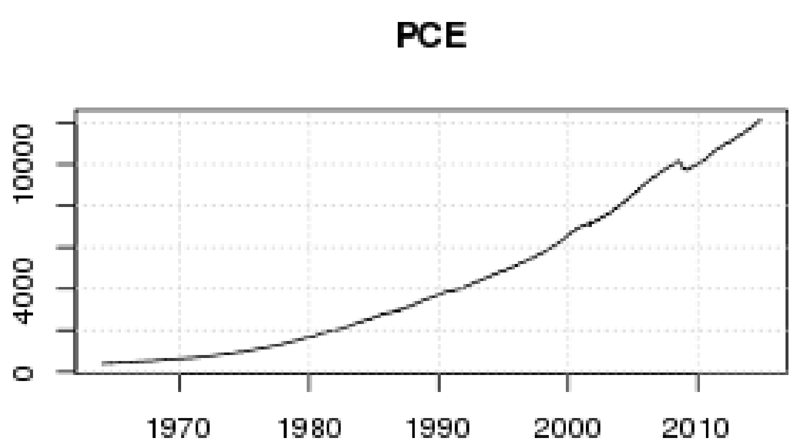


Figure 2: Time series of raw data

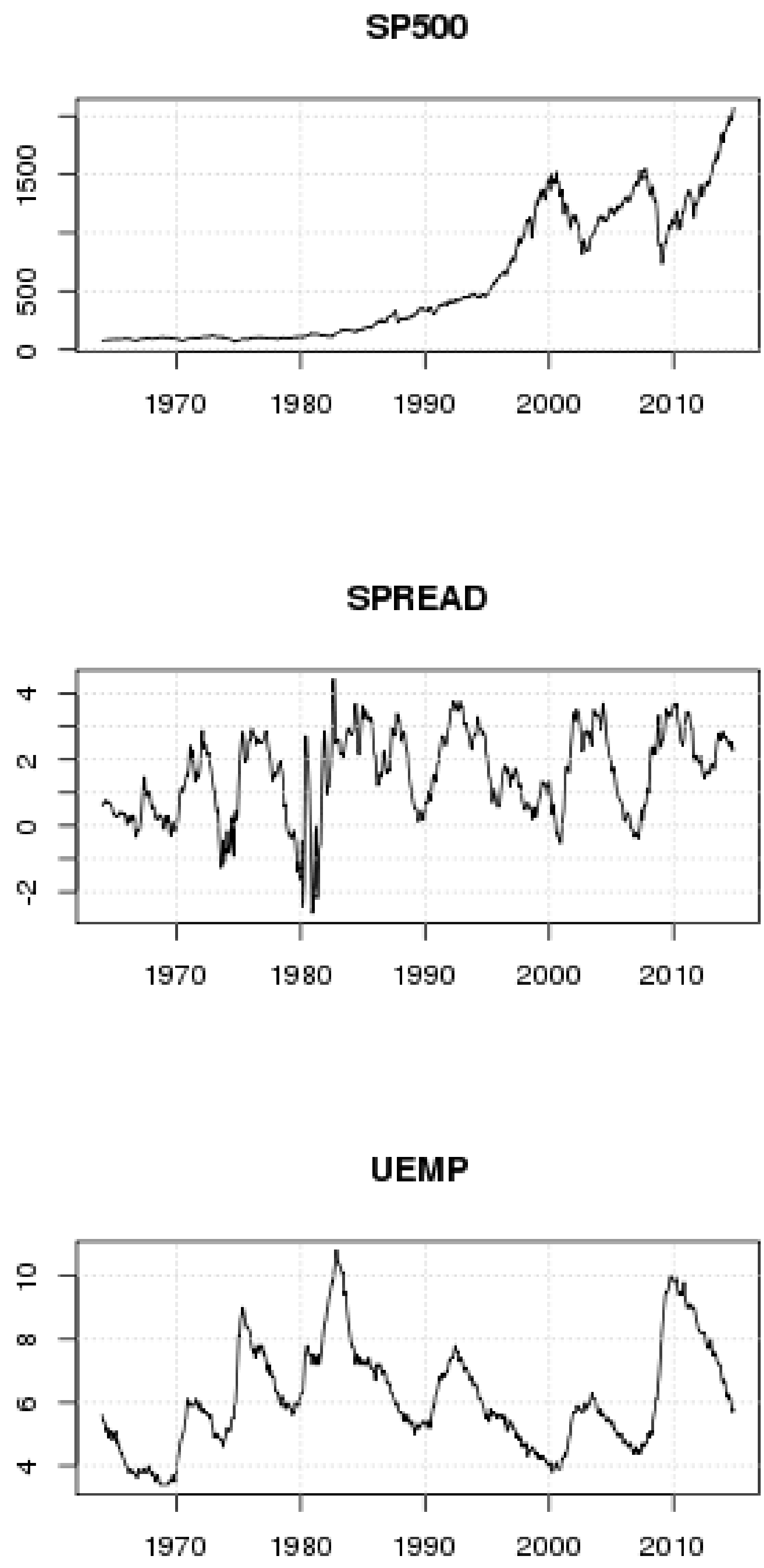


Figure 3: Time series of detrended data

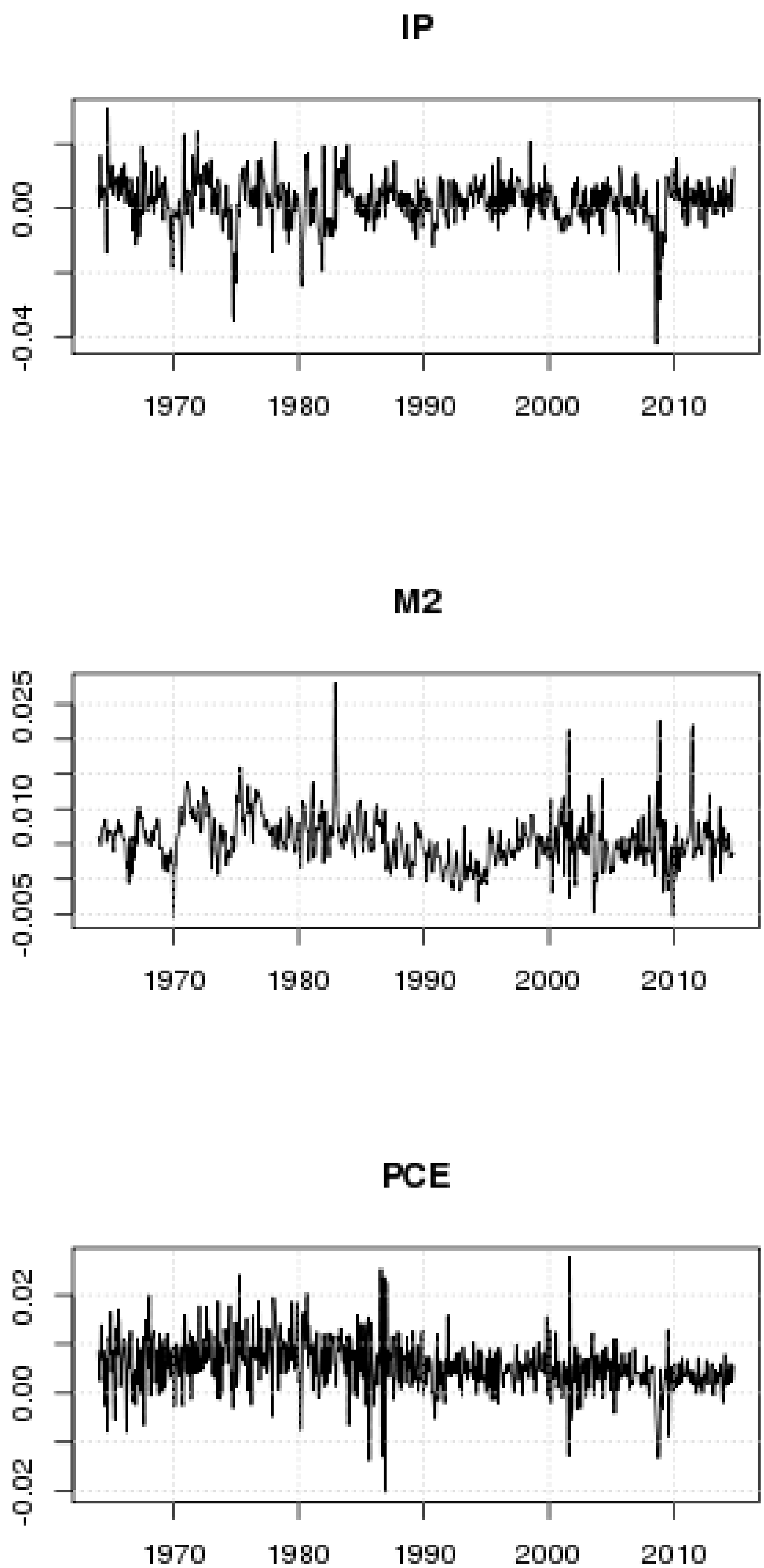


Figure 4: Time series of detrended data

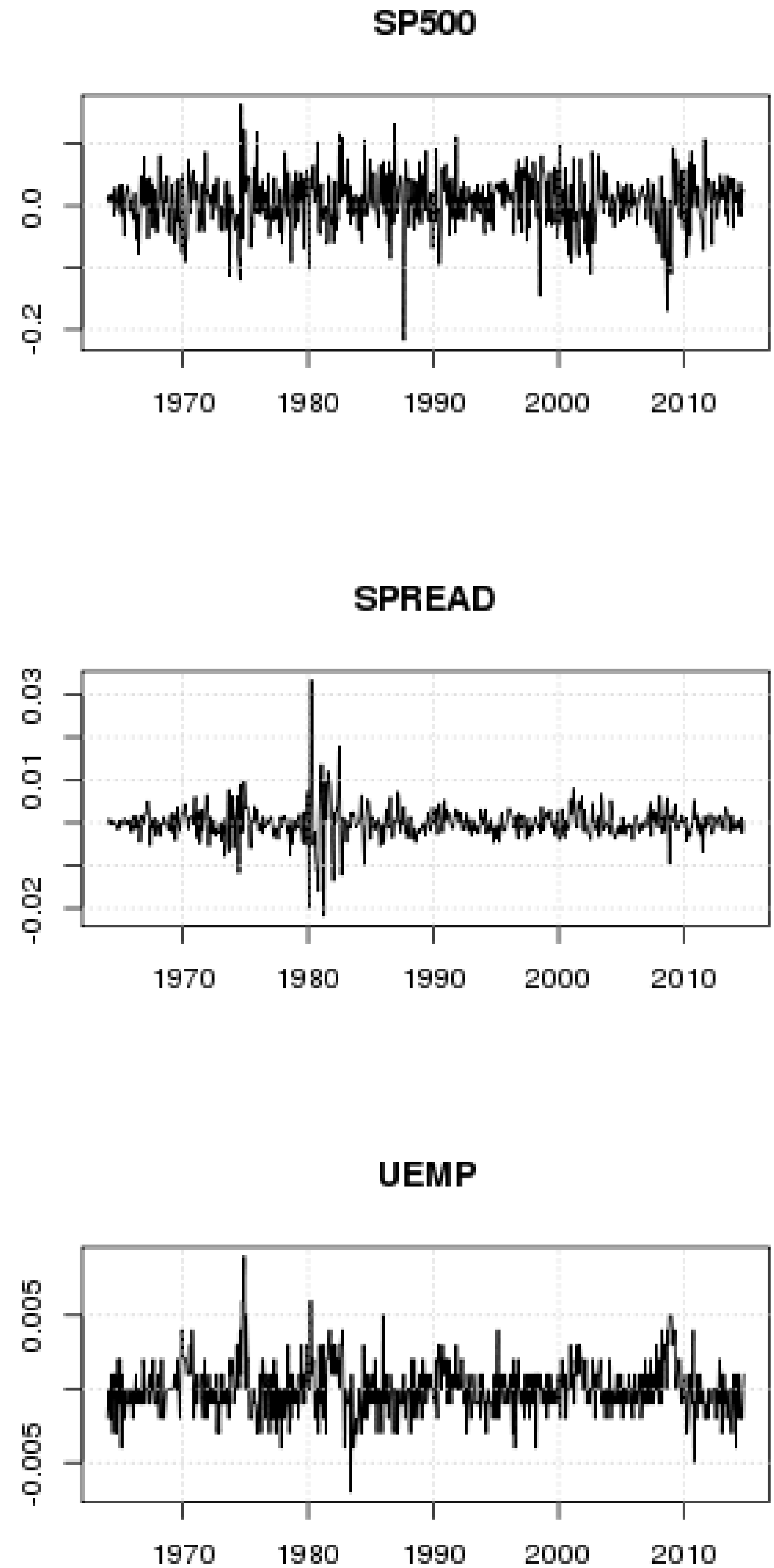


Figure 5: Time series of standard deviation of detrended data

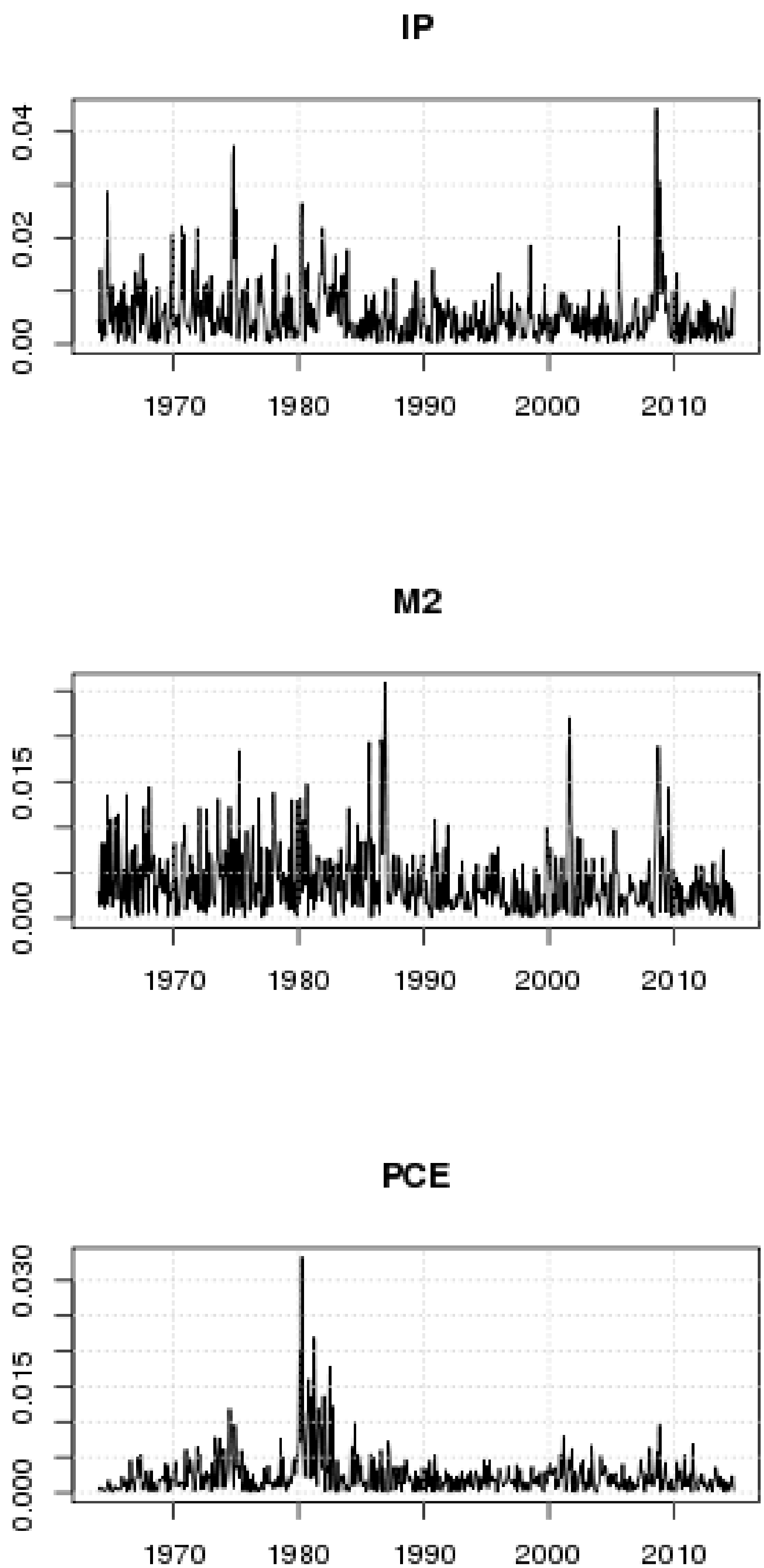


Figure 6: Time series of standard deviation of detrended data

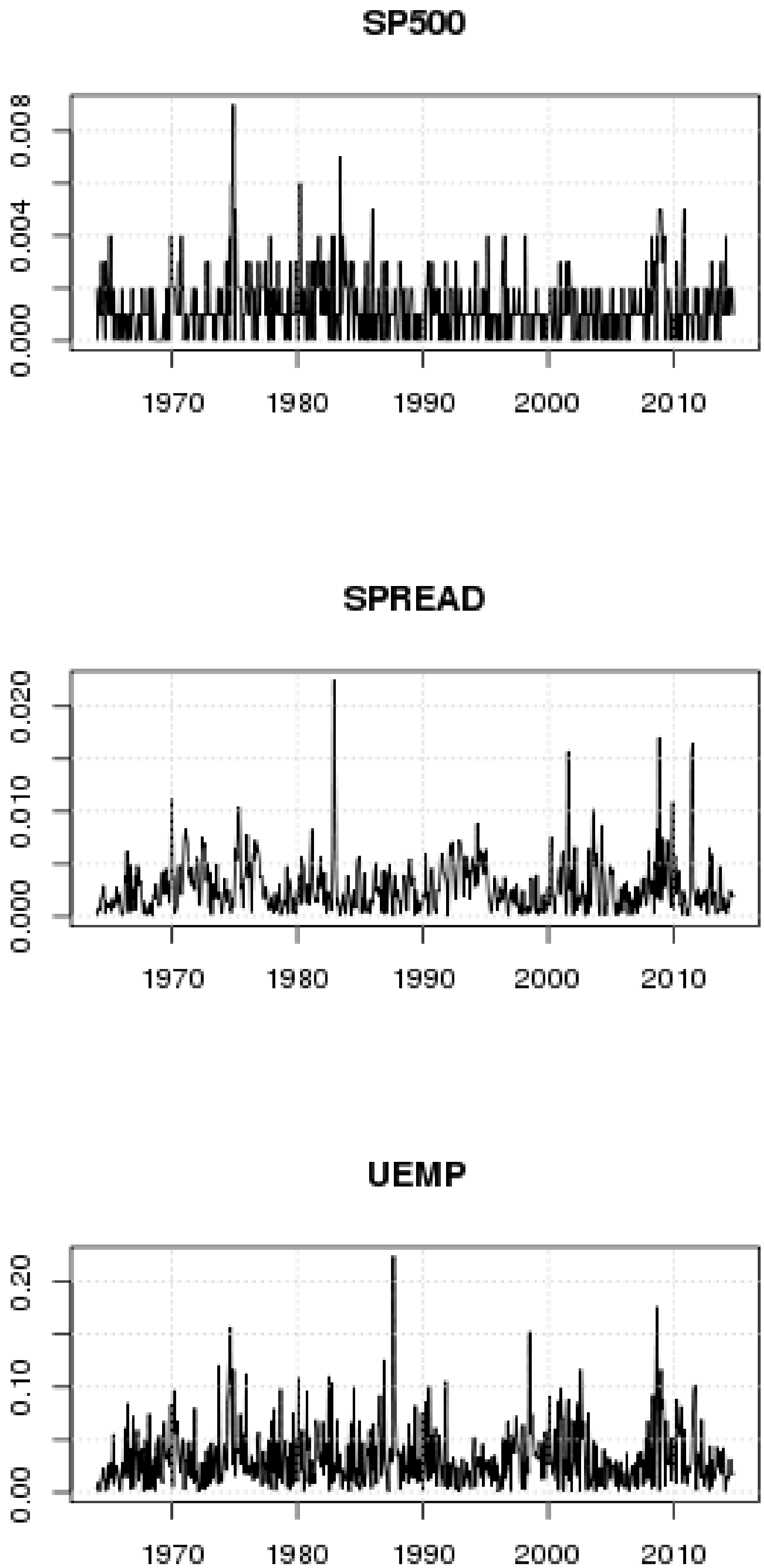


Figure 7: Density plots of detrended data using the normal kernel

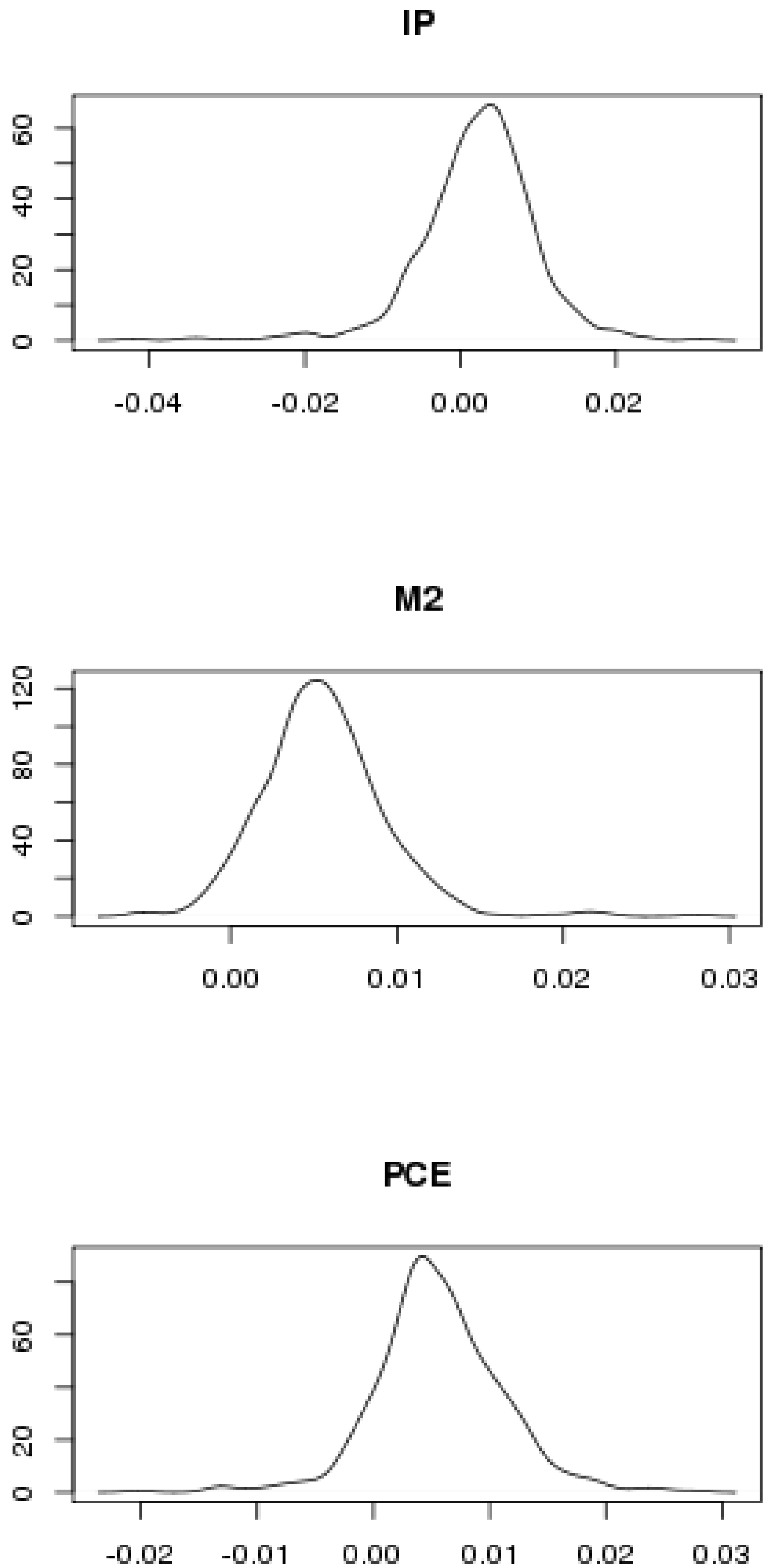


Figure 8: Density plots of detrended data using the normal kernel

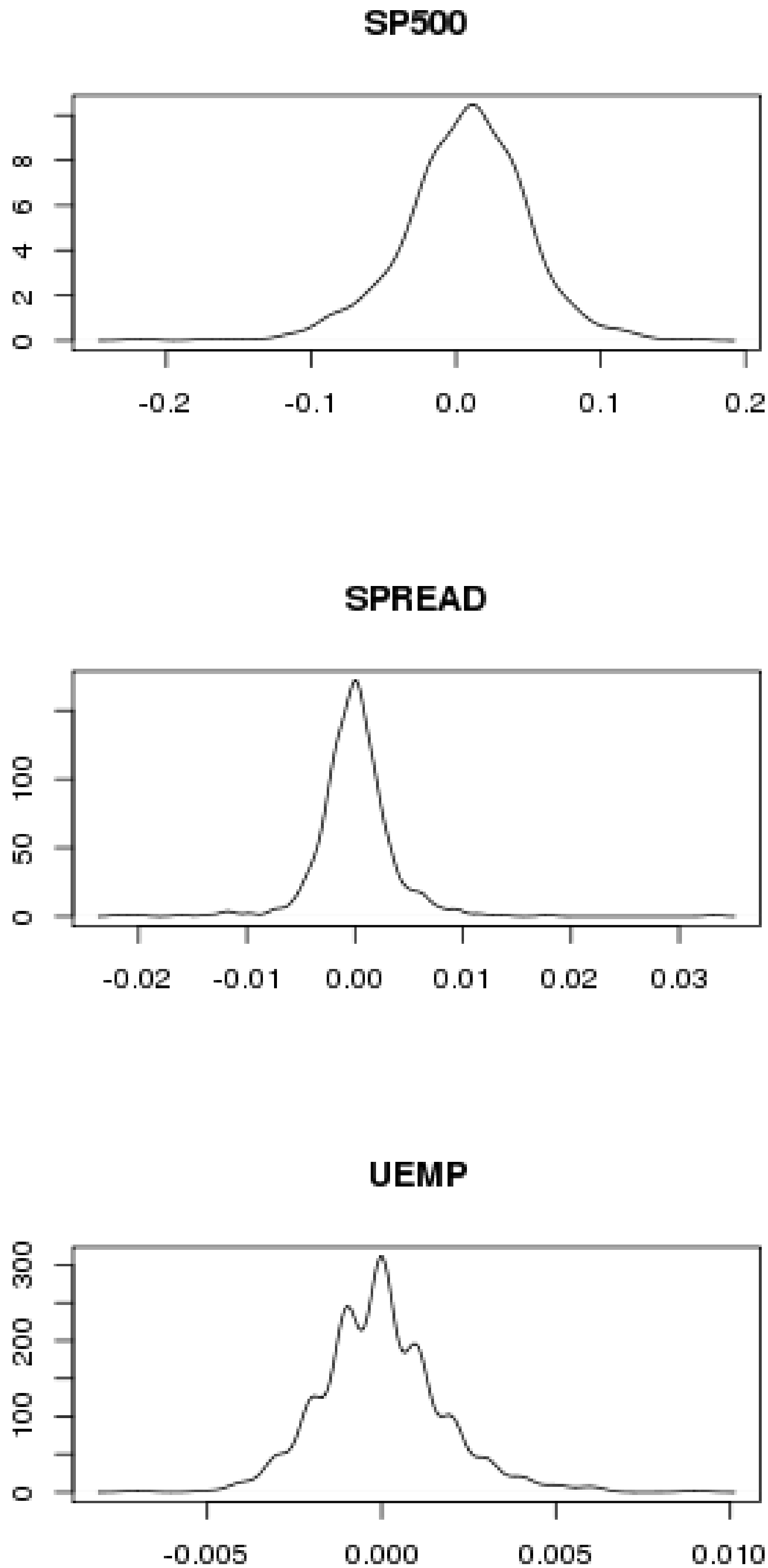


Table 1: Descriptive statistics of detrended data

Variables/Moments	Mean	Sd	Skew	Ex.Kurt
IP	0.002240	0.007461	-0.960578	4.635627
M2	0.005557	0.003706	0.920305	3.885949
PCE	0.005637	0.005485	-0.011374	2.130117
SP500	0.006341	0.043023	-0.439606	1.871890
SPREAD	0.000027	0.003714	0.679522	15.289696
UEMP	0.000003	0.001759	0.542755	1.916910

Table 1 displays the descriptive statistics of the detrended values. IP, PCE and SP500 exhibit negative skewness and positive excess kurtosis, which is confirmed by visual inspection of the density plots in Figure 7 and 8. UEMP, M2 and SPREAD have positive skewness, and SPREAD has a relatively high kurtosis. The fact that all the variables are leptokurtic indicates that the time series are not normally distributed. Therefore, a different distribution model has to be taken into account when we create the univariate models. Visual inspection of the density plot of UEMP leads to uncertainty when choosing an optimal distribution model, due to the signs of UEMP being multimodal. The relative value between mean and standard deviation for UEMP are very high, which may cause the variable to have a significantly larger impact on the other variables in the VAR model that will be modeled.

4 Methodology

4.1 Research Study Design

The authors will use four different types of models to predict volatility. The first model will be the following naive model

$$\hat{\sigma}_t^2 = \sigma_{t-1}^2, \tag{1}$$

where the volatility of the previous period σ_{t-1}^2, is the predicted volatility in the next period. This model will be used as a benchmark for the other models to see if they are able to do a better job at predicting volatility for our financial time series.

The second model will be an ARIMA(0,0,0)-GARCH(1,1) model with a normal distribution. Here we use ARIMA to model the mean the following way

$$X_t = \mu + \varepsilon_t, \tag{2}$$

where μ is the average arithmetic mean and ε_t is an error term assumed to be normally distributed. The GARCH is used to model the variance with normal distribution density. The model can be written as

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{3}$$

this is the equation for the conditional variance where ω is a constant, α is the coefficient for the ARCH parameter and β is the coefficient for the GARCH parameter. This will be used by the authors as it is a commonly used model to predict volatility for financial time series with a univariate model, which previous research have shown to often be inadequate(Schwert,1989). The third model will be an ARIMA(k,d,v)-EGARCH(p,q) model with the GED. The terms for the ARIMA(k,d,v) and the EGARCH(p,q) model will be chosen based on the information criterion HQIC except the d for ARIMA, which will be chosen prior to the optimization. HQIC is explained in the ARIMA model section that follows. The reasoning behind using an ARIMA, where we choose the AR and MA terms is based on the idea that the authors wish the model to depend on the time series available and not a preconceived notion on what is an adequate model. The authors wish to use EGARCH to see if we can create a better model by taking into account that some financial time series show signs of asymmetry in volatility which it can model. The choice of the GED distribution comes from the fact that the density of all the time series are leptokurtic in the in-sample period which is possible to model with a GED distribution unlike the normal distribution.

The fourth model used will be a VAR model, the reasoning behind this model is that the authors wish to see if the variables can help in forecasting each other and is the only multivariate model included in the thesis. The VAR model will attempt to predict the volatility for each variable using lagged values as regressors and a constant. An example of a VAR model with two variables and one lag follows

$$\sigma_{1,t} = \beta_1 + a_{1,1}\sigma_{1,t-1} + a_{1,2}\sigma_{2,t-1} + \varepsilon_{1,t} \tag{4}$$

$$\sigma_{2,t} = \beta_2 + a_{2,1}\sigma_{1,t-1} + a_{2,2}\sigma_{2,t-1} + \varepsilon_{2,t}, \tag{5}$$

where σ denotes the sample standard deviations, β the intercept terms, a coefficients for the lagged values and ε denotes the error terms. The lag length will be estimated based on the information criterion HQIC.

The models mentioned above require the time series to follow a stationary process. What is meant by a stationary process and how we will test it, can be found in the Appendix. The following subsections in the methodology explain in detail the models that will be used and the estimation process which will be performed out-of-sample, as well as how we will compare the performance of the different models. In addition to what follows in the methodology, there are several tests to see how well the ARIMA-EGARCH model fit in-sample to the time series we use which are explained in detail in the Appendix.

4.2 ARIMA Model

Yule (1927) used a pendulum analogue as a tool for gaining insight about the dynamic behavior of time series models, and used the pendulum as an inspiration to formulate an AutoRegressive model (AR) for the dependency in Wolfer's Sunspot numbers. Slutzky (1937) showed by using random numbers from a lottery that random processes can form cyclical processes. Slutzky demonstrated that if these uncorrelated variables were Moving Averages (MA) of the past, then the variables will become serial correlated. This sequence produces characteristics of a macroeconomic business cycle.

Wold (1938) combined AR(k) and MA(v) into ARMA(k,v) formulations from Yule and Slutzky. When the order of k and v terms are specified the ARMA (k,v) process can be used to model all stationary time series as a combination of past lagged values and past lagged error terms. Box and Jenkins (1976) popularized the ARIMA model and developed a methodology for identifying and estimating models that can integrate both AR and MA terms, where d is the order of difference. Box-Jenkins suggests to do the following: (1) Examine stationary data; (2) Identify model; (3) Estimate parameters; (4) Diagnostic checking; (5) Forecasting.

The general forecasting equation that will be used is

$$\hat{y}_t = \mu + \sum_{l_1=1}^{k} \phi_{l_1} y_{t-l_1} + \sum_{l_2=1}^{v} \theta_{l_2} y_{t-l_2}, \tag{6}$$

$$d \geq 1 : y_t = Y_t - 2Y_{t-1} + Y_{t-d}, \tag{7}$$

$$d = 0 : y_t = Y_t, \tag{8}$$

where Y_t denotes a variable at time t, d is the amount of times the model is differenced, k is the amount of AR terms and v the amount of MA terms. The μ term will be included if the data appears to trend around a mean $\neq 0$. The distribution model used will be the GED. When we have chosen the value of d, and decide whether to include μ or not, we will choose the values of k and v that minimize the HQIC. Although it is possible to choose the amount of AR and MA terms based on visual inspection of the ACF/PACF plots the authors feel it is better to use information criterion's due to the lack of knowledge from the authors part of the underlying process. The authors have chosen the HQIC, because it has been shown that on large sample size data it outperforms both of the commonly used Akaike's Information Criterion (AIC) and Schwartz Information Criterion (SIC) when it comes to correctly choosing the true unknown lag length of the process as shown by Liew (2004).

The differences between the information criterions come from how they penalize complex models, the HQIC is computed as

$$HQIC = \frac{-2LL}{n} + \frac{2ln(ln(n))k}{n}, \tag{9}$$

where LL is the log-likelihood, n the number of observations in the time series and k the number of parameters. It is also necessary to set an upper bound for the amount of lags the model can include. The authors have limited this to a maximum of 2 AR lags and 2 MA lags; this is done due to computational complexity of computing larger models, although the authors do not expect that including more lags would significantly change the predictive power of the model. It can also be beneficial to limit the possibility of the model including AR and MA terms that cancel each other out.

4.3 GARCH Model

Basic least squares models assume constant expected value of all error terms, which is the assumption of homoscedasticity. The violation of this assumption is heteroscedasticity. When estimating an Ordinary Least Squares (OLS) model, the standard errors and confidence intervals will give a false sense of accuracy if the error terms are heteroscedastic. This assumption is the focus of ARCH/GARCH models, because they treat heteroscedasticity as a variance to be modelled. Engle (1982) introduced a new class of stochastic processes called ARCH. This was introduced to model stochastic processes with homoscedastic error terms and made it possible to describe non-linear dynamics of financial data. The new ARCH processes can be described as mean zero, serially uncorrelated processes with non-constant variances conditional on the past, but constant unconditional variances. For these processes, the recent past gives information about one-period forecast variance (Engle,1982). The purpose of the study was to estimate the means and variances using an ARCH model, by using the inflation rate in the UK. Engle found evidence for the ARCH effect to be significant. The ARCH model was extended by Bollerslev (1986) in order to allow for a more flexible lag structure, and he proposed the GARCH process. A GARCH(p,q) model consists of a p number of AR lags (GARCH-terms), and q numbers of MA lags (ARCH-terms). The specification of this model allows for reducing the numbers of estimated parameters, and the model is mean reverting and conditionally heteroscedastic with a constant unconditional variance.

The original GARCH model developed by Bollerslev (1986) can be written the following way

$$\sigma_t^2 = \omega + \sum_{j=1}^{m} \zeta_j \upsilon_{jt} + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2. \tag{10}$$

When we work with the normal distribution the ζ, parameter is set equal to 0 and we end up with

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2. \tag{11}$$

σ_t^2 is the conditional variance at time t, α the ARCH parameter coefficient and β the GARCH parameter coefficient. The length of the model depends on the amount of ARCH(q) and GARCH(p) terms included, as a baseline we have set $q = 1$ and $p = 1$.

4.4 EGARCH Model

Nelson (1991) pointed out three main drawbacks when applying GARCH in asset pricing: (1) GARCH rules out the assumption of a negative correlation between current returns and future return volatility; (2) Violation of the parameter restrictions that GARCH models inflict can be caused by estimated coefficients, which may unnecessarily restrict the dynamics of the conditional variance process; (3) It can be difficult to illuminate whether shocks to conditional variance "persist" or not in the standard GARCH model. To meet these drawbacks, Nelson proposed a new form of GARCH, the EGARCH. The model, which is an asymmetric GARCH model, has no parameter restrictions, and invariably produces positive conditional variance.

The authors have decided to use EGARCH, because it can be used to model asymmetric behavior in volatility. The EGARCH is defined by Nelson (1991), as

$$ln(\sigma_t^2) = \omega + \sum_{j=1}^{m} \zeta_j \upsilon_{jt} + \sum_{j=1}^{q} (\alpha_j z_{t-j} + \gamma_j(|z_{t-j}| - E|z_{t-j}|)) + \sum_{j=1}^{p} \beta_j ln(\sigma_{t-j}^2). \qquad (12)$$

We optimize the GARCH model by keeping the ARIMA model constant with the appropriate specification found when optimizing the ARIMA model. The distribution model will be GED, which is the one used when modelling the ARIMA. The coefficients that are estimated is α for ARCH effects, β for GARCH effects, ω the intercept of the conditional variance, ζ to model shape of density and γ used to model volatility asymmetry.

Combinations of EGARCH(p,q) parameters will be tested with p ranging from 0-2, and q ranging from 1-2 in-sample. The reason for this is that it would take too long and require too much computation if we increased the range. $q = 0$ is not feasible when implementing the EGARCH model and is therefore, not tested. The model that minimizes the HQIC will be chosen as the preferred model.

4.5 Vector Autoregressive Model

During the 1970's, most of the western countries experienced stagflation (The Great Inflation). This was a period that could be described with high inflation rates, high unemployment rates and slow output growth. In the 1970's empirical methods in macroeconomics were mainly linear statistical systems, which were built around the Keynesian model. Due to including the stagflation period these models became instable, and received criticism, because these models relied on heavy identification assumptions and were difficult to interpret because of misestimating which rendered them obsolete in forecasting and monetary policy analysis.

Sims (1980) advocated the use of VAR models, which is an extension of a univariate AR model. This linear system is an n-equation, n-variable system and describes dynamics in multivariate time series, where each variable depends on past values of itself and past values of the other macroeconomic variables. The VAR system evolved to deal with the problems of previous models, and provided a feasible technique to forecasting, monetary policy analysis, and data description in a macroeconomic environment.

There has been written several technical and practical economic application studies about VAR. Lutkepohl (1991) and Lutkepohl (2011) gives the technical reference for VAR models, whereas Waggoner and Zha (1999) developed Bayesian method to computing probability distributions and error band in the light of conditional forecasts in the VAR framework. Stock and Watson (2001) evaluated how well VAR deals with data description, forecasting, structural inference, and policy analysis. As for describing data and generating benchmark forecasts VAR systems work as useful techniques. Granger causality tests, impulse responses and forecast error variance decompositions are standard VAR summary statistics that capture co-moments of the lagged values of multiple time series. "Yet, without modification VAR models miss conditional heteroscedasticity, non-linearity and drifts or breaks in parameters."(Stock and Watson,2001)

VAR models have the feature of being sensitive to variables. Two or three variables can be unstable, and thus be inadequate predictors, while too many variables create complications. Stock and Watson emphasize that VAR systems can analyse two types of policies which are changes in policy rules and surprise monetary policy. Whether VAR can contribute to structural inference and policy analysis or not, is questionable.

Poon and Granger (2003) provides an overview of 93 different papers on volatility forecasting, and examined whether volatility is predictable and which method will provide the best forecasts. Historical volatility models (which are models based on historical volatility using moving averages, exponential weights, autoregressive models) is better than GARCH models at forecasting in 56% of the cases, when both GARCH and historical volatility models in the studies were included. GARCH models outperform ARCH, and asymmetric GARCH models (such as EGARCH) perform better than GARCH.

While the ARIMA-EGARCH estimations are univariate, we will also create a VAR to model a potential dynamic interrelationship between the variables. In practice this means we try to explain each variable by itself and also the other variables. The general VAR model that will be used is the following

$$\begin{pmatrix} Y_{1,t} \\ \vdots \\ Y_{k,t} \end{pmatrix} = \begin{pmatrix} \beta_{1,t} \\ \vdots \\ \beta_{k,t} \end{pmatrix} + \sum_{l=1}^{p} \left[\begin{pmatrix} a^l_{1,1} & \dots & a^l_{1,k} \\ \vdots & \ddots & \vdots \\ a^l_{k,1} & \dots & a^l_{k,k} \end{pmatrix} \begin{pmatrix} Y_{1,t-l} \\ \vdots \\ Y_{k,t-l} \end{pmatrix} \right] + \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{k,t} \end{pmatrix}, \tag{13}$$

where k is the amount of time series in the VAR model, l the amount of lags included, Y the time series variables, a the coefficients for the lagged variables, β the constants, and ε the error terms. All variables are considered endogenous, so there are no exogenous variables. The VAR model will be estimated in-sample using maximum likelihood estimation (MLE). It is important to note that the variables in this model will be the sample standard deviations of the detrended variables used in the ARIMA-EGARCH, and that VAR also requires the variables to be stationary.

4.6 Estimation Procedure

To make the estimation process similar for all statistical models, they will be refitted every month using a recursive window. All previous information is used, and the models will therefore be refitted 239 times as the out-of-sample consists of 240 months. There is cause for concern if there are structural breaks, as the relationships between the coefficients that existed in the past may not necessarily exist for the future, if there is lack of structural stability.

4.7 Granger Causality

To interpret and analyze the dynamics of a VAR model, different types of structural analysis is used. As mentioned earlier in the thesis, Granger causality suggested by Granger (1969) is a test for describing data in multivariate time series. Assuming only stationary series, Granger defines causality as: "Variable Y is causing X, if we are better able to predict X using all available information than if the information apart from

Y has been used."(Granger,1969) In other words, the Granger causality statistics can tell if lagged values of variable X, help to predict variable Y. Granger did not wish to imply true causality. Causality can be misleading, as it implies a correlation between current value and past value. We will run a test for Granger causality to determine if the variables are useful in predicting each other. It is possible to do this in a multivariate setting, but the most common way to do it is bivariate, which the authors have opted to do where the Residual Sum Squares (RSS) is computed for the following models

$$y_{i,t} = \alpha + \sum_{l=1}^{p} \beta_p y_{i,t-1} + \sum_{l=1}^{p} \gamma_p x_{i,t-p} + \varepsilon_{i,t}, \tag{14}$$

$$y_{i,t} = \alpha + \sum_{l=1}^{p} \beta_p y_{i,t-1} + \varepsilon_{i,t}. \tag{15}$$

The lag length is equal for both y and x and is based on the lag length, found when creating the VAR model. If there is no Granger causality, the past values of x should not affect y. H_0 : all $\gamma_l = 0$. This is tested using an F-test, which is estimated using OLS.

$$F = \frac{\left(\frac{RSS_1 - RSS_2}{p_2 - p_1}\right)}{\left(\frac{RSS_2}{n - p_2}\right)}. \tag{16}$$

Where RSS is computed for both the model that has x_t included in it and the one that is a simple AR model.

4.8 Forecast Performance Evaluation

When evaluating the different estimation results, we will compare them using Mean Square Error (MSE), this is computed as the difference between predicted variance and realized variance in the out-of-sample period.

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (\hat{Y}_t - Y_t)^2. \tag{17}$$

Table 2: ADF test on raw data in the in-sample period. t-stat denotes the value of the test statistic. Crit.values denote critical values for the test at the significance levels 1%,5%,10% respectively, which are dependent on whether or not a trend term is included. The H_0 is rejected if $|t.stat| > |Crit.value|$. Lag denotes the lag length chosen that minimize MAIC. ϕ denotes the estimated root of the series.

	t-stat	Crit.values	Lag	ϕ
PCE	-1.0573	(-3.96, -3.41, -3.13)	10	1.0020
IP	-2.9299	(-3.96, -3.41, -3.13)	4	0.9953
SP500	-1.1989	(-3.96, -3.41, -3.13)	3	0.9981
SPREAD	-3.8836	(-3.96, -3.41, -3.13)	12	0.9521
M2	2.9366	(-3.96, -3.41, -3.13)	18	1.0047
UEMP	-2.9471	(-3.96, -3.41, -3.13)	4	0.9942

*$p < 0.1$,**$p < 0.05$,***$p < 0.01$,$n = 371$

Where n is the out-of-sample observations, $\hat{Y}$ a vector of predictions and Y a vector of realized values. By using MSE, we are penalizing all types of deviations from the observed value equally. There is a case to be made for using difference performance measures that penalize the big deviations relatively more than the small deviations, that being said the authors feel this is not of significant importance for this thesis, as the conclusion should remain the same regardless of choice of performance measure.

5 Results

5.1 Stationarity

Looking at Figure 1 and 2 in the data section, it becomes clear that the raw data does not follow a stationary process; they also show clear signs of trending/cyclical behavior so we include a trend component in the ADF test. The ADF test and how it is implemented is explained in detail in the Appendix. Table 2 reports the results for the ADF test on the raw variables. We fail to reject H_0 for all variables but SPREAD. Looking at the graph of SPREAD and having the prior knowledge that the ADF test is biased towards stationarity when using monthly data, the authors decide to de-trend all the variables.

Figure 3 and 4 in the data section shows the detrended variables.B ased on visual inspection, the authors decide to include a trend term in the ADF for PCE and M2.

Table 3: ADF test on detrended data in the in-sample period. t-stat denotes the value of the test statistic. Crit.values denotes critical values for the test, at the significance levels 1%,5%,10% respectively, which are dependent on whether a trend term is included or not. The H_0 is rejected if $|t.stat| > |Crit.value|$. Lag denotes the lag length chosen that minimize MAIC. ϕ denotes the estimated root of the series.

	t-stat	Crit.values	lag	ϕ
PCE	-4.1648***	(-3.96, -3.41, -3.13)	18	-0.1585
IP	-5.4516***	(-3.43, -2.86, -2.57)	17	0.4927
SP500	-6.5684***	(-3.43, -2.86, -2.57)	11	0.04505
SPREAD	-19.2212***	(-3.43, -2.86, -2.57)	0	0.2481
M2	-3.5627**	(-3.96, -3.41, -3.13)	16	0.5426
UEMP	-4.5850***	(-3.43, -2.86, -2.57)	18	0.6002

$^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$,$n = 370$

Table 3 report the results of the ADF test run on the detrended variables. H_0 is rejected for all the time series and we will from this point onwards treat them as stationary.

Table 4: ADF test on standard deviation of the detrended data in the in-sample period. t-stat denotes the value of the test statistic. Crit.values denotes critical values for the test at the significance levels 1%,5%,10% respectively, which are dependent on whether a trend term is included or not. The H_0 is rejected if $|t.stat| > |Crit.value|$. Lag denotes the lag length chosen that minimize MAIC. ϕ denotes the estimated root of the series.

	t-stat	Crit.values	lag	ϕ
PCE	-5.1364***	(-3.43, -2.86, -2.57)	13	0.3860
IP	-4.1665***	(-3.43, -2.86, -2.57)	18	0.3034
SP500	-3.4923***	(-3.43, -2.86, -2.57)	11	0.5628
SPREAD	-4.6965***	(-3.43, -2.86, -2.57)	18	0.3580
M2	-4.2078***	(-3.43, -2.86, -2.57)	17	0.6779
UEMP	-4.4789***	(-3.43, -2.86, -2.57)	14	0.5447

$^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$,$n = 370$

Looking at Figure 5 and 6 in the data section, which shows the standard deviation of the detrended variables, the authors see no indications of trending or cyclical behavior. Table 4 report the results for the ADF test on the standard deviation of the detrended variables. We reject H_0 for all the variables and treat them as covariance stationary.

5.2 ARIMA

Based on Figure 3 and 4 in the data section, the authors decide to include a mean term for IP, M2 and PCE, as there appear to be a constant mean different from 0 in these time series. All the variables are integrated of order 1, based on the results from the ADF and visual inspection. Due to the large sample size and the fact that the p-values given assumes normality, most of the variables are highly statistically significant with the exceptions being SP500 and UEMP. As the assumption that the variables are normally distributed is violated, the information obtained from the p-values is limited. Table 5 shows the optimized ARIMA models that minimize HQIC. The authors are unable to create a useful ARIMA model for UEMP as we have no significant coefficients for the ARIMA model. This is a cause for concern, as this can impact the EGARCH model. It is important to note that the distribution parameter ζ, in spite of being highly significant, in no way proves that the choice of distribution is correct, but rather that the parameters are $\neq 0$. Although the ARIMA model impacts the EGARCH model, it is possible to obtain useful estimates for volatility, even if one fails to estimate the mean.

Table 5: Optimal ARIMA coefficients for the ARIMA model that minimize HQIC. μ denotes the coefficient of the intercept term, when it is included, AR1(2) denotes the coefficient for the AR term(s). MA (2) denotes the coefficients of the MA term(s).Sigma denotes the coefficient for the constant variance. ζ denotes the coefficient of the shape parameter for the distribution model.

	IP	M2	PCE	SP500	SPREAD	UEMP
μ	0.003***	0.006***	0.007***	–	–	–
AR1	0.756***	1.525***	0.990***	–	–	–
AR2	–	-0.530***	–	–	–	–
MA1	-0.462***	-0.746***	-1.220***	0.039	0.218***	0.000
MA2	–	-0.181***	0.256***	–	-0.149***	–
Sigma	0.007***	0.002***	0.006***	0.043***	0.004***	0.014***
ζ	1.056***	1.213***	1.333***	1.357***	0.778***	0.111***

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$, $n = 370$

Table 6: Optimal ARIMA-EGARCH coefficients for the ARIMA-EGARCH model that minimize HQIC. The coefficients up to and including MA1(2) belong to the mean model, ζ belongs to the distribution model and the remaining coefficients belong to the variance model. μ denotes the coefficient of the intercept term when it is included in, AR1(2) denotes the coefficient for the autoregressive term(s). MA1(2) denotes the coefficients of the MA term(s). ω denotes the intercept term for the variance model, $\alpha 1(2)$ denotes the coefficient for the ARCH effect(s), $\beta 1(2)$ denotes the coefficient for the GARCH effect(s), $\gamma 1(2)$ denotes the coefficient for the symmetry term and ζ denotes the coefficient of the shape parameter for the distribution model.

	IP	M2	PCE	SP500	SPREAD	UEMP
μ	0.003***	0.006***	0.007***	–	–	–
AR1	0.814***	1.526***	1.000***	–	–	–
AR2	–	-0.537***	–	–	–	–
MA1	-0.579***	-0.738***	-1.229***	-0.008	0.345***	0.000
MA2	–	-0.152***	0.258***	–	-0.078	–
ω	-0.920***	-7.806***	-0.910*	-0.552***	-0.754**	-1.004**
$\alpha 1$	-0.226***	0.203**	-0.048	-0.363***	0.045	-0.027
$\alpha 2$	–	–	–	0.279***	–	–
$\beta 1$	0.909***	0.361***	0.913***	0.968***	0.935***	0.789***
$\beta 2$	–	–	–	-0.056***	–	–
$\gamma 1$	0.127***	0.545***	0.195***	-0.253**	0.552***	5.248
$\gamma 2$	–	–	–	0.478***	–	–
ζ	1.251***	1.402***	1.440***	1.662***	1.389***	0.100***

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$, $n = 370$

5.3 EGARCH

Table 6 shows the optimized ARIMA-EGARCH models that minimize HQIC. We see that the ARIMA coefficients have changed slightly. This happened because they have been optimized simultaneously with the EGARCH coefficients. The reason for the original ARIMA optimization is to choose AR and MA lags. Similarly, to the ARIMA coefficients most of the EGARCH coefficients are statistically significant due to the large sample size and the fact that the normality assumption is violated. We see that $\alpha 1$ is not significantly different from 0 for PCE, SPREAD and UEMP. This indicates that our estimations of ARCH effects are not statistically different from 0 and that if there are any ARCH effects present, our model does a poor job of estimating them.

All models show statistically significant GARCH effects as seen by looking at the β coefficients. The biggest immediate concern is the lack of significance in the $\gamma 1$ for UEMP. Looking at Figure 8 for UEMP it appears the multimodality shown is mistaken for asymmetry, as the distribution model is not able to deal with time series that are multimodal. In order to get useful estimates, the authors have therefore decided to enforce the restriction $\gamma 1 = 0$ for UEMP from this point onwards. It is important to keep in mind that these coefficients will change throughout the out-of-sample period when based on the new information obtained after each period.

5.4 Weighted Ljung-Box

The first variables we use to implement the Ljung-Box test on are the standardized residuals of our model tested in-sample; this is done to see if the residuals exhibit ARIMA effects. The Weighted Ljung-Box test and how it is implemented is explained in detail in the Appendix. Sign of ARIMA effects indicates our model is misspecified, as the model is meant to model the ARIMA effects the variables exhibit. Table 7 reports the results from the Weighted Ljung-Box test. We see that the only ARIMA-EGARCH that shows signs of this is the model for UEMP. This is not surprising, as we have already shown that we do a poor job of modelling UEMP due to the nature of the variables density distribution. While we do not detect any ARIMA effects for the other variables, this is tested in-sample, and the model is optimized in-sample, so we cannot say that this will hold for the out-of-sample period. It is, however, a good sign that it appears that we have managed to capture the ARIMA effects in the in-sample period.

Table 7: Weighted Ljung-Box test on standardized residuals, t-values for the tests are shown with p-values in parenthesis. The lag lengths vary case by case, as the models have different amounts of AR and MA terms.

Variables/Lag	1	2*(p+q)+(p+q)-1	4*(p+q)+(p+q)-1
IP	0.018 (0.894)	3.511 (0.202)	7.189 (0.111)
M2	0.064 (0.800)	5.561 (0.762)	11.187 (0.289)
PCE	0.214 (0.644)	1.442 (1.000)	3.12 (0.996)
SP500	0.266 (0.606)	0.332 (0.992)	1.114 (0.928)
SPREAD	0.159 (0.690)	1.334 (1.000)	2.221 (0.973)
UEMP	27.800*** (0.000)	44.540*** (0.000)	59.590*** (0.000)

$^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$,$n = 370$

Table 8 reports the results of the Weighted Ljung-Box test on the standardized squared residuals. The results here are similar to what we found in the previous Ljung-Box test, where it is clear that our ARIMA-EGARCH model of UEMP is misspecified. We also see that the model for IP at 10% significance level rejects the H_0 of no auto-correlation indicating that there may be ARCH/GARCH effects not captured by this ARIMA-EGARCH model.

5.5 Sign Bias

Table 9 reports the results of the Sign bias tests. The Sign bias test and how it is implemented is explained in detail in the Appendix. The only statistical evidence towards sign bias is found for the SPREAD variable, although we are unable to say what type of sign bias it is. While this test does not prove the absence of sign bias on our coefficients it is useful as a pointer on how our estimated coefficients capture the effect the variables have on the endogenous variable. That being said when we look at sign bias in-sample;

Table 8: Weighted Ljung-Box tests on standardized squared residuals, t-values for the tests are shown with p-values in parenthesis. The lag lengths vary case by case as the models have different amounts of ARCH and GARCH terms.

Variables/Lag	1	2*(p+q)+(p+q)-1	4*(p+q)+(p+q)-1
IP	3.443* (0.064)	4.691 (0.180)	6.556 (0.239)
M2	0.046 (0.830)	0.812 (0.901)	2.071 (0.896)
PCE	0.385 (0.535)	1.953 (0.630)	3.120 (0.726)
SP500	0.325 (0.569)	4.582 (0.645)	8.609 (0.606)
SPREAD	0.126 (0.723)	1.578 (0.721)	2.617 (0.820)
UEMP	20.310*** (0.000)	127.270*** (0.000)	188.650*** (0.000)

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$, $n = 370$

we would expect that there is little or no evidence for bias as long as the data do not exhibit structural breaks where the coefficients change significantly.

Table 9: Sign bias test, t-values for the tests are shown with p-values in parenthesis.

Variables	Sign Bias	Negative Sign Bias	Positive Sign Bias	Joint Effect
IP	1.325 (0.186)	0.338 (0.736)	0.425 (0.671)	3.666 (0.300)
M2	0.616 (0.539)	0.496 (0.620)	0.416 (0.678)	0.463 (0.927)
PCE	1.559 (0.120)	0.824 (0.410)	1.615 (0.107)	3.448 (0.328)
SP500	1.105 (0.270)	0.012 (0.990)	0.677 (0.499)	1.727 (0.631)
SPREAD	2.402** (0.017)	1.071 (0.285)	0.617 (0.538)	6.014 (0.111)
UEMP	0.285 (0.776)	0.005 (0.996)	0.375 (0.708)	0.165 (0.983)

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$, $n = 370$

5.6 Nyblom Stability

While Nybloms stability test is useful to test the assumption of structural stability for our coefficients, its biggest flaw is the fact that it says nothing of when a potential structural break has occurred. We, therefore, cannot say anything about when the actual structural breaks have occurred other than that we have statistical evidence for it happening. The Nyblom Stability test and how it is implemented is explained in detail in the Appendix. Due to the long time horizon the model is estimated over, the authors find it not surprising that we see multiple coefficients show signs of structural instability. Table 10 reports the results of the Nyblom stability tests. H_0 is rejected for most of the coefficients for UEMP which indicates structural instability. This can help explain why the models density-distribution is multimodal and why we fail to create a sufficient ARIMA-EGARCH model for the variable. This also enforces the belief that we will encounter issues, when we attempt to use the model out-of-sample.

Table 10: Nyblom stability test χ^2 statistics shown for the ARIMA-EGARCH coefficients. The coefficients up to and including MA1(2) belong to the mean model, ζ belongs to the distribution model, and the remaining coefficients belong to the variance model. μ denotes the coefficient of the intercept term, when it is included in, AR1(2) denotes the coefficient for the AR term(s). MA1(2) denotes the coefficients of the MA term(s). ω denotes the intercept term for the variance model, $\alpha 1(2)$ denotes the coefficient for the ARCH effect(s), $\beta 1(2)$ denotes the coefficient for the GARCH effect(s), $\gamma 1(2)$ denotes the coefficient for the symmetry term and ζ denotes the coefficient of the shape parameter for the distribution model. Joint is an approximation of the overall structural stability of the model.

Statistics	IP	M2	PCE	SP500	SPREAD	UEMP
μ	0.582**	0.541**	0.007	–	–	–
AR1	0.216	0.241	0.069	–	–	–
AR2	–	0.244	–	–	–	–
MA1	0.191	0.108	0.073	0.106	0.551**	0.139
MA2	–	0.160	0.102	–	0.116	–
ω	1.025***	0.059	0.387*	0.163	0.122	1.452***
$\alpha 1$	0.028	0.449*	0.029	0.206	0.130	0.217
$\alpha 2$	–	–	–	0.376*	–	–
$\beta 1$	1.068***	0.058	0.402*	0.175	0.120	1.596***
$\beta 2$	–	–	–	0.175	–	–
$\gamma 1$	0.333	0.034	0.843***	0.067	0.485**	–
$\gamma 2$	–	–	–	0.125	–	–
ζ	0.066	0.050	0.111	0.164	0.064	13.483***
Joint	2.649***	2.878**	1.721	1.562	1.641	106.573***

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$, $n = 370$, individual χ^2 statistics shown

IP, M2 and SPREAD all shows sign of structural breaks for the ARIMA coefficients at a 5% significance level. This is not detrimental to the model, however as our main concern is modelling the volatility and not the mean of the model. The authors are most concerned about PCE and SPREAD showing sign of structural breaks in the asymmetry variable $\gamma 1$. If $\gamma 1$ changes significantly; it means we will either be over or underestimating the difference between positive and negative shocks to the volatility. We also see that ω and $\beta 1$ show sign of structural instability for PCE, which would indicate that our conditional volatility intercept and the GARCH effect have changed in the past. This is a cause for concern, as we may be working with coefficients that are useless when estimating future volatility.

5.7 Pearson Goodness of Fit

Table 11 reports the results of the Pearson goodness of fit test. The Pearson goodness of fit test and how it is implemented is explained in detail in the Appendix. For most variables we fail to reject that GED can be used to model the variables. This is not evidence that the variables follow the exact GED found from optimization, but rather indicates how good the distribution fits to the data in-sample. Considering the fact

Table 11: Goodness of fit test, table shows t-values and their respective p-values in parenthesis. The estimated distribution model is tested against histograms with 20-50 groups.

Variables/Group	20	30	40	50
IP	21.351 (0.318)	34.757 (0.213)	43.189 (0.297)	49.730 (0.444)
M2	18.757 (0.473)	25.676 (0.643)	39.081 (0.466)	44.865 (0.641)
PCE	21.568 (0.306)	33.946 (0.241)	36.703 (0.575)	64.054* (0.073)
SP500	24.811 (0.167)	37.189 (0.141)	39.730 (0.437)	48.649 (0.487)
SPREAD	9.892 (0.956)	23.405 (0.758)	21.351 (0.990)	42.973 (0.715)
UEMP	755.946*** (0)	1141.514*** (0.000)	1427.838*** (0.000)	1640.000*** (0.000)

$^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$,$n = 370$

that we had to enforce earlier that UEMP's $\gamma 1 = 0$, it is no surprise that we have clear statistical evidence that the GED is not suitable to model the distribution of UEMP. We also have signs when creating a histogram with 50 groups that we can reject that PCE can be modelled by GED on a 10% significance level, but not at the 5% level. It therefore, looks as though our choice of GED suits modelling the distribution of all variables except UEMP.

5.8 Vector Autoregressive Model

Table 12 reports the results of choosing lag length with different information criterions. The lag length of the model is set to be 1, as this is the lag length that minimize HQIC. It is also the choice for the other information criterions the authors have available. This is not a result the authors expected as we would expect AIC to prefer a longer lag length than the other choices.

Table 12: Lag length selection for VAR, all numbers have been rounded to the closest digit, originally 10 lags were tested, although it may appear that the first lags for AIC are identical, this is simply due to the rounding.

IC/Lags	1	2	3	4	5	6	7	8
AIC	-66.4*	-66.4	-66.4	-66.3	-66.3	-66.2	-66.1	-66.1
HQIC	-66.2*	-66.0	-65.9	-65.6	-65.5	-65.3	-65.0	-64.8
SC	-65.9*	-65.5	-65.1	-64.6	-64.2	-63.8	-63.3	-62.9

*Denotes chosen lag length

Table 13 reports the optimal VAR model coefficients found in-sample using a lag length of 1. These coefficient estimates will change as the model obtains new data for each subsequent period. The diagonal of the matrix (excluding R^2_{adj}) gives the coefficients for the AR(1) term for each equation. It is important to keep in mind that the variables we are working with are the standard deviations of return for IP, PCE, M2 and SP500, and standard deviations of the change in % for SPREAD and UEMP. The authors find it interesting that there are no variables showing negative coefficients indicating that we see no signs of negative correlation between any variables except SP500 in the SPREAD equation, where the coefficient is statistically insignificant. There are also significant differences when it comes to the degree of standard deviation for these variables, while it is feasible that a variable such as SP500 can change for more than 10% in one month; this is not feasible for any of the other variables. Most of the variables are found to be statistically different from 0. UEMP is the only variable where the AR term is statistically insignificant which further explains why we have had trouble modelling it univariate previously. We also have to be cautious, when it comes to interpreting the

Table 13: Estimated coefficients in-sample VAR, coefficient estimates are shown for the variables lagged once. R^2_{adj} is computed individually for each equation as one would in a univariate linear regression.

Equations/Variables	IP	PCE	SPREAD	UEMP	M2	SP500	R^2_{adj}
IP	0.343***	0.175***	0.163**	0.677***	0.183*	0.024**	0.591
PCE	0.085**	0.268***	0.104	0.194	0.248***	0.033***	0.530
SPREAD	0.074**	0.101**	0.376***	0.298**	0.106*	-0.001	0.474
UEMP	0.091***	0.026	0.033	0.078	0.057**	0.006**	0.518
M2	0.035*	0.001	0.071**	0.097	0.668***	0.007**	0.701
SP500	0.136	1.661***	1.438***	3.790***	1.831***	0.173***	0.510

*$p < 0.1$,**$p < 0.05$,***$p < 0.01$,$n = 369$

significance; the coefficients show as it is likely that this result is not robust and would likely change, if we chose another lag length or another in-sample period.

5.9 Granger Causality

The Granger causality test have been run on the out-of-sample data, based on the lag length chosen in-sample; this is done to see if the variables in our model are useful at predicting each other. Looking at the test we see that all variables are useful in predicting at least one other variable except SPREAD and M2. We have statistical evidence towards these variables not being able to predict other variables in the out-of-sample period and that SP500 can be used to predict SPREAD and M2 at a 5% significance level. It is important to keep in mind the limitations of the test, as it is unlikely that all of the significance found is highly robust, meaning if we had changed the out-of-sample period or the lag length of the model chances are, we would find different variables with predictive power. That being said considering all variables are either useful in explaining other variables or can be explained by another variable, the use of a multivariate model compared to a univariate model seems justified.

5.10 Forecast Performance Evaluation

The authors have decided to show the different forecast performances graphically and numerically as this is to be considered the final result of the thesis. Table 15 reports the MSE for the different models tested in this thesis. The best performing model for

Table 14: Granger causality test done on standard deviations of the detrended data, all tests are done bivariate and their F-statistic with the following p-value are shown.

	F-statistic	p-value
PCE ->IP	15.980^{***}	0.000
SPREAD ->IP	0.568	0.452
UEMP ->IP	4.458^{**}	0.036
M2 ->IP	0.864	0.354
SP500 ->IP	0.215	0.644
IP ->PCE	3.358^{*}	0.068
SPREAD ->PCE	0.011	0.915
UEMP ->PCE	1.263	0.262
M2 ->PCE	2.128	0.146
SP500 ->PCE	14.671^{***}	0.000
IP ->SPREAD	4.387^{**}	0.037
PCE ->SPREAD	6.044^{**}	0.015
UEMP ->SPREAD	0.003	0.955
M2 ->SPREAD	1.207	0.273
SP500 ->SPREAD	5.313^{**}	0.022
IP ->UEMP	16.781^{***}	0.000
PCE ->UEMP	6.469^{**}	0.012
SPREAD ->UEMP	0.001	0.971
M2 ->UEMP	2.727	0.100
SP500 ->UEMP	6.447^{**}	0.012
IP ->M2	3.350^{*}	0.068
PCE ->M2	3.251^{*}	0.073
SPREAD ->M2	0.113	0.737
UEMP ->M2	0.039	0.843
SP500 ->M2	4.958^{**}	0.027
IP ->SP500	27.773^{***}	0.000
PCE ->SP500	1.103	0.295
SPREAD ->SP500	3.273^{*}	0.072
UEMP ->SP500	1.857	0.174
M2 ->SP500	3.570^{*}	0.060

$^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$,$n = 240$

all variables is the VAR model. This is interesting, as the authors might expect the univariate models to outperform the multivariate on some of the variables. It is hard to say if the difference in MSE from GARCH and EGARCH is due to chance as the models switch between predicting better than one another. The naive model is shown to outperform both GARCH models for all variables. Looking at UEMP it is clear that EGARCH does a very poor job in estimating it, even though we set $\gamma 1 = 0$. Had we let the $\gamma 1$ be optimized the results would most likely be even worse.

Table 15: MSE comparison between models rounded to nearest 7 digits

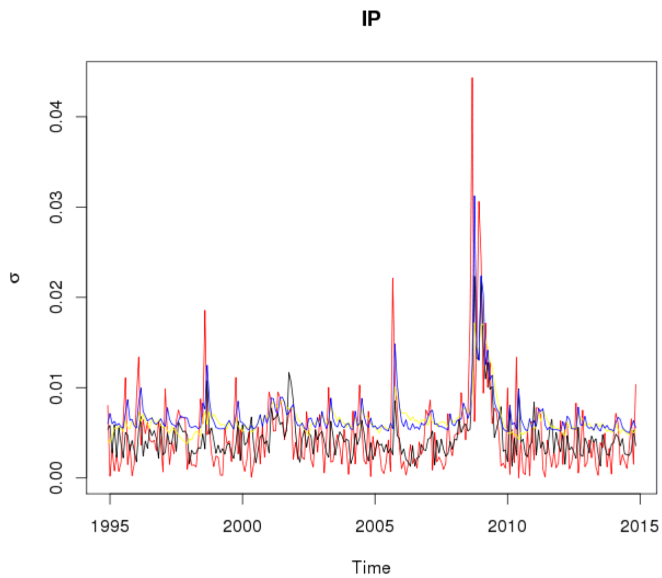
Model/Variables	IP	PCE	SPREAD	UEMP	M2	SP500
NAIVE	$0.296e^{-04}$	$0.120e^{-04}$	$0.039e^{-04}$	$0.018e^{-04}$	$0.094e^{-04}$	$1.245e^{-03}$
GARCH	$0.474e^{-04}$	$0.218e^{-04}$	$0.061e^{-04}$	$0.026e^{-04}$	$0.140e^{-04}$	$1.903e^{-03}$
EGARCH	$0.381e^{-04}$	$0.182e^{-04}$	$0.054e^{-04}$	$0.138e^{-04}$	$0.157e^{-04}$	$1.951e^{-03}$
VAR	$0.205e^{-04*}$	$0.085e^{-04*}$	$0.028e^{-04*}$	$0.011e^{-04*}$	$0.070e^{-04*}$	$0.835e^{-03*}$

*Denotes best performing model for each variable

While MSE is useful to say something about the overall difference between estimated volatility and realized volatility, it does not say whether or not the volatilities are overestimated or underestimated, which is why it is useful to look at the graphs in addition to the MSE. The authors have opted not to show the naive model in the graphs, as it would only clutter the graphs further, due to the nature of the naive model it will be identical to the realized volatility in the previous month. Figures 9 to 14 show the different model estimations of volatility and the realized volatility in the out-of-sample period.

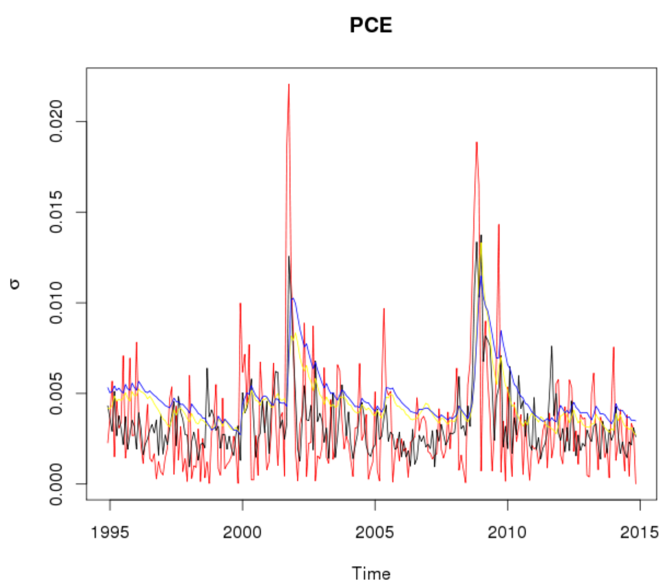
It can be seen that both GARCH and EGARCH seem to overestimate volatility compared to the realized volatility. GARCH and EGARCH seem to follow each other closely except when it comes to UEMP where EGARCH estimates significantly different from GARCH around 2000-2005. It is also interesting to note that the VAR model seem to be better at estimating the peak volatilities in addition to having a lower overall MSE. This indicates that the VAR model is superior regardless of volatility being relatively high or low in the estimated period.

Figure 9: MSE comparison between models for IP and realized volatility



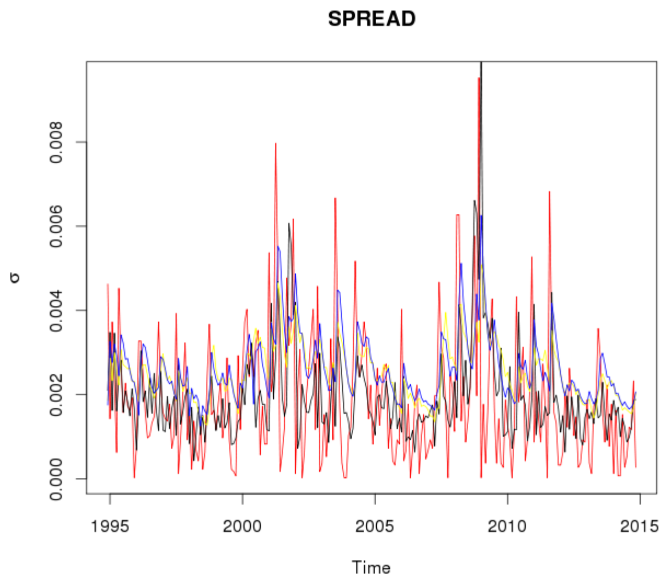
red = realized volatility, blue = GARCH, black = VAR, yellow = EGARCH

Figure 10: MSE comparison between models for PCE and realized volatility



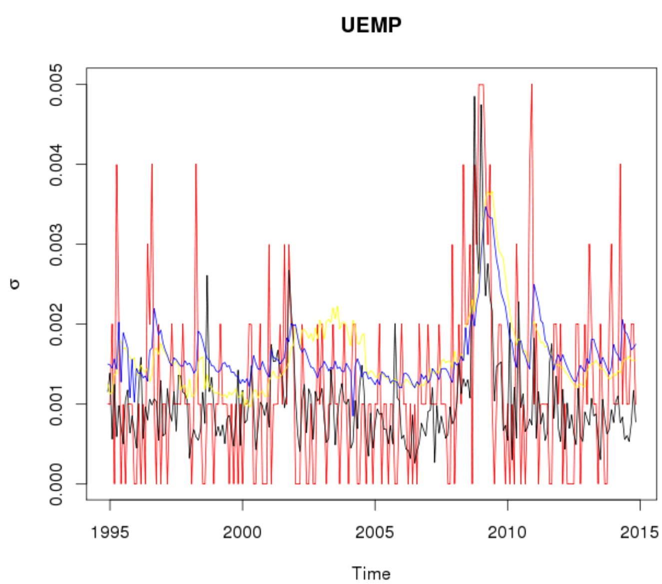
red = realized volatility, blue = GARCH, black = VAR, yellow = EGARCH

Figure 11: MSE comparison between models for SPREAD and realized volatility



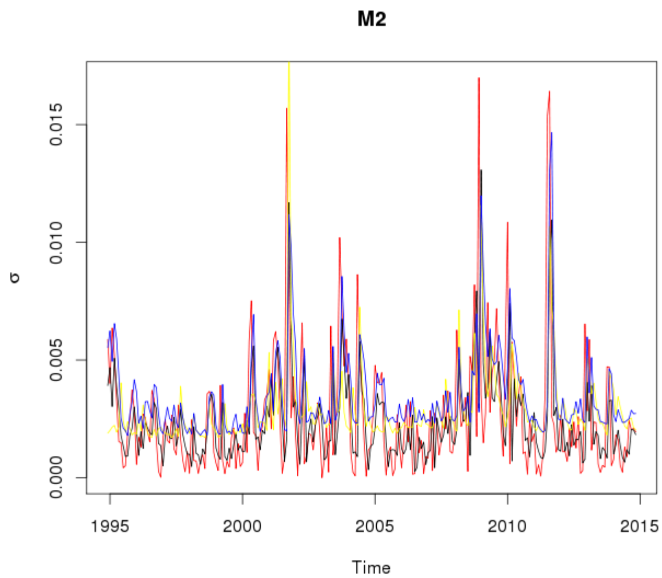
red = realized volatility, blue = GARCH, black = VAR, yellow = EGARCH

Figure 12: MSE comparison between models for UEMP and realized volatility



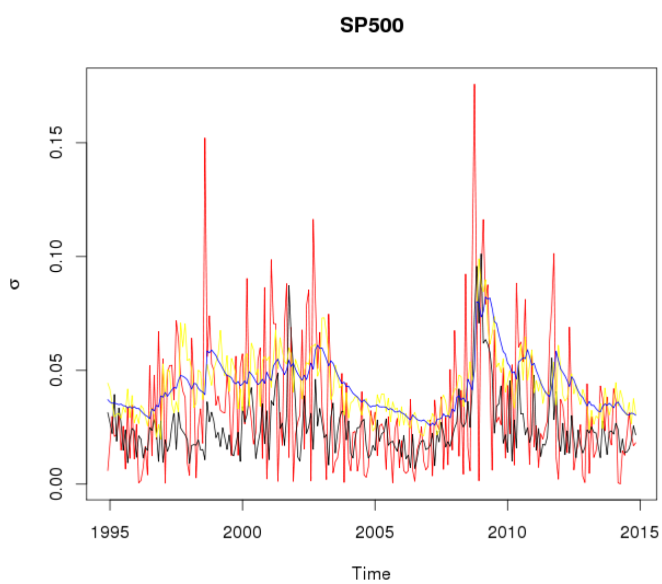
red = realized volatility, blue = GARCH, black = VAR, yellow = EGARCH

Figure 13: MSE comparison between models for M2 and realized volatility



red = realized volatility, blue = GARCH, black = VAR, yellow = EGARCH

Figure 14: MSE comparison between models for SP500 and realized volatility



red = realized volatility, blue = GARCH, black = VAR, yellow = EGARCH

6 Discussion

Table 14 shows statistical evidence of Granger causality for the variables used in the VAR model. The results show similarities with Schwert (1989), where he found that stock price volatility is better at predicting macroeconomic volatility than vice versa.

Our Granger causality tests show that lagged volatilities of SP500 can help to predict PCE, SPREAD, UEMP and M2 volatility at a 5% significance level. On the other hand, the only macroeconomic variable that helps to predict SP500 volatility at a 5% significance level is the IP volatility. Even if a variable is found not to be statistically significant, it can still be useful in predicting another variable. The significance only refer to the statistical evidence we have for the coefficient of the variable being different from 0. This is in agreement with Schwerts findings, and can be supported by economic theory, which states that assets react quickly to new information. Our results are not consistent with the research done by Beltratti and Morana (2006) and Christiansen et al. (2012), which shows evidence of macroeconomic volatility helping to predict stock market volatility.

At a 5% level, the Granger causality tests indicate that volatility of IP, PCE and SP500 helps to predict UEMP volatility. Based on economic theory, this can be explained by investments being sensitive to the inflation rate. When there is uncertainty in inflation (high volatility), the investment in the general economy tends to decline and so does the income growth. In the long term, low investment in future projects can cause the unemployment rate to increase. The results also show that UEMP volatility can help to predict IP volatility. These findings are interesting, because two variables help to predict each other. When investment fluctuates, causing the volatility of industrial production to increase, the unemployment will also fluctuate. If the unemployment rate increases, it can lead to lower industrial production. I.e. the volatility in unemployment rate and industrial production are closely related.

SPREAD volatility shows no signs of being useful in predicting the other variables. But at a 5% level, lagged variables of IP, PCE, and SP500 volatility Granger cause the

SPREAD volatility. Even though this is not in line with the conclusion of Wheelock and Wohar (2009) that the term spread can forecast output growth, it is important to keep in mind that several studies Wheelock and Wohar (2009) reviewed, found that the term spread's ability to predict output growth has been reduced since the mid-1980s.

The forecasting performance of the models is consistent with the review presented by Poon and Granger (2003), where the VAR models outperform GARCH models. Regardless of consensus with previous literature, there are some uncertainties concerning the estimation of the VAR model. This thesis does not take into account questions regarding co-integration. Having seen evidence for structural instability, the assumption was made that the variables which are not co-integrated should be challenged, and if variables are found to be co-integrated a Vector Error Correction Model (VECM) could be created to forecast future volatility. Other concerns regarding the VAR model is that estimation errors can occur because of the length of the out-of-sample period used, the wrong estimated lag lengths, or because of the sample size used in this thesis. In addition, the authors had no way of creating a suitable distribution model for UEMP as it had multiple modes. The authors are aware that this is possible to model in theory although very technical, and it would be interesting to see if it is possible to find a distribution model fitting UEMP.

Unlike Poon and Granger (2003), we find no statistical evidence of the EGARCH model outperforming the standard GARCH. This can be connected to the sample size and frequency of the data. Sadorsky (2003) suggests using daily data when calculating conditional volatility, because the daily data show more persistency than the monthly. The choice of a recursive window was made on the basis that a GARCH model, which is structurally stable, will be consistent in the sense that more observations should increase the power of the model. But as we see evidence towards it not being structurally stable, a moving window could be used instead to help alleviate the effects that this has on the estimation process.

There is also a possibility that EGARCH does not provide the best estimation for conditional volatility for all of the variables, as there are over 150 different ARCH models. In addition to test the performance between a multivariate model (VAR) and a univariate model (GARCH), an extension of this thesis would be to compare multivariate models, such as VAR and a multivariate GARCH. This method may provide a better understanding of how conditional volatility of a variable is influenced by other variables.

Periods such as the Great Inflation (1965-1982), the Great Moderation (1982-2007), Black Monday (1987) and the recent financial crisis of 2008 may have significant impact on forecasting volatility. The sample period could be divided into these sub-periods to alleviate the problems caused by the apparent structural breaks.

When comparing the estimation results, we have used MSE to evaluate performance. It would be interesting to compare different evaluation measures that separate between over and underestimation of volatility.

The use of a more flexible VAR model, such as an Asymmetric Vector AutoRegressive model (AVAR) can also be implemented. This model can contain different number of lags in the model, and is more flexible in modelling dynamic systems than a standard VAR and is helpful to eliminate variables that are not statistically significant. There are a number of macroeconomic and financial variables that work as economic indicators. Such as; M1 as a measure for money supply, the oil price or different sorts of exchange rates are leading economic indicators which can be used in volatility forecasting models for future research.

7 Conclusion

In this thesis we have predicted the volatility for 6 macroeconomic variables in the US. We found that our optimized ARIMA-EGARCH models seem to be well-specified in-sample for the most part except the model for UEMP, as it shows signs of structural instability. We found many cases of Granger causality, although the robustness of these results was questioned by the authors. When comparing our models, we found out that the VAR model outperforms all the models for all variables, while the naive model outperforms both EGARCH and GARCH. This suggests that the variables are useful in predicting each other and that the univariate GARCH framework does a poor job of volatility prediction when the data frequency is monthly. The authors find no conclusive evidence for a significant difference in the performance of the GARCH and the EGARCH.

The findings in the thesis align with previous research in that our multivariate model is superior to the univariate model. The univariate GARCH models fail to beat the naive model, which previous research indicate could be either due to the lack of external regressors or due to limited persistence in volatility clustering in monthly macroeconomic data.

8 Acknowledgment

We would like to thank Valeriy I. Zakamulin for always being available and providing extensive insight, Jochen Jungeilges for assistance with several econometric issues encountered and Alexios Ghalanos for creating Rugarch, which enabled much of the empirical work in our thesis. Any errors in this thesis should be attributed to the authors.

References

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). "Volatility and Correlation Forecasting",. Vol. 1 of *Handbook of Economic Forecasting*, pp. 777 – 878. Elsevier.

Apergis, N. and Miller, S. (2005). "Money volatility and output volatility: any asymmetric effects?", *Journal of Economic Studies*, *32*(6), 511–523.

Beltratti, A. and Morana, C. (2006). "Breaks and persistency: macroeconomic causes of stock market volatility", *Journal of Econometrics*, *131*(1–2), 151 – 177.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, *31*(3), 307–327.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

Brownlees, C., Engle, R., and Kelly, B. (2011). "A practical guide to volatility forecasting through calm and storm", *The Journal of Risk*, *14*(2), 3–22.

Christiansen, C., Schmeling, M., and Schrimpf, A. (2012). "A comprehensive look at financial volatility prediction by economic variables", *Journal of Applied Econometrics*, *27*(6), 956–977.

Christoffersen, P. F. and Diebold, F. X. (2000). "How Relevant is Volatility Forecasting for Financial Risk Management?", *The Review of Economics and Statistics*, *82*(1), pp. 12–22.

Dickey, D. A. and Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series With a Unit Root", *Journal of the American Statistical Association*, *74*(366), pp. 427–431.

Engle, R. F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, *50*(4), pp. 987–1007.

Engle, R. F., Ghysels, E., and Sohn, B. (2008). "On the Economic Sources of Stock Market Volatility",.

Engle, R. F. and Ng, V. K. (1993). "Measuring and Testing the Impact of News on Volatility", *The Journal of Finance*, *48*(5), 1749–1778.

Ewing, B. T. and Thompson, M. A. (2008). "Industrial production, volatility, and the supply chain", *International Journal of Production Economics*, *115*(2), 553 – 558. Institutional Perspectives on Supply Chain Management.

Fallahi, F., Pourtaghi, H., and Rodríguez, G. (2012). "The unemployment rate, unemployment volatility, and crime", *International Journal of Social Economics*, *39*(6), 440–448.

Fisher, T. J. and Gallagher, C. M. (2012). "New Weighted Portmanteau Statistics for Time Series Goodness of Fit Testing", *Journal of the American Statistical Association*, *107*(498), 777–787.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). "The Economic Value of Volatility Timing", *Journal of Finance*, *56*(1), 329–352.

Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", *Econometrica*, *37*(3), pp. 424–438.

Granger, C. and Newbold, P. (1974). "Spurious regressions in econometrics", *Journal of Econometrics*, *2*(2), 111 – 120.

Judson, R. and Orphanides, A. (1996). "Inflation, volatility and growth", Tech. rep..

Kearney, C. and Daly, K. (1998). "The Causes of Stock Market Volatility in Australia", *Applied Financial Economics*, *8*, 597 – 605.

Liew, V. K.-S. (2004). "Which Lag Length Selection Criteria Should We Employ?", *Economics Bulletin*, *3*(33), 1–9.

Lutkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

Lutkepohl, H. (2011). "Vector Autoregressive Models", Economics working papers ECO2011/30, European University Institute.

Marquering, W. and Verbeek, M. (2004). "The Economic Value of Predicting Stock Index Returns and Volatility", *The Journal of Financial and Quantitative Analysis*, *39*(2), pp. 407–429.

Mele, A. (2009). "Financial Volatility and Economic Activity", Fmg discussion papers dp642, Financial Markets Group.

Morelli, D. (2002). "The relationship between conditional stock market volatility and conditional macroeconomic volatility: Empirical evidence based on UK data", *International Review of Financial Analysis*, *11*(1), 101 – 110.

Mullineaux, D. J. (1980). "Unemployment, Industrial Production, and Inflation Uncertainty in the United States", *The Review of Economics and Statistics*, *62*(2), pp. 163–169.

Nelson, D. B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, *59*(2), pp. 347–370.

Ng, S. and Perron, P. (2001). "LAG Length Selection and the Construction of Unit Root Tests with Good Size and Power", *Econometrica*, *69*(6), 1519–1554.

Nyblom, J. (1989). "Testing for the Constancy of Parameters Over Time", *Journal of the American Statistical Association*, *84*(405), 223–230.

Palm, F. C. (1996). "7 GARCH models of volatility", *Handbook of Statistics*, *14*, 209,240.

Poon, S.-H. and Granger, C. W. (2003). "Forecasting Volatility in Financial Markets: A Review", *Journal of Economic Literature*, *41*(2), 478–539.

Sadorsky, P. (2003). "The macroeconomic determinants of technology stock price volatility", *Review of Financial Economics*, *12*(2), 191 – 205.

Said, S. E. and Dickey, D. A. (1984). "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order", *Biometrika*, *71*(3), pp. 599–607.

Schwert, G. (1989). "Why Does Stock Market Volatility Change over Time?", *Journal of Finance*, *44*(5), 1115–53.

Sims, C. A. (1980). "Macroeconomics and Reality", *Econometrica*, *48*(1), pp. 1–48.

Slutzky, E. (1937). “The Summation of Random Causes as the Source of Cyclic Processes”, *Econometrica*, *5*(2), pp. 105–146.

Sollis, R. (2005). “Predicting returns and volatility with macroeconomic variables: evidence from tests of encompassing”, *Journal of Forecasting*, *24*(3), 221–231.

Stock, J. H. and Watson, M. W. (2001). “Vector Autoregressions”, *Journal of Economic Perspectives*, *15*(4), 101–115.

Waggoner, D. F. and Zha, T. (1999). “Conditional Forecasts in Dynamic Multivariate Models”, *The Review of Economics and Statistics*, *81*(4), pp. 639–651.

Wheelock, D. C. and Wohar, M. E. (2009). “Can the Term Spread Predict Output Growth and Recessions? A Survey of the Literature”, *Federal Reserve Bank of St. Louis Review*.

Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Almqvist and Wiksell.

Yule, G. U. (1927). “On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer’s Sunspot Numbers”, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *226*, pp. 267–298.

9 Appendix

9.1 Stationary Processes

If non-stationary time series are used in regression modelling, it can lead to obtaining significant relationships between variables that are unrelated and is referred to as “spurious regression”. Granger and Newbold (1974) made us aware of the consequences of estimating “spurious regression” models, where the results give unlikely high R^2 values and $t-values$, which does not follow a t-distribution. Because the presence of a unit root in AR models indicate that the time series are not stationary, a number of methods have been proposed to check for this including, Dickey-Fuller (DF) test by Dickey and Fuller (1979), Augmented Dickey-Fuller (ADF) test by Said and Dickey (1984), Phillips-Perron test by Phillips and Perron in 1988 and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test by Kwiatkowski, Phillips, Schmidt and Shin in 1992.

The first test for detecting unit roots in a autoregressive processes were proposed by Dickey and Fuller (1979), and later extended by Said and Dickey (1984). The DF test, which is based on known finite orders of AR models, was augmented to permit for unknown order of finite ARMA models. This test is known as the ADF-test and together with the Phillips-Perron test is currently the most used tests for detecting unit root.

In order to make meaningful estimates, it is imperative that we are working with a stationary process; with a stationary process we are referring to a process that is covariance stationary. A process is covariance stationary iff

$$E[Y_t] = \mu, \tag{18}$$

for all t,

$$cov(Y_t, Y_{t-j}) = E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j, \tag{19}$$

for all t and any j,

where Y_t is a variable at time t. From this it follows that the mean and covariance is independent of time. To make sure our variables are stationary, the authors will examine the raw data, through inspection, to see whether there is evidence for obvious trending or cyclical behaviour. Because observation can be subject to bias, a unit root test will also be conducted.

9.2 Augmented Dickey-Fuller

The authors expect that both the visual inspection and the unit root test will come to the same conclusion with regards to stationarity, although there is evidence that the ADF test is biased, when using monthly time series. "The estimation error in the monthly volatility estimates biases the unit root estimates toward stationarity."(Schwert,1989) We test this regression in the ADF test for each time series

$$Y_t = \beta' D_t + \phi Y_{t-1} + \sum_{j=1}^{p} \lambda_j \Delta Y_{t-j} + \varepsilon_t, \tag{20}$$

where Y_t is a variable at time t, D_t includes our deterministic constants and can include a constant, or a constant and a time trend. Under the null hypothesis, $\phi = 1$, which implies the presence of a unit root while the alternative hypothesis is that $\phi = 0$. If we fail to reject the null hypothesis the data is considered non-stationary and must be differentiated, in order to make it stationary. The lag length p is set as high as needed for the error terms to be serially uncorrelated. The optimization is done by using OLS and the t-statistic is given by

$$ADF_t = t_{\phi=1} = \frac{\phi - 1}{SE(\phi)}. \tag{21}$$

When choosing lag length for the ADF test, we first choose an upper bound for the amount of lags to include in the ADF test based on the paper by Schwert (1989)

$$p_{max} = 12 \left(\frac{n}{100} \right)^{0.25}, \tag{22}$$

where n is the number of observations in the time series. Once the upper bound is set, we use the modified Akaike's information criterion (MAIC) to find the lag length that minimizes the MAIC. The reason for choosing MAIC is that although it has a tendency of overestimating the lag order on large sample sizes, it is better to overestimate than underestimate when we are testing for the presence of unit roots as shown by Ng and

Perron (2001). The MAIC is computed as

$$MAIC(l) = ln(\hat{\sigma}_p^2) + \frac{2(\tau_n(p) + p)}{n - p_{max}}, \tag{23}$$

$$\tau_n(p) = \frac{1}{\hat{\sigma}_p^2}\hat{\beta}_0^2 \sum_{t=p_{max}+1}^{n} y_{t-1}^2, \tag{24}$$

$$\hat{\sigma}_l^2 = \frac{1}{(n - p_{max})} \sum_{t=p_{max}1}^{n} +\hat{e}_{tp}^2, \tag{25}$$

where p denotes the amount of lags, n denotes the observations, β is a scalar and y_t^2 are squared observations at time t. σ_p^2 is the volatility and $\hat{\varepsilon_{tp}^2}$ is the predicted error term. To implement the ADF test we have to choose, if we want to include a trend term or not as this impacts the rejection power of the test on the null hypothesis. This choice will be done based on visual inspection of the raw data, if we observe trending or cyclical behaviour a trend term will be included. As described in the data section our raw data consists of data variables, given in both real numbers and in percentages. We, therefore, employ different transformations to the variables to make them stationary. Data in raw form, are differentiated by computing the returns,

$$dY_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}, \tag{26}$$

while for the variables that are already given in percentages the difference will be computed as

$$dY_t = Y_t - Y_{t-1}. \tag{27}$$

The reasoning behind this is one of parsimony, where the authors wish all the transformed variables to be given in percentage terms. Although the authors are aware, the variables could be co-integrated; this possibility will not be explored as it is outside the scope of this thesis and it is therefore assumed that the variables are not co-integrated.

9.3 Normal Distribution

The normal distribution is a distribution that is explained solely by its first two central moments $mean = \mu$ and $variance = \sigma^2$ with density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-0.5(x-\mu)^2}{\sigma^2}}. \tag{28}$$

This distribution is commonly used in many sciences, including economics. It is difficult to justify the use of the normal distribution in the field of economics, as you will rarely find data that do not depend on the third and the fourth moments skewness and kurtosis. The main reason for using a normal distribution used to be the fact that it is relatively easy to work with, as it requires less computation than other distributions. It is also the most commonly used distribution, which is the reason we will use it for our baseline model, with which to compare the models suggested by the authors.

9.4 Generalized Error Distribution

The density plots shown in the data section of our variables give us an indication of what kind of distribution would be suitable to model the time series. The kernel used to show the densities is the default kernel in R, the Gaussian kernel

$$K(\bar{Y};\sigma) = \frac{1}{(\sqrt{2\pi\sigma})^2} e^{-\frac{|\bar{Y}|}{2\sigma^2}}. \tag{29}$$

From the density plots it is clear that a normal distribution is unsuitable due to it showing signs of being leptokurtic; meaning it has positive excess kurtosis, this suggests we need a more flexible distribution model. We have chosen to use the GED, which is useful for modelling leptokurtic behavior. It cannot model asymmetric behavior in the density, however, there is some visual evidence from the density plots, but the authors believe it is the best fit of the available distributions.

Looking at the last density plot, it is clear we are unable to model this distribution sufficiently, as we only have unimodal distributions available and the density appears to be multimodal, meaning it has multiple modes, which is observed by the different peaking values. This gives cause for concerns, as it could severely impact the ARIMA-EGARCH estimation process. The conditional density of the GED is given by

$$f(Y) = \frac{\kappa e^{-\frac{1}{2}|(\frac{x-\alpha}{\beta})|^{\kappa}}}{2^{1+\kappa^{-1}}\beta\Gamma(\kappa^{-1})}, \tag{30}$$

α represents the location parameter. β represents the scale parameter and Γ is the Gamma function. κ represent the shape parameter.

The moments are given by

$$Mean(Y) = \alpha, \tag{31}$$

$$Var(Y) = \beta^2 2^{\frac{2}{k}} \frac{\Gamma(3\kappa^{-1})}{\Gamma(\kappa^{-1})}, \tag{32}$$

$$Skew(Y) = 0, \tag{33}$$

$$Kurt(Y) = \frac{\Gamma(5\kappa^{-1})\Gamma(\kappa^{-1})}{\Gamma(3\kappa^{-1})\Gamma(3\kappa^{-1})}. \tag{34}$$

The location parameter is equal to the first central moment, as the distribution is symmetrical and unimodal. The third central moment is 0, as the distribution is symmetric. The GED will be estimated based on our in-sample period and the estimated model will then be used when forecasting out-of-sample.

9.5 Weighted Ljung-Box

The weighted Ljung-Box test is used to check for serial correlation and was suggested by Fisher and Gallagher (2012). In this thesis it will be used both to test for serial correlation in the standardized residuals to check for ARIMA effects, and in the standardized squared residuals to test for ARCH/GARCH effects. H_0 : No serial correlation in data, if we reject the H_0, there is evidence for ARIMA or ARCH/GARCH effects not captured by our ARIMA-EGARCH model and is a sign of misspecification. The test statistic is given by

$$\widetilde{Q}_W = n(n+2)\sum_{k=1}^{m}\frac{(m-k+1)}{m}\frac{\hat{r}_k^2}{n-k}, \tag{35}$$

the test statistic is asymptotically distributed as χ^2 random variable with $p+q$ degrees of freedom. Lag lengths tested are lag 1, lag[2(p+q)+(p+q)-1] and lag[4(p+q)+(p+q)-1]. If $P-value < 0,05$ we reject H_0 indicating that there are ARIMA or ARCH/GARCH effects not captured by our model depending on which variables are tested.

9.6 Sign Bias

We will use the sign bias test of Engle and Ng (1993) to test for leverage effects in the squared residuals. This is tested by regressing squared residuals on lagged negative and positive shocks.

$$\hat{z}_t^2 = c_0 + c_1 I_{\hat{\varepsilon}_{t-1}<0} + c_2 I_{\hat{\varepsilon}_{t-1}<0}\hat{\varepsilon}_{t-1} + c_3 I_{\hat{\varepsilon}_{t-1}\geq 0}\hat{\varepsilon}_{t-1} + \mu_t. \tag{36}$$

Where $H_0 : c_i = 0\,(for\; i = 1,2,3)$,

and jointly $H_0 : c_1 = c_2 = c_3 = 0$.

I is the indicator function, $\hat{\varepsilon}_t$ the estimated residuals of the GARCH. If we observe a $P-value < 0,05$ for any of the null hypotheses, there is evidence for misspecification in the EGARCH model.

9.7 Nyblom Stability

We will use Nybloms stability test from Nyblom (1989) to test for structural stability in the in-sample period. If the variables show sign of structural instability, it indicates that we have one or more structural breaks in the data. This is tested with the Lagrange multiplier test, where the test statistics is computed as

$$\widetilde{L} = \frac{1}{n\hat{\sigma}^2} tr[V^{-1}\sum_{t=1}^{n} S_t S_t^{'}]. \tag{37}$$

$H_0 : \sigma^2_{ti} = 0$ for all i. H_0 is rejected if $\widetilde{L} > cv_{0,05}$.

σ^2_{ni} is the variance at time t for parameter i, where S_t is the cumulative first-order conditions (MLE scores) and $V = n^{-1}X'X$, where n denotes observations and X denotes a matrix of variables.

9.8 Pearson Goodness of Fit

Pearson goodness of fit test is used to compare the observed standardized residuals to what would be expected given that the chosen distribution model is correct. The test implements the suggestions made by Palm (1996), which is used to adjust for observations that are not identically independently distributed by classifying the standardized residuals by magnitude and not value. By default, the chosen numbers of columns in the histogram that are tested are 20,30,40,50.

H_0 : The data follow the given distribution. If $P - value < 0,05$ we reject H_0.

9.9 R-code

```
# install and load required packages
ipak <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg, dependencies = TRUE)
  sapply(pkg, require, character.only = TRUE)
}

packages <- c("PerformanceAnalytics","rugarch","FinTS","tsDyn",
    "zoo","urca","vars","xts","nortest","stargazer","MSBVAR","
    vcd")
```

```
ipak(packages)

# Read data
PCE.data = read.csv("~/master/Consumption.csv")
IP.data = read.csv("~/master/IP.csv")
M2.data = read.csv("~/master/M2_2.csv")
YIELD_S.data = read.csv("~/master/YIELD_S.csv")
sp500.data = read.csv("~/master/sp500.csv")
UEMP.data = read.csv("~/master/Unemployment_Rate.csv")

# Create vectors for variables
SP500 = sp500.data[,7]
SP500 = rev(SP500)

PCE = PCE.data[,2]
IP = IP.data[,2]
M2 = M2.data[,2]
SPREAD = YIELD_S.data[,4] # div by 100 for %
UEMP = UEMP.data[,2]     # div by 100 for %
raw.data = cbind(IP,M2,PCE,SP500,SPREAD,UEMP)
raw.data = ts(raw.data,start=c(1964,1),frequency=12)
# plot raw data
par(mfrow=c(3,1))
for (j in 1:ncol(raw.data)) {
  plot(raw.data[,j],main=colnames(raw.data)[j],ylab=NULL,type="
      l")
  grid(lty=3)
}
```

```r
# change SPREAD and UEMP to %
SPREAD=SPREAD/100
UEMP= UEMP/100
# make raw data stationary
nobs=nrow(UEMP.data)




Detrended.PCE = PCE[2:nobs]/PCE[1:(nobs-1)]-1
Detrended.IP = IP[2:nobs]/IP[1:(nobs-1)]-1
Detrended.M2 = M2[2:nobs]/M2[1:(nobs-1)]-1
Detrended.SP500 = SP500[2:nobs]/SP500[1:(nobs-1)]-1
Detrended.SPREAD = diff(SPREAD)
Detrended.UEMP = diff(UEMP)
#out-of-sample data matrix form
sample.matrix = ts((cbind(Detrended.IP, Detrended.M2, Detrended.
    PCE, Detrended.SP500, Detrended.SPREAD, Detrended.UEMP)), start=
    c(1964,1), frequency=12)
#plot ts transformed stationary data returns
par(mfrow=c(3,1))
for (j in 1:ncol(sample.matrix)) {
  plot(sample.matrix[,j], main=colnames(sample.matrix)[j], ylab=
     NULL)
  grid(lty=3)
 }
```

```
# Create vectors for realized standard deviation

Sd.PCE = abs(Detrended.PCE-mean(Detrended.PCE))
Sd.IP = abs(Detrended.IP-mean(Detrended.IP))
Sd.M2 = abs(Detrended.M2-mean(Detrended.M2))
Sd.SP500 = abs(Detrended.SP500-mean(Detrended.SP500))
Sd.SPREAD = abs(Detrended.SPREAD-mean(Detrended.SPREAD))
Sd.UEMP = abs(Detrended.UEMP-mean(Detrended.UEMP))



sd.matrix = ts((cbind(Sd.IP,Sd.PCE,Sd.SPREAD,Sd.UEMP,Sd.M2,Sd.
    SP500)),start=c(1964,1), frequency=12)
sd1.matrix = sd.matrix[370:609,]
#plot ts transformed stationary standard deviation
par(mfrow=c(3,1))
for (j in 1:ncol(sd.matrix)) {
  plot(sd.matrix[,j],main=colnames(sd.matrix)[j],ylab=NULL)
  grid(lty=3)
}

# Test for stationarity in returns

# Adf test for stationarity


# Sebastian Fossati
# 1/2013
```

```
ur.test = function(yt,trend,method,penalty="MAIC",kmax=NULL,
  kmin=NULL){

 # Implements unit root tests discussed in Ng, S.
 # and P. Perron (2001): "Lag Length Selection and
 # the Construction of Unit Root Tests with Good
 #     Size and Power", Econometrica, 69, 6, 1519-1554.
 #
 #     Inputs:
 #                       yt -> univariate time series
 #            trend -> the order of the polynomial trend
 #                 a) trend=c for constant
 #                 b) trend=ct for constant and linear trend
 #                       method -> unit root test to be
    constructed
 #                                a) method=adf.ols for augmented
     Dickey-Fuller test
 #                                b) method=adf.gls for Dickey-
    Fuller GLS test
 #                                c) method=pt for feasible point
     optimal test
 #            penalty -> penalty for lag length selection
 #                                a) penalty -> MAIC for MAIC
 #                                b) penalty -> AIC for AIC
 #                                c) penalty -> BIC for BIC
 #            kmax -> maximum number of lags
```

```
#            kmin -> minimum number of lags
#
#      Output:
#                        stat:   the test statistic
#                        cv:     the corresponding .01, .05, and
    .10 critical values
#               krule:  the order of the autoregression
   selected by
#                        the data dependent rule (used in
   constructing
#                        the spectral density estimator at
   frequency zero)
#                        phi:    estimate of phi

# set data as matrix!
yt = as.matrix(yt)
nt = nrow(yt)
if (is.null(kmax)){kmax = floor(12*(nt/100)^(.25))}
if (is.null(kmin)){kmin = 0}
# get deterministic components
if(trend=="c"){z = matrix(1,nr=nt); cbar = -7}
if(trend=="ct"){
  z = cbind(matrix(1,nr=nt),matrix(seq(1:nt)))
  cbar = -13.5
}
# transform the data
tmp = glsd(yt,z,cbar)
yd = tmp$yd
```

```
phi = solve(t(yd[1:(nt-1)])%*%yd[1:(nt-1)])%*%
  t(yd[1:(nt-1)])%*%yd[2:nt]
ssra = tmp$ssr
# get optimal number of lags
if(kmax==kmin){krule = kmax}
else{krule = s2ar(yd,penalty,kmax,kmin)}

# adf-ols
if(method=="adf.ols"){
  # construct adf-ols
  adf.ols = adf.ols(yt,z,krule)
  # 5% critical values
  if(trend=="c"){cv = c(-3.43,-2.86,-2.57)}
  if(trend=="ct"){cv = c(-3.96,-3.41,-3.13)}
  listss = list(method=method,stat=adf.ols$tstat,cv=cv,
                penalty=penalty,k=krule,phi=phi)
}
# adf-gls of ERS (1996)
if(method=="adf.gls"){
  # construct adf-ols
  adf.gls = adfk(yd,krule)
  # 5% critical values
  if(trend=="c"){cv = c(-2.58,-1.98,-1.62)}
  if(trend=="ct"){cv = c(-3.42,-2.91,-2.62)}
  listss = list(method=method,stat=adf.gls$tstat,cv=cv,
                penalty=penalty,k=krule,phi=phi)
}
# point optimal test of ERS (1996)
```

```
  if(method=="pt"){
    # construct adf-gls for sar
    adf.gls = adfk(yd,krule)
    sar = adf.gls$s2vec
    # construct pt
    tmp = glsd(yt,z,0)
    ssr1 = tmp$ssr
    pt = (ssra - (1+cbar/nt)*ssr1)/sar
    # 5% critical values
    if(trend=="c"){cv = c(1.78,3.17,4.45)}
    if(trend=="ct"){cv = c(4.03,5.48,6.67)}
    listss = list(method=method,stat=pt,cv=cv,
                      penalty=penalty,k=krule,phi=phi)
  }

  return( listss )

}

print.ur.test = function(test){
  method = test$method
  # print results
  cat(sprintf("\n Test for Unit Root: "))
  if(method=="adf.ols"){cat(sprintf("ADF Test"))}
  if(method=="adf.gls"){cat(sprintf("ADF Test with GLS 
     Detrending"))}
  if(method=="pt"){cat(sprintf("ERS Test"))}
  cat(sprintf("\n\n Null Hypothesis: there is a unit root"))
```

```
  cat(sprintf("\n Test Statistic: %.3f", test$stat))
  cat(sprintf("\n Critical Values (.01,.05,.10): %.2f %.2f %.2f
     ",
                  test$cv[1], test$cv[2], test$cv[3]))
  cat(sprintf("\n\n Lag Order Selection Rule: %s", test$penalty)
     )
  cat(sprintf("\n Selected Lag Order: %1.0f", test$k))
  cat(sprintf("\n Estimated Coefficient: %.4f \n\n", test$phi))
}

adfk = function(yt, kstar){
  reg = lagn(yt,1)
  dyt = diffk(yt,1)
  if(kstar>0){for(i in 1:kstar){reg = cbind(reg, lagn(dyt, i))}}
  dyt = trimr(dyt, kstar+1,0)
  reg = trimr(reg, kstar+1,0)
  b = solve(t(reg)%*%reg)%*%t(reg)%*%dyt
  ee = dyt - reg%*%b
  nef = nrow(dyt)-kstar-1
  s2e = t(ee)%*%ee/nef
  xx = solve(t(reg)%*%reg)
  sre = xx[1,1]*s2e
  sumb = 0
  if(kstar>0){sumb = sum(b[2:(kstar+1)])}
  return(list(tstat=b[1]/sqrt(sre),rho=b[1]+1,s2vec=s2e/(1-sumb
     )^2))
}
```

```
adf.ols = function(yt,z,kstar){
  reg = cbind(lagn(yt,1),z)
  dyt = diffk(yt,1)
  if(kstar>0){for(i in 1:kstar){reg = cbind(reg,lagn(dyt,i))}}
  dyt = trimr(dyt,kstar+1,0)
  reg = trimr(reg,kstar+1,0)
  b = solve(t(reg)%*%reg)%*%t(reg)%*%dyt
  ee = dyt - reg%*%b
  nef = nrow(dyt)-kstar-1-ncol(z)
  s2e = t(ee)%*%ee/nef
  xx = solve(t(reg)%*%reg)
  sre = xx[1,1]*s2e
  return(list(tstat=b[1]/sqrt(sre)))
}

s2ar = function(yt,penalty,kmax,kmin){
  tau = rep(0,kmax+1)
  s2e = 999999*rep(1,kmax+1)
  dyt = diffk(yt,1)
  reg = lagn(yt,1)
  if(kmax>0){for(i in 1:kmax){reg = cbind(reg,lagn(dyt,i))}}
  dyt = trimr(dyt,kmax+1,0)
  reg = trimr(reg,kmax+1,0)
  sumy = t(reg[,1])%*%reg[,1]
  nef = nrow(dyt)
  for(k in kmin:kmax){
    b = solve(t(reg[,1:(k+1)])%*%reg[,1:(k+1)])%*%
      t(reg[,1:(k+1)])%*%dyt
```

```
    e = dyt - reg[,1:(k+1)]%*%b
    s2e[k+1] = t(e)%*%e/nef
    tau[k+1] = b[1]*b[1]*sumy/s2e[k+1]
  }
  kk = seq(0,kmax)
  if(penalty=="AIC"){ic = log(s2e)+2*kk/nef}
  else if(penalty=="BIC"){ic = log(s2e)+log(nef)*kk/nef}
  else {ic = log(s2e)+2*(kk+tau)/nef}
  tmp = which.min(ic)
  return(tmp-1)
}

glsd = function(yt,z,cbar){
  nt = nrow(yt)
  abar = 1 + cbar/nt
  ya = matrix(0,nt,1)
  za = matrix(0,nt,ncol(z))
  ya[1,1] = yt[1,1]
  za[1,] = z[1,]
  ya[2:nt,1] = yt[2:nt,1]-abar*yt[1:(nt-1),1]
  za[2:nt,] = z[2:nt,]-abar*z[1:(nt-1),]
  # construct gls detrended series
  bhat = solve(t(za)%*%za)%*%t(za)%*%ya
  yd = yt - z%*%bhat
  ssr = t(ya-za%*%bhat)%*%(ya-za%*%bhat)
  return(list(yd=yd,ssr=ssr))
}
```

```
olsd = function(yt,z){
  # construct ols detrended series
  bhat = solve(t(z)%*%z)%*%t(z)%*%yt
  yd = yt - z%*%bhat
  return(yd=yd)
}

diffk = function(x,k){
  if(k==0){return(x)}
  else{return(rbind(matrix(0,k,ncol(x)),diff(x,lag=k)))}
}

lagn = function(x,n){
  if(n<0){return(rbind(trimr(x,abs(n),0),
                          matrix(0,abs(n),ncol(x))))}
  else{return(rbind(matrix(0,n,ncol(x)),trimr(x,0,n)))}
}

trimr = function(x,n1,n2){
  return(as.matrix(x[(n1+1):(nrow(x)-n2),],nc=ncol(x)))
}
# adf test raw data
raw.PCEtest = ur.test(PCE,trend="ct",method="adf.ols",kmax=NULL
    )
raw.print.ur.test(raw.PCEtest) # stationary 1%
raw.IPtest = ur.test(IP,trend="ct",method="adf.ols",kmax=NULL)
print.ur.test(raw.IPtest) # stationary 1%
raw.SP500test = ur.test(SP500,trend="ct", method="adf.ols",kmax
```

```
    =NULL)
print.ur.test(raw.SP500test) # Stationary 1%
raw.SPREADtest = ur.test(SPREAD, trend="ct",method="adf.ols",
    kmax=NULL)
print.ur.test(raw.SPREADtest)    # stationary 1%
raw.M2test = ur.test(M2, trend="ct",method="adf.ols",kmax=NULL)
print.ur.test(raw.M2test) # stationary 5%
raw.UEMPtest = ur.test(UEMP, trend="ct",method="adf.ols",kmax=
    NULL)
print.ur.test(raw.UEMPtest)# stationary  1%



#adf test returns
PCEtest = ur.test(Detrended.PCE, trend="ct",method="adf.ols",
    kmax=NULL)
print.ur.test(PCEtest) # stationary 1%
IPtest = ur.test(Detrended.IP, trend="c",method="adf.ols",kmax=
    NULL)
print.ur.test(IPtest) # stationary 1%
SP500test = ur.test(Detrended.SP500, trend="c", method="adf.ols"
    ,kmax=NULL)
print.ur.test(SP500test) # Stationary 1%
SPREADtest = ur.test(Detrended.SPREAD, trend="c",method="adf.ols
    ",kmax=NULL)
print.ur.test(SPREADtest)    # stationary 1%
M2test = ur.test(Detrended.M2, trend="ct",method="adf.ols",kmax=
    NULL)
```

```
print.ur.test(M2test) # stationary 5%
UEMPtest = ur.test(Detrended.UEMP, trend="c",method="adf.ols",
    kmax=NULL)
print.ur.test(UEMPtest)# stationary  1%

#adf test sd
PCE1test = ur.test(sd.matrix[,1],trend="c",method="adf.ols",
    kmax=NULL)
print.ur.test(PCE1test) # stationary 1%
IP1test = ur.test(sd.matrix[,2],trend="c",method="adf.ols",kmax
    =NULL)
print.ur.test(IP1test) # stationary 1%
SP5001test = ur.test(sd.matrix[,3],trend="c", method="adf.ols",
    kmax=NULL)
print.ur.test(SP5001test) # Stationary 1%
SPREAD1test = ur.test(sd.matrix[,4],trend="c",method="adf.ols",
    kmax=NULL)
print.ur.test(SPREAD1test)   # stationary 1%
M21test = ur.test(sd.matrix[,5],trend="c",method="adf.ols",kmax
    =NULL)
print.ur.test(M21test) # stationary 5%
UEMP1test = ur.test(sd.matrix[,6],trend="c",method="adf.ols",
    kmax=NULL)
print.ur.test(UEMP1test)# stationary  1%

# compute excess kurtosis, skewness and plot density of
    stationary data
par(mfrow=c(3,1))
```

```
Ex.Kurt = c(rep(0,6))
Skew = c(rep(0,6))
Mean = c(rep(0,6))
Sd = c(rep(0,6))
for (j in 1:ncol(sample.matrix)) {
Mean[j] = mean(sample.matrix[,j])
Sd[j] = sd(sample.matrix[,j])
Skew[j] = skewness(sample.matrix[,j])
Ex.Kurt[j] = kurtosis(sample.matrix[,j])
Moments = cbind(Mean,Sd,Skew,Ex.Kurt,rownames=colnames(raw.data
    ))
plot(density(sample.matrix[,j]),main=colnames(raw.data)[j])
}
rownames(Moments) = colnames(sample.matrix)
stargazer(Moments, digits=3)

rownames(Moments) = colnames(sample.matrix)
# test for mean model order used to find variance model based
    on AIC, highest mean model allowed 2,2

arima.IP= autoarfima(data=sample.matrix[1:(nobs-241),1], ar.max
     = 2, ma.max = 2, criterion = "HQIC",
           method = "partial", include.mean = TRUE,
           distribution.model = "ged",
           solver = "solnp",
           return.all = FALSE)
arima.M2= autoarfima(data=sample.matrix[1:(nobs-241),2], ar.max
     = 2, ma.max = 2, criterion = "HQIC",
```

```
                  method = "partial", include.mean = TRUE,
                  distribution.model = "ged",
                  solver = "solnp",
                  return.all = FALSE)
arima.PCE= autoarfima(data=sample.matrix[1:(nobs-241),3], ar.
   max = 2, ma.max = 2, criterion = "HQIC",
                  method = "partial", include.mean = TRUE,
                  distribution.model = "ged",
                  solver = "solnp",
                  return.all = FALSE)
arima.SP500= autoarfima(data=sample.matrix[1:(nobs-241),4], ar.
   max = 2, ma.max = 2, criterion = "HQIC",
                  method = "partial", include.mean = FALSE,
                  distribution.model = "ged",
                  solver = "solnp",
                  return.all = FALSE)
arima.SPREAD= autoarfima(data=sample.matrix[1:(nobs-241),5], ar
   .max = 2, ma.max = 2, criterion = "HQIC",
                  method = "partial", include.mean = FALSE,
                  distribution.model = "ged",
                  solver = "solnp",
                  return.all = FALSE)
arima.UEMP= autoarfima(data=sample.matrix[1:(nobs-241),6], ar.
   max = 2, ma.max = 2, criterion = "HQIC",
                  method = "partial", include.mean = FALSE,
                  distribution.model = "ged",
                  solver = "solnp",
                  return.all = FALSE)
```

```
arima.coef = cbind(coef(arima.IP$fit),coef(arima.M2$fit),coef(
    arima.PCE$fit),coef(arima.SP500$fit),coef(arima.SPREAD$fit),
    coef(arima.UEMP$fit))
colnames(arima.coef) = colnames(sample.matrix) # remember that
    colnames here will be wrong when importing to latex due to
    different arima terms

#test to find optimal garch model with parameters ranging from
    0-2, optimal model chosen based on AIC

garch.IP.spec = ugarchspec(distribution="ged",variance.model =
    list(model = "eGARCH",garchOrder=c(1,1)), mean.model = list(
    armaOrder=c(1,1),arfima=FALSE, include.mean=TRUE))

garch.IP.fit = ugarchfit(spec=garch.IP.spec, data=Detrended.IP,
     solver="hybrid",out.sample=240)

class(garch.IP.fit)
slotNames(garch.IP.fit)
names(garch.IP.fit@fit)
names(garch.IP.fit@model)

# show garch fit

garch.IP.fit

# Estimate M2 garch (1,1) arma(2,2) im | Hannan-Quinn -9.3325
```

```
garch.M2.spec = ugarchspec(distribution="ged",variance.model =
    list(model = "eGARCH",garchOrder=c(1,1)), mean.model = list(
    armaOrder=c(2,2),arfima=FALSE, include.mean=TRUE))

garch.M2.fit = ugarchfit(spec=garch.M2.spec, data=Detrended.M2,
      solver="hybrid",out.sample=240)

class(garch.M2.fit)
slotNames(garch.M2.fit)
names(garch.M2.fit@fit)
names(garch.M2.fit@model)

# show garch fit
garch.M2.fit



# Estimate PCE garch (1,1) arma(1,2) im | Hannan-Quinn -7.5103

garch.PCE.spec = ugarchspec(distribution="ged",variance.model =
      list(model = "eGARCH",garchOrder=c(1,1)), mean.model = list
    (armaOrder=c(1,2),arfima=FALSE, include.mean=TRUE))

garch.PCE.fit = ugarchfit(spec=garch.PCE.spec, data=Detrended.
    PCE, solver="hybrid",out.sample=240)

class(garch.PCE.fit)
slotNames(garch.PCE.fit)
names(garch.PCE.fit@fit)
```

```
names(garch.PCE.fit@model)

# show garch fit
garch.PCE.fit


# Estimate SP500 garch (2,2) arma (1,1) im | Hannan–Quinn
    −3.5337

garch.SP500.spec = ugarchspec(distribution="ged",variance.model
    = list(model = "eGARCH",garchOrder=c(2,2)), mean.model =
    list(armaOrder=c(0,1),arfima=FALSE, include.mean=FALSE))

garch.SP500.fit = ugarchfit(spec=garch.SP500.spec, data=
    Detrended.SP500, solver="hybrid",out.sample=240)

class(garch.SP500.fit)
slotNames(garch.SP500.fit)
names(garch.SP500.fit@fit)
names(garch.SP500.fit@model)

# show garch fit
garch.SP500.fit


# Estimate SPREAD garch (1,1) arma (0,2) | Hannan–Quinn −8.8185

garch.SPREAD.spec = ugarchspec(distribution="ged",variance.
```

```
model = list(model = "eGARCH",garchOrder=c(1,1)), mean.model
    = list(armaOrder=c(0,2),arfima=FALSE, include.mean=FALSE))

garch.SPREAD.fit = ugarchfit(spec=garch.SPREAD.spec, data=
    Detrended.SPREAD, solver="hybrid",out.sample=240)

class(garch.SPREAD.fit)
slotNames(garch.SPREAD.fit)
names(garch.SPREAD.fit@fit)
names(garch.SPREAD.fit@model)

# show garch fit
garch.UEMP.fit

# Estimate UEMP garch (1,1) arma (0,1) | Hannan-Quinn -13.467

garch.UEMP.spec = ugarchspec(distribution="ged",variance.model
    = list(model = "eGARCH",garchOrder=c(1,1)), mean.model =
    list(armaOrder=c(0,1),arfima=FALSE, include.mean=FALSE),
    fixed.pars=list(gamma1=0))

garch.UEMP.fit = ugarchfit(spec=garch.UEMP.spec, data=Detrended
    .UEMP, solver="hybrid",out.sample=240)

class(garch.UEMP.fit)
slotNames(garch.UEMP.fit)
names(garch.UEMP.fit@fit)
names(garch.UEMP.fit@model)
```

```
# show garch fit
garch.UEMP.fit


# pearson goodness of fit ARIMA-EGARCH
gof.IP = gof(garch.IP.fit, c(20,30,40,50))
gof.M2 = gof(garch.M2.fit, c(20,30,40,50))
gof.PCE = gof(garch.PCE.fit, c(20,30,40,50))
gof.SP500 = gof(garch.SP500.fit, c(20,30,40,50))
gof.SPREAD = gof(garch.SPREAD.fit, c(20,30,40,50))
gof.UEMP = gof(garch.UEMP.fit, c(20,30,40,50))

# sign bias ARIMA-EGARCH
sign.IP = signbias(garch.IP.fit)[,1:2]
sign.M2 =signbias(garch.M2.fit)[,1:2]
sign.PCE = signbias(garch.PCE.fit)[,1:2]
sign.SP500 =signbias(garch.SP500.fit)[,1:2]
sign.SPREAD =signbias(garch.SPREAD.fit)[,1:2]
sign.UEMP = signbias(garch.UEMP.fit)[,1:2]

#GARCH(1,1) spec

# Estimate GARCH(1,1) as a baseline with mean model : arma 0,0
  IM

garch11.spec = ugarchspec(distribution="norm",variance.model =
  list(model = "sGARCH",garchOrder=c(1,1)), mean.model = list(
```

```
    armaOrder=c(0,0),arfima=FALSE, include.mean=TRUE))

garch11.M2.fit = ugarchfit(spec=garch11.spec, data=Detrended.M2
    , solver="hybrid")
garch11.IP.fit = ugarchfit(spec=garch11.spec, data=Detrended.IP
    , solver="hybrid")
garch11.PCE.fit = ugarchfit(spec=garch11.spec, data=Detrended.
    PCE, solver="hybrid")
garch11.SP500.fit = ugarchfit(spec=garch11.spec, data=Detrended
    .SP500, solver="hybrid")
garch11.SPREAD.fit = ugarchfit(spec=garch11.spec, data=SPREAD,
    solver="hybrid")
garch11.UEMP.fit = ugarchfit(spec=garch11.spec, data=UEMP,
    solver="hybrid")

# Garch forecasts based on optimal Garch specs for univariate
    garch models
IP.roll=ugarchroll(spec=garch.IP.spec, data=sample.matrix[,1],
    n.ahead = 1, forecast.length=240,
            , refit.every = 1, refit.window = c("recursive")
        , solver = "hybrid")

M2.roll=ugarchroll(spec=garch.M2.spec, data=sample.matrix[,2],
    n.ahead = 1, forecast.length=240,
              , refit.every = 1, refit.window = c("recursive")
              , solver = "hybrid")

PCE.roll=ugarchroll(spec=garch.PCE.spec, data=sample.matrix
```

```
    [,3], n.ahead = 1, forecast.length=240,
                  , refit.every = 1, refit.window = c("recursive")
                  , solver = "hybrid")

SP500.roll=ugarchroll(spec=garch.SP500.spec, data=sample.matrix
    [,4], n.ahead = 1, forecast.length=240,
                  , refit.every = 1, refit.window = c("recursive")
                  , solver = "hybrid")

SPREAD.roll=ugarchroll(spec=garch.SPREAD.spec, data=sample.
    matrix[,5], n.ahead = 1, forecast.length=240,
                  , refit.every = 1, refit.window = c("recursive")
                  , solver = "hybrid")

UEMP.roll=ugarchroll(spec=garch.UEMP.spec, data=sample.matrix
    [,6], n.ahead = 1, forecast.length=240,
                  , refit.every = 1, refit.window = c("recursive")
                  , solver = "hybrid")
pred.opt.garch = cbind(IP.roll,M2.roll,PCE.roll,SP500.roll,
    SPREAD.roll,UEMP.roll)


#GARCH roll 11
IP.roll11=ugarchroll(spec=garch11.spec, data=sample.matrix[,1],
     n.ahead = 1, forecast.length=240
                        , refit.every = 1, refit.window = c("
                           recursive")
                        , solver = "hybrid")
```

```
M2.roll11=ugarchroll(spec=garch11.spec, data=sample.matrix[,2],
    n.ahead = 1, forecast.length=240
                        , refit.every = 1, refit.window = c("
                          recursive")
                        , solver = "hybrid")

PCE.roll11=ugarchroll(spec=garch11.spec, data=sample.matrix
    [,3], n.ahead = 1, forecast.length=240
                        , refit.every = 1, refit.window = c("
                          recursive")
                        , solver = "hybrid")

SP500.roll11=ugarchroll(spec=garch11.spec, data=sample.matrix
    [,4], n.ahead = 1, forecast.length=240
                        , refit.every = 1, refit.window = c("
                          recursive")
                        , solver = "hybrid")

SPREAD.roll11=ugarchroll(spec=garch11.spec, data=sample.matrix
    [,5], n.ahead = 1, forecast.length=240
                        , refit.every = 1, refit.window = c("
                          recursive")
                        , solver = "hybrid")

UEMP.roll11=ugarchroll(spec=garch11.spec, data=sample.matrix
    [,6], n.ahead = 1, forecast.length=240
                        , refit.every = 1, refit.window = c("
```

```
                recursive")
              , solver = "hybrid")
pred.11.garch = cbind(IP.roll11,M2.roll11,PCE.roll11,SP500.
    roll11,SPREAD.roll11,UEMP.roll11)

# select optimal lag order VAR based on first HQ = 1 then HQ =
    2
VARselect(sd.matrix[1:370,])
stargazer(VARselect(sd.matrix[1:370,]),digits=1)

# estimate var model in-sample first to get coefficients

var.in.sample = VAR(sd.matrix[1:370,], p=1, type = "none")
# estimate var model whole for rolling forecast
var = lineVar(sd.matrix, lag=1, include = "none",
              model = "VAR", I = "level",
              estim = "ML")
stargazer(var.in.sample,)
## predict standard deviation using vecm and var
pred.var= predict_rolling(var,n.ahead=1, nroll=240, refit.every
    =1)
pred.sig.var=ts(pred.var$pred,end=c(2014,11),frequency=12)
true.sig=ts(pred.var$true,end=c(2014,11),frequency=12)

summary(var.in.sample)
# plot
par(mfrow=c(1,2))
for (j in 1:ncol(sd.matrix)) {
```

```
plot((pred.sig.var[,j]), col="black", type="s",main=colnames(sd
    .matrix[,j]),ylab=expression(sigma),cex.lab=1.5)
lines((true.sig[,j]), col="red", type="s",main=colnames(sd.
    matrix[,j]),ylab=expression(sigma),cex.lab=1.5)
legend(x="topleft",c("pred","true"),col=c("black","red"),lty=1,
    cex=1)
}

# Test for granger causality
granger.test(sd.matrix[370:610,],p=1)

# out-of-sample period
Realised.sd = sd.matrix[371:610,]

# calculate MSPE
MSE.naive = t((colSums((sd1.matrix-true.sig)^2))/240)
MSE.var = t((colSums((pred.sig.var-true.sig)^2))/240)
MSE.sgarch = cbind(fpm(IP.roll11),fpm(PCE.roll11),fpm(SPREAD.
    roll11),fpm(UEMP.roll11),fpm(M2.roll11),fpm(SP500.roll11))
MSE.egarch = cbind(fpm(IP.roll),fpm(PCE.roll),fpm(SPREAD.roll),
    fpm(UEMP.roll),fpm(M2.roll),fpm(SP500.roll))
MSE.egarch = MSE.egarch[1,]
MSE.sgarch = MSE.sgarch[1,]
colnames(MSE.egarch) = c("IP","PCE","SPREAD","UEMP","M2","SP500
    ")
colnames(MSE.sgarch) = c("IP","PCE","SPREAD","UEMP","M2","SP500
    ")
colnames(MSE.var) = c("IP","PCE","SPREAD","UEMP","M2","SP500")
```

```
colnames(MSE.naive) = c("IP","PCE","SPREAD","UEMP","M2","SP500"
    )
# matrix for comparing MSE
options(scipen=6)
MSE.matrix = rbind(MSE.naive,MSE.sgarch,MSE.egarch,MSE.var)
rownames(MSE.matrix)= c("naive","sGARCH","eGARCH","VAR")
# plot garch, var and realized data
eGarch.estimates = cbind(as.data.frame(IP.roll)[,2],as.data.
    frame(PCE.roll)[,2],as.data.frame(SPREAD.roll)[,2],as.data.
    frame(UEMP1.roll)[,2],as.data.frame(M2.roll)[,2],as.data.
    frame(SP500.roll)[,2])
eGarch.estimates = ts(eGarch.estimates,frequency=12,end=c
    (2014,11))
colnames(Garch.estimates) = colnames(MSE.egarch)
sGarch.estimates = cbind(as.data.frame(IP.roll11)[,2],as.data.
    frame(PCE.roll11)[,2],as.data.frame(SPREAD.roll11)[,2],as.
    data.frame(UEMP.roll11)[,2],as.data.frame(M2.roll11)[,2],as.
    data.frame(SP500.roll11)[,2])
sGarch.estimates = ts(sGarch.estimates,frequency=12,end=c
    (2014,11))
naive.estimates = ts(sd1.matrix,frequency=12,end=c(2014,11))
colnames(Garch.estimates) = colnames(MSE.sgarch)
par(mfrow=c(1,1))
for (j in 1:6) {
plot(ts(true.sig[,j],frequency=12,end=c(2014,11)),col="red",
    ylab=expression(sigma),main=colnames(MSE.garch)[j])
lines(ts(pred.sig.var[,j],frequency=12,end=c(2014,11)),col="
    black")
```

```
lines(eGarch.estimates[,j],col="yellow1")
lines(sGarch.estimates[,j],col="blue")
}
MSE.matrix
```