A-MF-EPiC-GARCH:

Comparing the predictive power of an Asymmetric Mixed-Frequency Econometric Principal Component GARCH and other GARCH-type models in R and Python environments

Lauri Teerimäki & Frans Mäkelä

Department: Department of Finande and Economics	**Type of work**: Term paper
Authors: Lauri Teerimäki & Frans Mäkelä	**Date**: 9.11.2023

Title of paper:
A-MF-EPiC-GARCH: Comparing the predictive power of an Asymmetric Mixed-Frequency Econometric Principal Component GARCH and other GARCH-type models in R and Python environments

Abstract:

In this paper, we evaluate the forecasting performance of various first-order GARCH model extensions, in R and Python software environments, including a modern asymmetric mixed frequency component GARCH, drawing on methods of Asgharian, Hou, & Javed (2013) and Condrad & Kleen (2020). **Combining the Mixed Data Sampling (MIDAS) framework by Engle, Ghysels & Sohn (2013) and PCA as a dimensionality reduction technique, we integrate lower frequency principal components of a selection of macroeconomic variables as covariates in the asymmetric GARCH-model.**

Our study spans from 2006 to 2023, including significant market events, and employs a mixed-frequency dataset comprising weekly S&P 500 returns and RV and monthly macroeconomic variable observations. We assess the GARCH, EGARCH, and GJR-GARCH models using the `rugarch` and `arch` packages, investigating the impact of asymmetric innovations and the default-setting-optimisers on model fit and out-of-sample forecasting performance.

We use out-of-sample-**R^2, Mean Directional Accuracy, MAPE and RMSPE** to evaluate the models, and employ a **Diebold Mariano test** to introduce statistical significance testing of forecasting performance differences, introducing a method to validate model superiority by proxy.

Our analysis challenges the conventional emphasis on software differences in GARCH modelling, redirecting focus to the estimation method as one of the key determinants of model accuracy and highlihts, that for identical models, any fitted differences can largely be attributed to differing estimation methods. The humorously named EPiC-GARCH hopefully offers a new perspective on volatility forecasting, and is intended to work as a stepping stone for the emergence of a standardised framework for PCA-enabled mixed-frequency models in volatility forecasting.

Keywords: Volatility forecasting, GARCH, MIDAS, PCA, MLE

TABLE OF CONTENTS

1 INTRODUCTION

Volatility is an elusive concept, often associated with or defined as the uncertainty of an outcome or the propensity for rapid change of state. Inherently, volatility is an unobservable latent force, that can only be inferred through its effects. There is a large body of literature dedicated to the attempt of quantifying, modelling, and forecasting volatility, from which one can roughly extract two different classes of models: i) the discrete, autoregressive models that compute volatility as a function conditional of an observable information set, and ii) the latent models that compute volatility as a function including unobservable qualities.

In this paper we will compare the out-of-sample forecasting performance of models belonging to the first class, more specifically, extensions to the Generalised Autoregressive Conditional Heteroscedasticity (GARCH) model. The GARCH was proposed by Bollerslev (1986) as a generalisation of his mentor's and field pioneer Engle's ARCH model (1982). Several extensions to the GARCH class have been proposed since, most notably the EGARCH by Nelson (1991) and the GJR-GARCH by Glosten, Jagannathan & Runkle (1993) to address the asymmetric volatility phenomenon, also known as the leverage effect (discussed in the following section).

Connections between volatility and macroeconomic variables have also been extensively studied, with research showing how various economic indicators can impact market volatility. For example, Schwert (1989) showed that bond returns, short-term interest rates, producer prices, and the growth rate of industrial production have explanatory power in monthly market volatility. As a result of the many proposed component models in 1990's, notably by Ding & Granger (1996) and later by Engle & Lee (1999) that split the conditional volatility into long- and short-term components, the insights of e.g., Schwert came to inspire a macro-enabled component GARCH by Engle, Ghysels & Sohn (2013). In this model, a mixed data sampling (MIDAS) technique introduced by Ghysels, Santa-Clara, & Valkanov (2005) is used to reconcile the high frequency volatility process (short-term component) with a lower-frequency macroeconomic variable (long-term component). The GARCH-MIDAS model class has since been further developed by many researchers, for example by Asgharian, Hou, & Javed (2013), who proposed using principal components (PCA) as a dimensionality reduction exercise to capture several macroeconomic variables into one regressor and Condrad & Kleen (2020) where an asymmetric innovation in the spirit of GJR is included in the GARCH-MIDAS model.

The technical implementation of GARCH-type models is done with various software. Among the most commonly used are the R package `rugarch` (Galanos, 2023), Ox (G@RCH), Eviews, Matlab and Python package `arch` (Sheppard, 2021). Charles & Darné (2019), demonstrated that asymmetric GARCH model fitting is highly dependent on the software used, and highlight the need for detailed reporting of software use in research for better comparability and reliability of results.

We concur with the need for detailed reporting on model construction but argue that that comparing different software packages for GARCH modelling is essentially a red herring. Aside from computer engineering issues like differences in binary floating-point arithmetic operations and computer hardware, the differences in model fit results between different software packages, given identical mean and variance equations, is *completely* contingent on the maximum likelihood estimation (MLE) method. Since there are no closed-form solutions to the MLE, the use of numeric methods is necessary. Different non-linear optimisation algorithms, or different configurations for the same algorithm might therefore yield varying solutions to the MLE.

The core of the issue lies not in the software packages themselves, but in the optimiser and hyperparameter selection. Reformulating the issue from this standpoint leads to an entirely different research question: "Which optimisers (and hyperparameters) provide the most accurate MLE (for time series models)?". Analysing the topic solely on a software implementation level is therefore, in our view, a low-resolution endeavour that does not paint a complete picture.

1.1 Hypothesis formulation

In this paper, we test the model fit and out-of-sample volatility forecasting performance of three established GARCH-type models (GARCH, EGARCH, GJR-GARCH) with **default configurations** in **`rugarch`** (R) and **`arch`** (Python) and a **modern asymmetric mixed frequency component GARCH** with **principal component regressors** from a collection of macroeconomic variables (A-MF-EPiC-GARCH). Our hypotheses are:

$H_{1(0)}$:	*Asymmetric innovations (EGARCH and GJR-GARCH) provide incremental out-of-sample performance to a standard GARCH*

$H_{1(0)}$:	*Lower frequency macroeconomic variables via PCA & MIDAS provide incremental out-of-sample performance to an asymmetric GARCH*

1.2 Stylised facts about volatility

For the purposes of this paper, the most widely acknowledged stylised facts about volatility are briefly recited in the accordance with Masset's (2011) summary paper:

Horizontal dependence: Volatility is dynamic and exhibits mean-reversion, often clustering in patterns where periods of high volatility are likely to be followed by similar periods, and the same for low volatility. This characteristic is reinforced by significant autocorrelation over extended periods, creating what is called long-memory.

Leptokurtosis: The frequency of extreme returns, exceeds predictions of standard financial theories, indicating a distribution with fat tails. Additionally, market shocks tend to generate significant volatility and subsequent aftershocks, a phenomenon not fully captured by many financial models.

The Leverage Effect: There is a notable inverse relationship between volatility and returns, known as the leverage effect or asymmetric volatility phenomenon. This relationship is asymmetric; negative returns typically lead to a sharper increase in volatility compared to the more moderate decrease in volatility following positive returns.

Vertical dependence: The informational content of volatility varies with the frequency at which it is measured (e.g., intraday, daily, or monthly). Lower-frequency volatility tends to have a greater influence on subsequent higher-frequency volatility than vice versa. Moreover, increases in low-frequency volatility have a more pronounced effect on high-frequency volatility compared to decreases.

2 MODEL DESCRIPTIONS

In the previous chapter, we briefly looked at the stylised facts of volatility – notably the long memory and leverage effect. **In this section we will present the relevant model definitions in an *attempted unified notational framework*.** Most GARCH-type models are defined with a mean equation and a variance equation, along with the relevant parameter restriction equations. For simplicity and their historically demonstrated ability we will only consider first-order autoregressive terms in this paper.

For the models without exogenous covariates, we adopt a standard interpretation of the mean equation, assuming that the (log) return follows the process $r_t = \mu + \epsilon_t$, where μ represents the (time-invariant) expected return and $\epsilon_t = \sigma_t z_t$ is an innovation. In this formulation, σ_t is the conditional volatility of the return series and z_t is an independent and identically distributed (i.i.d.) standard random variable $z_t \sim N(0,1)$. In cases where $E[r_t] = 0$, the mean equation simplifies to $r_t = \epsilon_t$.

The **GARCH** (Bollerslev, 1986) conditional variance equation can be expressed as follows:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2,$$

where α and β are parameters for the previous period's innovation and the previous conditional variance respectively. The term ω can be viewed as a representation of the unconditional variance, which is given by $\sigma^2 = \frac{\omega}{1-\alpha-\beta}$. The model is constrained by $\alpha, \beta, \omega \geq 0$ to ensure that the variance is non-negative and by $\alpha + \beta < 1$, to ensure that the process is covariance stationarity does not explode.

An explicit notational form of the **EGARCH** (Nelson, 1991)[1] conditional variance equation can be defined as follows:

$$\ln \sigma_t^2 = \omega + \alpha\left(|e_{t-1}| - \sqrt{\frac{2}{\pi}}\right) + \gamma e_{t-1} + \beta \ln \sigma_{t-1}^2$$

Here, γ is added as a parameter alongside α, β and ω, as a weight for $e_{t-1} = \frac{\epsilon_{t-1}}{\sigma_{t-1}}$, the standardised lagged innovation, modelling the leverage effect by assigning a direction to the previous period's shock. The term $(|e_{t-1}| - \sqrt{2/\pi})$ models the (directionless) magnitude of the previous period's shock, where the absolute innovation (of a standard normal distribution) is mean corrected by its expectation, $E[|Z|] = \sqrt{2/\pi}$ to capture the incremental effect of a shock. The logarithmic formulation negates the non-negativity constraint and allows for negative α or γ values for different reactions to positive or negative innovations. The model is still

[1] Nelson (1991) uses a different, less explicit notation for the EGARCH framework, with a separately specified asymmetric response function.

bound by covariance stationarity[2], which in the logarithmic form is not as easily stated as a set of constraints. Due of this, the stationarity condition for EGARCH models is often assessed empirically through estimation and diagnostic checking rather than through a simple rule of thumb.

The **GJR-GARCH** (Glosten et al., 1993) conditional variance equation is defined as follows:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \gamma\epsilon_{t-1}^2 I[\epsilon_{t-1} < 0] + \beta\sigma_{t-1}^2$$

As with the EGARCH, also the GJR-model employs an asymmetric term $\gamma\epsilon_{t-1}^2$ in addition to ω, α and β as in the standard GARCH specification. The asymmetric term is activated with an indicator function $I[\epsilon_{t-1} < 0]$, which equals one if $\epsilon_{t-1} < 0$ and zero otherwise. The GJR model is a nested version of the standard GARCH, such that when $\gamma = 0$, the model simplifies to a standard GARCH. The model parameters are constrained by non-negativity: $\alpha, \beta, \omega, \gamma \geq 0$ and covariance stationarity $\alpha + \frac{\gamma}{2} + \beta < 1$, assuming shocks are symmetrically distributed (half of γ is therefore considered). This allows for a new interpretation of the unconditional variance $\sigma^2 = \frac{\omega}{1-\alpha-\frac{\gamma}{2}-\beta}$ that is also referenced in the following model.

A-MF-EPiC-GARCH

Drawing on the original proposal (Robert F Engle et al., 2013) while combining the inclusion of an asymmetric innovation (Conrad & Kleen, 2020) and macroeconomic variables via PCs (Asgharian et al., 2013) we explore the framework presented below. In this section the index t refers to a lower frequency time period and the index i refers to a higher frequency time period within the index t.

In this multiplicative interpretation, the conditional variance is modelled as the product of a long- and a short-term component denoted by τ_t and $g_{i,t}$ respectively in the following manner: $\sigma_t^2 = \tau_t g_{i,t}$. In the case of $E[r_t] = 0$ or of a demeaned return series, the mean equation $r_{i,t} = \epsilon_{i,t} = \sigma_{i,t} z_{i,t}$, substituting in $\tau_t g_{i,t}$ can be expressed as follows:$\frac{\epsilon_{i,t}}{\sqrt{\tau_t}} = \sqrt{g_{i,t}} z_{i,t}$, where and z_t is an independent and identically distributed (i.i.d.) standard random variable $z_t \sim N(0,1)$ and $\frac{\epsilon_{i,t}}{\sqrt{\tau_t}}$ is assumed to follow a GARCH process[3].

The short-term component of the conditional variance equation $\sigma_t^2 = \tau_t g_{i,t}$ is further defined as follows:

$$g_{i,t} = \left(1 - \alpha - \frac{\gamma}{2} - \beta\right) + \alpha\left(\frac{\varepsilon_{t-1}^2}{\tau_t} + \right) + I[\epsilon_{t-1} < 0]\left(\frac{\varepsilon_{i-1}^2}{\tau_t}\right) + \beta g_{i-1,t}$$

---

[2] The EGARCH can often be assumed to be stationary with most estimated parameter values due to the logarithmic form. In any case, reasonable parameter values are sought after, as extreme values, even though not necessarily violating the stationarity condition, might lead to inaccurate forecasts, such as too persisting variance series.

[3] Reconciling this with standard GARCH models is to assume that $\tau_t = 1$, which would mean that the returns themselves would follow a GARCH process.

The short-term component is a covariance stationary GJR-GARCH unit variance process with the long-term component τ_t factored out of the equation, intended capture higher frequency (weekly, in our case) clustering and asymmetry of the modelled variance series. The model parameters are subject to the same constraints as the standard GJR-GARCH, i.e., that $\alpha, \beta, \omega, \gamma \geq 0$ and $\alpha + \frac{\gamma}{2} + \beta < 1$. The model simplifies to a standard GJR-GARCH if the long-run variance is selected as a constant $\tau_t = \frac{\omega}{1-\alpha-\frac{\gamma}{2}-\beta}$.

The long-term component τ_t is intended to describe smoother movements in the conditional variance and is often defined as a lag polynomial of some lower frequency (monthly, in our case) explanatory variable(s) X_t. The explanatory variable(s) X_t are assumed to stationary and independent of lower frequency innovations. Expressing the long-term component with two covariates, applying the MIDAS technique, τ_t is defined as follows:

$$\tau_t = m + \theta_1 \sum_{k=1}^{K} \phi_k\,(w_1, w_2) X_{t-k}^1 + \theta_2 \sum_{k=1}^{K} \phi_k\,(w_1, w_2) X_{t-k}^2,$$

where $X_{t-1}^1$and X_{t-1}^2 are lagged low-frequency (monthly) covariates, m is the constant, the parameters θ_1 and θ_2 determine the sign for the effect size of $X_{t-1}^1$and X_{t-1}^2. The beta lag polynomial weighting scheme denoted as $\phi_k(w_1, w_2)$ is the selected method to assign weights to lagged explanatory variables in the MIDAS (Mixed model. This scheme is defined as follows:

$$\phi_k(w_1, w_2) = \frac{\left({}^{l}\!/_{K+1}\right)^{w_1-1}\left(1 - {}^{l}\!/_{K+1}\right)^{w_2-1}}{\sum_{j=1}^{K}\left({}^{j}\!/_{K+1}\right)^{w_1-1}\left(1 - {}^{j}\!/_{K+1}\right)^{w_{2-1}}},$$

where $\phi_k(w_1, w_2)$ is the weight assigned to its corresponding lagged variable X_{t-k}, l is the specific lag position being considered (ranging from 1 to K), for which the weight is being calculated, K is the order of lags in the model, $w_1$andw_1 are parameters determining how the weights decay for more distant lags and j is the index in the denominator summation, ensuring that the weights sum to 1 across all lags.

3 DATA AND MODEL ESTIMATION

The paper considers two primary types of financial data: daily log-transformed S&P 500 returns resampled to a weekly level, and a selection of monthly macroeconomic variables, from which the first two principal components are extracted via PCA. The time frame for our analysis extends from **February 5, 2006, to January 1, 2023**, which we have split into **a training and test set.** The test set considers the latest 19.5 % of the total data length (883), leaving 169 observations in the out-of-sample series of the dataset. **Both subsets include considerable volatility shocks**; the subprime mortgage crisis of 2007-2008 and the COVID-19-related market uncertainties of 2020.

The returns, denoted by r are calculated as follows: $r_d = ln\left(\frac{I_d}{I_{d-1}}\right)$, $r_w = \sum_{i=1}^{n} r_d$, where I stands for the adjusted closing value of the S&P 500 index, n denotes the number of trading days within observed period and the indices d and w indicate daily and weekly observation windows respectively.

The distribution of returns can be seen on the graph to the right, displaying signs of overall symmetry, albeit with slight negative skew and lepto-kurtosis. Assuming an expected value of zero for daily returns (which even seems consistent on a weekly frequency) for the entire dataset, consistent with most of the literature, we use the sum of squared returns: $\boldsymbol{RV_w = \sqrt{\sum_{d=1}^{n} r_d^2}}$**, as a proxy for realised volatility**, based on Andersen & Bollerslev (1998), to evaluate model fits and out-of-sample performance.

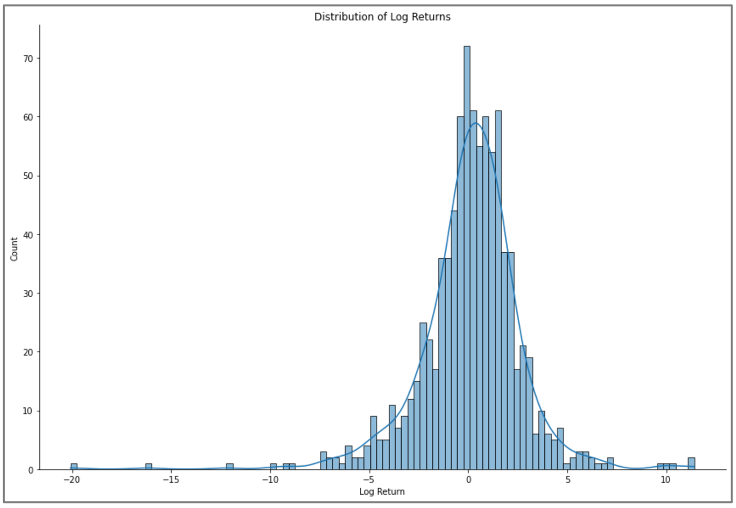

Figure 1: Distribution of Log Returns

The graph collection in figure 2 illustrates the weekly returns and the squared returns (RV) as line plots, as well as their respective autocorrelation functions. As is visible, the autocorrelation of returns is negligible, while the RV exhibits a high degree of autocorrelation. Additionally, the leverage effect is noticeable when comparing the line plots, since we see volatility spikes coinciding with large negative returns. Clustering is also visible from the line plots.

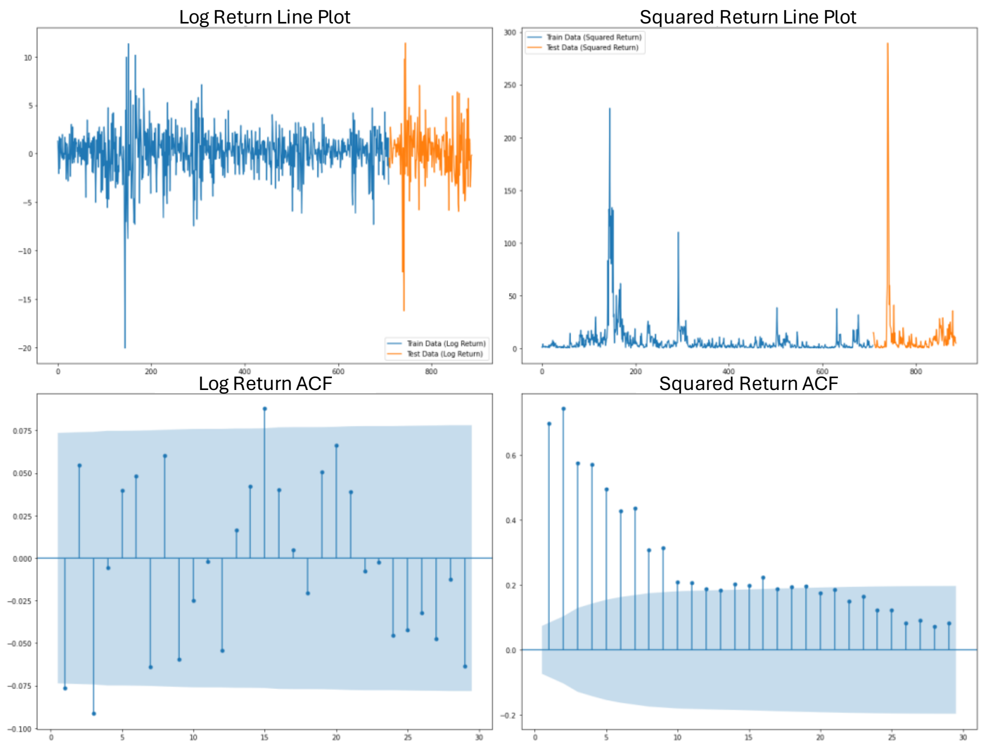

Figure 2: Split and ACF from Log Returns and Realized Volatility

For the purposes of this paper and considering the existing literature, for example Schwert (1989), **we have selected macroeconomic variables in the spirit of Asgharian et al. (2013) according to availability**:

Table 1: Macro Variables for the MIDAS Models

Variable	Note:	Source:
Yield Curve Slope	Spread between a 10-year Treasury note and a 13-week Treasury bill	Yahoo Finance
Δ REER (%)	Change in monthly USD real effective exchange rate	Bank for International Settlements
Δ Oil Price (%)	Change in monthly Crude (CL=F) price	Yahoo Finance
Δ Unemployment (%)	Change in monthly unemployment rate	S&P Capital IQ

As **this exercise focuses on extracting maximum predictive power for forecasting purposes**, we do not attempt to explain any underlying relationship, causal or otherwise, between the macroeconomic variables and volatility. Instead, our approach is limited to examining simple correlations over the in-sample period as part of the exploratory data analysis.

3.1 Principal Component Analysis and variable correlations

Continuing the application of techniques by of Asgharian et al. (2013) we perform a dimensionality reduction with PCA to extract the first two principal components as potential regressors for the GARCH-MIDAS model. The PCA process transforms the original data into a new set of orthogonal variables (PCs), which are linear combinations of the original variables. The process is as follows: Each variable in the dataset (matrix) is standardised $X_{\text{std}} = \frac{X-\mu}{\sigma}$ and the covariance matrix, $\Sigma = \frac{1}{m-1} X_{\text{std}}^{\top} X_{\text{std}}$ is computed. The eigenvalues and eigenvectors are obtained through eigen decomposition $\Sigma v_i = \lambda_i v_i$ and the top (two) eigenvectors are selected based on their corresponding eigenvalues. The linear transformation of the selected (two) eigenvectors is applied to the original dataset, creating a new two-dimensional space $Y = X_{\text{std}} V_k$ which is defined by the principal components.

Aside from reducing the number of variables in the model by transforming them into a smaller number of principal components, which simplifies model estimation and mitigates overfitting, multicollinearity is naturally avoided, since principal components are orthogonal to one another. The correlation matrix below confirms that PC1 has the highest absolute correlation (0.42) with weekly realised volatility out of all proposed macroeconomic variables, suggesting its utility as a regressor in the GARCH-MIDAS.

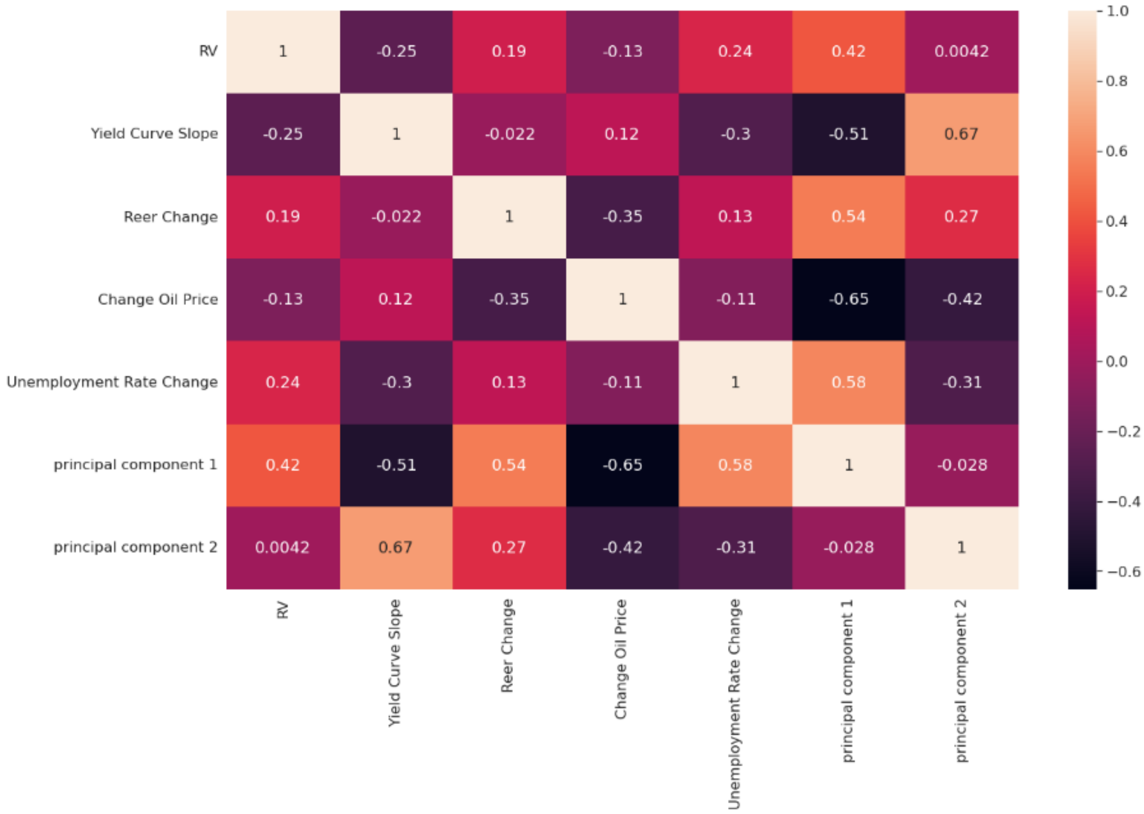

Figure 3: Principal Component Heat Map

3.2 Descriptive statistics and Augmented Dickey-Fuller unit-root test

As it can be seen from table 2, the mean log return is 0.12%, which is acceptable for our zero-mean approximation. The computed values for skew (-0.96) and kurtosis (8.27) confirm our visual observations of the data. Also, RV exhibits high kurtosis (17.02) and skewness (3.30) further suggesting the presence of extreme values in our volatility series, which is in line with the stylised facts recited in the previous chapter. Both principal components show zero means as a result of the centring. The standard deviations are relatively large (1.37 and 1.07 respectively), considering the first and third quartile values, which suggest they capture a good amount of variance in the underlying variables. **The ADF test confirms stationarity of the returns (p=0.00), variance and all macroeconomic variables along with their PCs.**

Table 2: Descriptive Statistics

	Weekly			Monthly					
	Log return (%)	**Realised volatility** (%)	**Realised Variance** (%)	**Yield Curve Slope**	**Δ REER** (%)	**Δ Oil Price** (%)	**Δ Unemploy-ment** (%)	**PC1**	**PC2**
N	883.00	883.00	883.00	883.00	883.00	883.00	883.00	883.00	883.00
Mean	0.12	2.16	34.90	-0.01	0.00	0.01	0.00	0.00	0.00
Std	2.57	1.79	81.59	0.31	0.01	0.12	0.16	1.37	1.07
Min	-20.08	0.18	2.01	-1.34	-0.03	-0.54	-0.18	-4.09	-3.91
25 %	-0.99	1.06	7.06	-0.17	-0.01	-0.06	-0.02	-0.65	-0.41
50 %	0.27	1.70	14.95	0.02	0.00	0.01	0.00	-0.20	0.05
75 %	1.48	2.65	31.35	0.16	0.01	0.07	0.02	0.38	0.41
Max	11.42	17.02	776.22	0.94	0.06	0.88	2.34	8.75	10.11
Skewness	-0.96	3.30	6.61	-0.69	0.52	1.51	13.36	3.32	3.42
Kurtosis	8.27	17.14	49.58	3.74	1.20	16.26	189.36	18.02	35.58
ADF (T)	-17.80	-6.19	-5.72	-7.41	-5.42	-7.91	-6.23	-6.50	-7.53
ADF (p)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

3.3 Model estimation

The GARCH variants without external regressors are estimated on the test data using both the rugarch and arch packages. The model parameters are estimated with maximum likelihood estimation (MLE).

3.3.1 Estimation of the GARCH-models without exogenous regressors

In table 3 we present the in-sample parameter estimates for the GARCH-models without exogenous regressors, estimated with rugarch and arch:

`Rugarch` uses as its standard MLE solver a hybrid strategy, in where it will try **`solnp`**, **`nlminb`**, **`gosolnp`**, **`nloptr`** (see the respective documentations) in sequence if the previous one fails to converge. **Since the hybrid strategy does not explicitly tell which optimiser it uses without specification, we cannot be certain of it, although the default option for `solnp` is the Augmented Lagrange Multiplier Method.**

The **`arch`** package in Python for GARCH modelling uses the **`optimize`** module from **`scipy`** for optimization. The default optimiser selected by the module depends on the specific characteristics of the problem, such as whether it has constraints or bounds. If no specific optimizer is provided, **`optimize.minimize`** selects L-BFGS-B (see scipy documentation).

Table 3: Parameter Estimates for the GARCH, EGARCH, and GJR-GARCH Models

		Estimates				T-value			
		α	β	γ	ω	α	β	γ	ω
rugarch	**GARCH**	0.282	0.665	-	0.373	3.881	10.137	-	2.981
		(0.073)	(0.066)	-	(0.125)				
	EGARCH	0.353	0.881	-0.233	0.164	4.137	22.686	-4.587	2.816
		(0.085)	(0.039)	(0.051)	(0.058)				
	GJR-GARCH	0.029	0.690	0.408	0.404	0.748	9.448	3.568	3.146
		(0.039)	(0.073)	(0.114)	(0.129)				
arch	**GARCH**	0.246	0.696	-	0.364	5.740	14.765	-	3.348
		(0.043)	(0.047)	-	(0.109)				
	EGARCH	0.341	0.873	-0.255	0.192	6.049	34.273	-7.271	5.042
		(0.056)	(0.025)	(0.035)	(0.038)				
	GJR-GARCH	0.016	0.694	0.456	0.435	0.532	15.115	5.376	4.772
		(0.031)	(0.046)	(0.085)	(0.091)				

The parameter estimates between the identical models do not differ greatly, although there are noticeable differences. The asymmetric models indicate the presence of leverage effect in both software environments, the arch estimates being somewhat more biased towards asymmetric shocks with larger absolute values for γ in both asymmetric models. **The arch estimates are also larger for ω in the asymmetric models, indicating a stronger propensity for mean reversion.** The GJR-model is the only model with a parameter T-value outside of a reasonable critical range (α), further underlining the importance of modelling the asymmetry.

3.3.2 Estimation of the GARCH-MIDAS model

We attempt three different regressor designs for the long-term component τ_t in the asymmetric GARCH-MIDAS model:

1. **PC1** and monthly **RV**
2. **PC2** and monthly **RV**
3. **PC1** and **PC2**

In table 4 we present the in-sample[4] parameter estimates for the three mixed-frequency models with-exogenous regressors, estimated with the **`mfGARCH`** (Conrad & Kleen, 2020; Kleen, 2021) package in R.

The package uses **`maxLik`**, as the standard optimising package, which will cycle through Newton-Raphson, Nelder-Mead and BFGS algorithms when attempting to solve the MLE.

For the beta weighting scheme, in the MIDAS estimation of the long-term component, we follow the default suggestion by Kleen (2021) and fix w_1 at one. Additionally, we opted to only use one lag of the final versions of the three selected specifications, as additional lags did not seem to provide any incremental predictive

[4] We opted to omit the other in-sample parameter estimates with multiple lags, to save space, as the first-order models performed significantly better already at that stage.

power in the in-sample testing – simultaneously abiding by the law of parsimony and keeping the comparison between the other first order GARCH models fair.

Table 4: Parameters Estimates for the MIDAS Models

		Estimates						T-value					
X_1	X_2	α	β	γ	m	θ_1	θ_2	α	β	γ	m	θ_1	θ_2
PC1	**PC2**	0.020	0.677	0.419	1.513	-0.091	-0.027	0.461	15.859	5.206	6.096	-1.388	-0.413
		(0.043)	(0.043)	(0.081)	(0.248)	(0.065)	(0.065)						
RV	**PC1**	0.017	0.672	0.429	1.503	0.001	-0.098	0.407	15.872	5.436	5.549	0.310	-1.510
		(0.042)	(0.042)	(0.079)	(0.271)	(0.002)	(0.065)						
RV	**PC2**	0.034	0.678	0.386	1.504	0.000	-0.030	0.865	16.326	5.059	5.711	-0.015	-0.424
		(0.040)	(0.042)	(0.076)	(0.263)	(0.002)	(0.070)						

Alpha is relatively small across all models and lack statistical significance, much in line with the corresponding values of the estimated GJR-GARCH models. The gamma estimates are also in line with those of the GJR-GARCH models. Beta is high in all models, exceeding 0.67, which suggests high weekly clustering. Somewhat surprisingly, the largest parameter estimates are found in the constant m. The order of magnitude ($m > 1.5, compared\ to\ \omega < 0.44$) is several factors larger than for the other GARCH models, highlighting the base level of weekly variance. The MIDAS parameters $\theta_1$and θ_1 are all negative for all PCs, with PC1 displaying the largest effect in the RV-PC1 model at -0.098 and RV showing none or negligible effects. The statistical significance is debatable for the PCs, although they still contribute to the model.

3.4 Out-of-sample forecasting and model evaluation

"A volatility model should be able to forecast volatility", is the opening sentence of by Engle and Patton (2007) in their paper *What good is a volatility model?* To test whether our constructed models are able to effectively forecast (realised) volatility, we perform recursive out-of-sample forecasts over the test set and assess the relative accuracy and inaccuracy of the models against the computed realised volatility (RV) defined in chapter 2.

We implement recursive out-of-sample forecasts in our models, where the models incrementally integrate new observations into the training data. Initially, a model is trained using the original training data. As new observations become available, the realised volatilities are added recursively to the training data, and the model is re-trained on the updated dataset. This ensures that the forecasts incorporate the most recent information, thereby increasing the model's accuracy.

Our performance evaluation framework is comprised of:

i) **TWO LOSS METRICS**; Mean Absolute Percentage Error (MAPE) and the Root of Mean Squared Percentage Error (RMSPE) to measure absolute and extreme errors respectively, defined below:

$$\boldsymbol{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{\sigma}_t - RV_t}{\hat{\sigma}_t}\right|, \qquad \boldsymbol{RMSPE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\hat{\sigma}_t - RV_t}{\hat{\sigma}_t}\right)^2}$$

ii) **TWO ACCURACY METRICS**; Mean Directional Accuracy (MDA) to measure how well models predict direction and out-of-sample-R^2 as a reference measure general goodness of fit:

$$\boldsymbol{R^2_{\text{oos}}} = 1 - \frac{\sum_{i=1}^{n}(\hat{\sigma}_t - RV_t)^2}{\sum_{t=1}^{n}(\overline{RV} - RV_t)^2}, \qquad \boldsymbol{MDA} = \frac{1}{n}\sum\nolimits_{t=1}^{n} \text{sgn}(RV_t - RV_{t-1}) == \text{sgn}(\hat{\sigma}_t - RV_{t-1})$$

$$\text{sgn}(\hat{\sigma}_i - RV_{i-1}) := \begin{cases} -1 \Rightarrow (\hat{\sigma}_t - RV_{t-1}) < 0 \\ \ \ 0 \Rightarrow (\hat{\sigma}_i - RV_{t-1}) = 0 \\ \ \ 1 \Rightarrow (\hat{\sigma}_i - RV_{t-1}) > 0 \end{cases}$$

iii) **DIEBOLD-MARIANO TESTING**; a statistical test proposed by Diebold and Mariano (1995) ("**DM-test**") to test the statistical significance of some loss differential[5] between two different forecast series. As we already deploy RMSPE as a loss metric, we compute the DM-test utilising the root of squared errors. The null and alternative hypotheses of the (two sided) DM-test is expressed below

$H_{DM(0)}$:	*The expected difference in the squared forecast errors of the models is zero*
$H_{DM(a)}$	*The expected difference in the squared forecast errors of the models is not zero*

3.4.1 Statistical hypothesis

Reiterating our research hypotheses: i) Asymmetric innovations (EGARCH and GJR-GARCH) provide incremental out-of-sample performance to a standard GARCH and ii) Lower frequency macroeconomic variables via PCA & MIDAS provide incremental out-of-sample performance to an asymmetric GARCH, we seek to validate previous results showing that introducing asymmetry to simple GARCH models can drastically improve their performance, and to investigate whether the optimiser differences between popular software packages can be large enough to produce significant differences in predictive performance. We also seek to find if the MIDAS-framework, extended by asymmetry and PCA can reliably outperform simpler and easier-to-estimate models.

The null hypothesis below will be rejected by a mix of critical loss and accuracy metric review, combined with the statistical DM-testing of RSE between models.

$\text{H}_{1(0)}$:	*There are no measurable differences in the predictive power between the models, and unlikely economic significance*
$H_{1(a)}$	*There are measurable differences in the predictive power between the models and likely economic significance*

[5] The DM test assumes stationarity of the loss differentials, and that they are serially uncorrelated, which we validated with the Augmented Dickey-Fuller and Ljung-Box tests respectively. All loss differentials are covariance stationary and serially uncorrelated ($p < 0.05$). We will not post the table here in order to save space.

4 RESULTS AND ANALYSIS

In this section, the out-of-sample performance of various GARCH models applied to S&P500 weekly realised volatility is analysed. The evaluation focuses on the error metrics MAPE and RMSPE, and the accuracy metrics R² and MDA, which can all be found in table 5. An additional part of the analysis is the use of the Diebold-Mariano test to determine the statistical significance of differences in forecast accuracy between models, with results detailed in table 6.

4.1 Model Accuracy

Notably, the GARCH(1, 1) model with static parameters exhibits the poorest performance, despite recording the highest R² value. This is reflected by its MAPE of 4.845 and RMSPE of 12.479. **Every model implementing the recursive out-of-sample forecasts outperform the fixed GARCH(1, 1) model.** We opted to include the fixed parameter forecast to illustrate the relative unreliability of R^2 as a standalone accuracy metric.

Table 5: Model Prediction Quality Metrics

		Loss		Accuracy	
	Model	MAPE	RMSPE	MDA	R^2
`arch`	GARCH (1,1) -FIXED	4.845	12.479	0.536	**0.645**
	GARCH (1,1)	0.521	0.789	0.649	0.427
	GJR-GARCH	0.444	0.655	0.696	0.539
	EGARCH	0.514	0.783	0.655	0.472
`r-ugarch`	R-GARCH	0.557	0.836	0.637	0.393
	R-GJR-GARCH	0.449	0.659	0.696	0.541
	R-EGARCH	0.436	0.655	**0.708**	0.582
`mf-Garch`	GARCH-MIDAS-RV-PC1	0.43	0.637	**0.714**	0.509
	GARCH-MIDAS-RV-PC2	0.432	0.648	0.69	0.523
	GARCH-MIDAS-PC1-PC2	**0.422**	**0.634**	**0.702**	0.553

It can be observed that models of the same type generally exhibit smaller differences in error metrics than models of varying complexity. Models that incorporate leverage effects demonstrate improved performance over the simple GARCH(1, 1) models, as evidenced by lower MAPE and RMSPE values, along with higher MDA and R². The Python E-GARCH model underperforms relative to its R-EGARCH equivalent, exhibiting a MAPE of 0.514 and an RMSPE of 0.783, which contrasts with the R-EGARCH's lower values of 0.436 and 0.655, respectively. While the `arch` package's EGARCH model is among the least accurate, the R-EGARCH model boasts one of the highest MDAs at 0.708.

The inclusion of lower frequency macro variables in the GARCH-MIDAS models appears to enhance accuracy beyond that of the asymmetric models, although the R-EGARCH model from `rugarch` exhibits a performance that is highly similar. **The GARCH-MIDAS model that incorporates**

both principal components registers the lowest values for MAPE and RMSPE, at 0.422 and 0.634 respectively, but it lags in terms of MDA and R^2. However, considering that the fixed GARCH(1, 1) model achieves the highest R^2 yet performs poorly on other metrics, there is reason to be sceptical about the reliability of R^2 as an indicator of model performance in the models.

4.2 Statistical testing of the differences

Table 6 shows the results of the Diebold-Mariano tests. Although the error metrics between identical models in different software packages are miniscule, the p-values from the Diebold-Mariano test for many equivalent models are close to zero, indicating a statistically significantly different time series. For instance, the p-value between GJR-GARCH and R-GJR-GARCH is 0.021, and 0.000 between R-GARCH and GARCH(1, 1).

Table 6: Diebold-Mariano Test P-values

	GARCH(1,1)	*GJR-GARCH*	*EGARCH*	*R-GARCH*	*R-GJR-GARCH*	*R-EGARCH*	*GM-RV-PC1*	*GM-RV-PC2*	*GM-PC1-PC2*
GARCH(1,1)	---	*0.000*	*0.327*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
GJR-GARCH	0.000	---	0.000	0.000	0.021	0.37	0.118	0.201	0.011
EGARCH	0.327	0.000	---	0.000	0.000	0.000	0.000	0.000	0.000
R-GARCH	0.000	0.000	0.000	---	0.000	0.000	0.000	0.000	0.000
R-GJR-GARCH	0.000	0.021	0.000	0.000	---	0.182	0.037	0.087	0.003
R-EGARCH	0.000	0.37	0.000	0.000	0.182	---	0.658	0.709	0.185
GM-RV-PC1	0.000	0.118	0.000	0.000	0.037	0.658	---	0.864	0.426
GM-RV-PC2	0.000	0.201	0.000	0.000	0.087	0.709	0.864	---	0.096
GM-PC1-PC2	**0.000**	**0.011**	**0.000**	**0.000**	**0.003**	0.185	0.426	**0.096**	---

4.3 Conclusions

In this study, we combined asymmetric innovations, MIDAS and macroeconomic variables via PCA methodology in an attempt to enhance future variance predictions. Through principal component analysis, changes in the yield curve slope, USD real effective exchange rate, crude price and the unemployment rate were used to inform the long-term variance component. A recursive forecast approach was utilised for out-of-sample variance forecasting, allowing for comparison with standard GARCH models, and the comparisons between identical models estimated with different methods.

The results of this study support the argument that the choice of software is secondary to model hyperparameter configuration, as the differences in error metrics between identical models are considerably smaller than those between models with varying complexity. When examining the forecasts from figure 5, identical models also yield visually almost indistinguishable results. Despite these facts, the DM test results often indicate that similar models in different software packages are statistically significantly different from each other, suggesting that **the choice of hyperparameters and model optimization still plays a significant role in model building.**

Our investigation confirms that enhancing GARCH models with asymmetric adjustments and mixed frequency macroeconomic variables leads to improved forecasting performance. This is evidenced by lower error metrics and higher Mean Directional Accuracy (MDA), supporting our hypotheses that such integrations are beneficial beyond simpler GARCH frameworks. This progression mirrors the advancements made in the field by Nelson (1991) with EGARCH and by Engle, Ghysels & Sohn (2013) with the GARCH-MIDAS model, reinforcing the notion that nuanced model extensions can provide substantial out-of-sample forecasting benefits. Our results indicate that the EPiC-GARCH, and component models in general, show great promise in variance forecasting, especially with the option to include large classes of variables from which one can extract the most efficient PCs. This improvement is evident in the comparison of loss metrics, where the PC1-PC2 model scored best in the cohort and to a lesser extent, in MDA, where the RV-PC1 model scored highest. Although the PC1-PC2 model's RSE superiority was confirmed by the DM test ($p < 0.05$) against most of the cohort, it still failed to do so in comparison to the R-estimated EGARCH ($p = 0.185$). Through the models, R^2 ranges unpredictably, making it a questionable choice for assessing model performance in GARCH models.

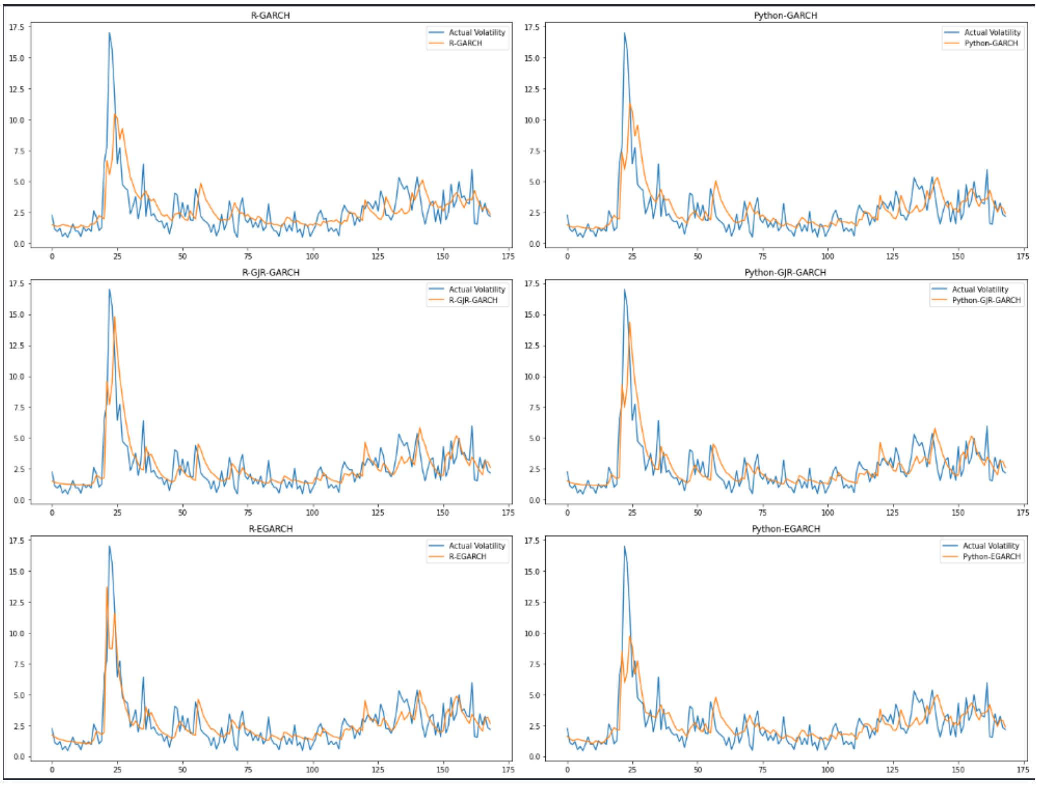

Figure 4: Python and R GARCH Predictions

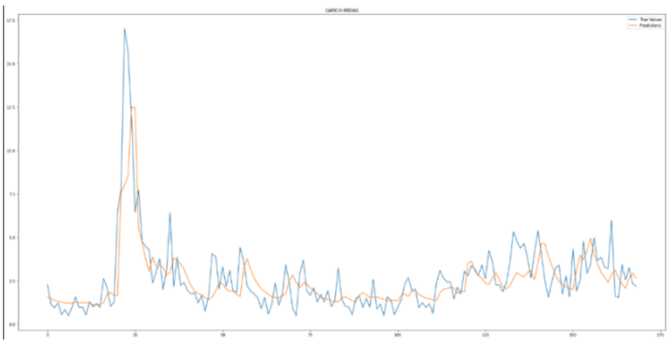

Figure 5: EPiC-GARCH (PC1-PC2) Predictions vs Realized Volatility

REFERENCES

Andersen, T. G., & Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. International Economic Review, 39(4), 885. https://doi.org/10.2307/2527343

Asgharian, H., Hou, A. J., & Javed, F. (2013). The importance of the macroeconomic variables in forecasting stock return variance: A GARCH-MIDAS approach. Journal of Forecasting, 32(7), 600–612. https://doi.org/10.1002/for.2256

Bollerslev, T. (1986). GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY. In Journal of Econometrics (Vol. 31).

Charles, A., & Darné, O. (2019). The accuracy of asymmetric GARCH model estimation. International Economics, 157, 179–202. https://doi.org/10.1016/j.inteco.2018.11.001

Conrad, C., & Kleen, O. (2020). Two are better than one: Volatility forecasting using multiplicative component GARCH-MIDAS models. Journal of Applied Econometrics, 35(1), 19–45. https://doi.org/10.1002/jae.2742

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Ding, Z., & Granger, C. W. J. (1996). Modeling volatility persistence of speculative returns: A new approach. Journal of Econometrics, 73(1), 185–215. https://doi.org/10.1016/0304-4076(95)01737-2

Engle, R.F., & Lee, G. (1999). A Permanent and Transitory Component Model of Stock Return Volatility. In Robert F. Engle & H. White (Eds.), Cointegration, Causality and Forecasting: A Festschrift in Honor of Clive W.J. Granger. Oxford University Press.

Engle, Robert F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987. https://doi.org/10.2307/1912773

Engle, Robert F, Ghysels, E., & Sohn, B. (2013). STOCK MARKET VOLATILITY AND MACROECONOMIC FUNDAMENTALS. http://direct.mit.edu/rest/article-pdf/95/3/776/1917282/rest_a_00300.pdf

Engle, Robert F., & Patton, A. J. (2007). What good is a volatility model? https://doi.org/10.1016/B978-075066942-9.50004-2

Galanos, A. (2023). rugarch: Univariate GARCH models (R package version 1.5-1.). https://github.com/alexiosg/rugarch

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2005). There is a risk-return trade-off after all. Journal of Financial Economics, 76(3), 509–548. https://doi.org/10.1016/j.jfineco.2004.03.008

GLOSTEN, L. R., JAGANNATHAN, R., & RUNKLE, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. The Journal of Finance, 48(5), 1779–1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x

Masset, P. (2011). Volatility Stylized Facts. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1804070

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica, 59(2), 347. https://doi.org/10.2307/2938260

SCHWERT, G. W. (1989). Why Does Stock Market Volatility Change Over Time? The Journal of Finance, 44(5), 1115–1153. https://doi.org/10.1111/j.1540-6261.1989.tb02647.x

Sheppard, K. (2021). bashtage/arch: Release 4.18 (Version v4.18). Zenodo. https://doi.org/https://doi.org/10.5281/zenodo.593254