Abstract

Financial markets exhibit time-varying volatility and cross-asset correlation. To construct and optimize a risk-adjusted portfolio and account for underlying collinearity between securities, a good understanding of these two dynamics is important. In this study global equity indices are modeled with the purpose of improving the forecasting accuracy of time-varying correlations, as well as distinguishing the difference between global multivariate and pairwise settings.

A core part of this is the difference between a DCC-GARCH baseline model and its numerous exogenous market signal customizations. Examples of these are the volatility index, dark-pool activity index and credit spreads. Smaller models are compared with the full model to question the efficacy of the multivariate approach.

The results show that the inclusion of some exogenous variables provide marginal improvement in predictive accuracy if added right, whilst some worsen it, either numerically or through added complexity. Credit spreads show modest but mostly beneficial improvements, both for in sample fit and out-of-sample forecast errors for both the full multivariate model and for the pairwise models. The volatility index showed nuanced results, improving correlation forecasts for markets pairs such as US-Europe, while worsening predictability for US-Asia pairs. DIX shows the weakest and least consistent effects, suggesting dark pool activity carries limited information for global correlation dynamics in our model. A recurring finding is that the DCC-GARCH baseline already captures much of the dynamics that the exogenous signals might otherwise explain. Furthermore, the model dimensionality affects the magnitude of market signal changes. A recurring theme across all signals is that the extent of improvement is small.

Sammanfattning

Finansiella marknader kännetecknas av tidsvarierande volatilitet och korrelationer mellan tillgångar. Inom risk-och portföljoptimering är det nyttigt att få en så bra uppfattning av sådana samrörelser. Dessa modelleras i denna studie mellan globala aktieindex i syfte att förbättra prognostiseringen av tidsvarierande korrelationer samt jämföra beteendet hos globala respektive parvisa modeller.
En central del av undersökningen utgörs av skillnaden mellan en baslinje och dess utbyggnad med exogena marknadssignaler. Exempel på dessa är volatilitetsindex, dark-pool index och kreditspreadar. Huruvida dessa modeller kan indikera på systemisk stress, samt ifall en stor korrelationsmodell fångar detta även på lokala marknader, diskuteras också. Resultaten visar en viss förbättring från vissa exogena signaler, i form av lägre prognosfel medan andra har ytterst liten påverkan eller tillochmed en försämring av prediktionen. Kreditspreadar visar modesta men genomgående förbättringar, med både bättre insample fit och lägre prediktionsfel out-of-sample. Volatilitetsindexet visade resultat som förbättrar prognosen för vissa marknadspar såsom US-Europa, medans försämrar den för US-Asien. DIX var den minst påverkande signalen, vilket kan tyda på att dark poolaktivitet har begränsad information för korrelationsdynamiker inom vår modell. Något återkommande är att DCC-GARCH basmodellen redan fångar mycket av den dynamik som exogena signaler annars skulle kunna förklara. Vidare visas att modelldimensionaliteten påverkar vilken effekt de exogena signalerna har. Ett genomgående tema för samtliga signaler att förbättringarnas omfattning är marginell.

Acknowledgements

We would like to express a special thank you to our supervisor Elena Westerdahl from AP3 for not only sourcing vast amounts of data and inspiration, but also for great guidance throughout the study.
We would also like to extend our gratitude to AP3 for giving us this opportunity.
Finally, thank you Ted Kronvall for your support, academic feedback and the many interesting discussions that helped shape this thesis.

Contents

List of Figures

List of Tables

1 Introduction

Financial markets are complex and dynamic systems where the prices of assets are continuously affected by different kinds of new information, investor behavior, and macroeconomic factors [1]. A pension fund for example is likely holding stocks in both the US and in Europe. In calm periods, these tend to move somewhat independently. But during the 2008 financial crisis or the COVID-19 crash in March 2020, virtually every market fell simultaneously and quickly. The diversification that normally exists in the portfolio quickly blurs out. The fact that correlations between markets rise during periods of stress, is well documented [2] and makes accurate modeling of time-varying correlation essential for risk management. Being able to anticipate when markets are likely to move more closely together allows investors to adjust their portfolios. A central property of financial time series and risk management is thus volatility, which refers to the degree of variation in returns over time, i.e. the percentage change in the price day to day. Unlike many other statistical assumptions, volatility in finance is rarely stationary, which can be problematic because many standard econometric models assume constant variance over time. Instead, it is more common than not that the market volatility is in ever-changing flux, something typically known as heteroskedasticity. It means that the variance is not constant [3]. Markets tend to show periods in which high volatility is usually followed by high volatility, and calmer periods of variation follow each other, also known as volatility clustering. [4]

As volatility changes over time, so do the relationships between assets, where assets in our case refer to market indices representing aggregates of multiple stocks. In times of financial stress, correlations between market indices tend to increase, which affects both diversification and risk optimization [2]. Modeling both volatility and correlation dynamics is therefore essential for risk management and tactical allocation decisions. These dynamics can often be connected to distinct states or periods in the market in which the statistical properties of returns are relatively stable but fundamentally different from those in other states, such as a crisis regime where uncertainty is high and correlations rise. This is what is known as market regimes [5].

To capture and forecast these time varying patterns, models that allow for conditional heteroskedasticity, meaning the variance changes over time but now also depending on past information and dynamic correlations have been developed. One of the more popular models is the GARCH model [3], in which volatility is modeled as a function of past shocks and past conditional variances. Univariate GARCH models on their own can effectively describe volatility for single assets but are often insufficient when the purpose is to analyze covariation and time varying correlations between multiple returns. To address this, multivariate extensions have been developed, such as the DCC-GARCH (Dynamic Conditional Correlation - Generalized Autoregressive Conditional Heteroskedasticity) model.

Traditional DCC-GARCH models are primarily driven by historical returns and past shocks. In practice, however, there are also observable macro financial indicators that may contain information about current risk conditions and thus potentially help to better detect market regimes [6]. These types of market signals are included as exogenous inputs, to establish whether or not they drive significant change in model performance. Depending on the nature of the market signal they can enter the DCC-GARCH in either the volatility end (GARCH-X) or the correlation end (DCC-X), essentially creating different permutations of a larger "DCC-X-GARCH-X" model.

For large institutional investors, understanding changing market correlations is particularly important because the benefits of portfolio diversification often weaken during periods of financial stress. [2] The Third Swedish National Pension Fund (AP3) manages a substantial portion of Sweden's public pension capital. They aim to provide maximum possible benefit through responsible investment and management of the pension system buffer. Since large institutional portfolios like theirs are exposed to global market movements, accurate modeling of time-varying volatility and correlations becomes an integral part for both the portfolio construction and the management of risk.

While traditional DCC-GARCH models have proven effective in capturing time-varying correlation, they may not fully incorporate all other market information that reflects the current financial condition. Other research has shown that GARCH models with exogenous variables can better explain volatility in return rates. [7] However, less attention has been given to how external market signals affect correlation forecasting for DCC-GARCH specifically. This study investigates whether adding specific market signals provide information that better helps explain and forecast dynamic correlations.

The remainder of this thesis is structured as follows:

Chapter 2 presents the theoretical framework and introduces the models used in the study. Chapter 3 describes the data, selected exogenous variables, and the methodology applied for estimation and forecasting. It also presents the empirical results and model evaluation. Chapter 4 presents the conclusions, and Chapter 5 contains suggestions for future research.

1.1 Research Proposal

- Can exogenous market signals improve the out-of-sample forecasting of DCC-GARCH models relative to a baseline model?
- Do different exogenous variables affect volatility and correlation forecasts differently across global equity markets?
- Does the efficiency of exogenous inputs differ between large multivariate DCC models and smaller pairwise DCC specifications?

2 Theory

2.1 GARCH

Financial time series often show characteristics that deviate from the assumptions of classical linear models. These include volatility clustering and excess kurtosis (more extreme outliers) which motivate the use of models that explicitly allow for conditional heteroskedasticity. One of the more prominent models is the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, introduced by Bollerslev in 1986 [3] as a generalization of the ARCH model proposed by Engle in 1982 [8]. The key idea behind GARCH is that the conditional variance at time t depends on both past squared shocks and past conditional variances, unlike ARCH which specifies the conditional variance solely as a function of past squared shocks. As a result, GARCH is able to capture the persistence in volatility that is commonly observed in financial data.

Let r_t denote the log return at time t, where P_t represents the daily closing price at time t. The log return is defined as

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{2.1}$$

The return process is assumed to follow:

$$r_t = \mu + \varepsilon_t, \tag{2.2}$$

$$\varepsilon_t = \sigma_t z_t, \tag{2.3}$$

Where μ is the conditional mean and ε_t is the error term and $z_t \sim \mathcal{N}(0,1)$ is an independent and identically distributed process with zero mean and unit variance.

The conditional variance at time t in a general GARCH(p,q) model is given by

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{2.4}$$

where $\omega > 0$, $\alpha_i \geq 0$, and $\beta_j \geq 0$.

The model order, p and q, determines the number of lags included in the variance equation. A higher order specification may be subject to overfitting due to the increased number of parameters. In empirical applications, the GARCH(1,1) specification is typically enough and will be assumed throughout the paper [3, 6]. The GARCH(1,1) model is specified as follows.

The conditional variance at time t for a GARCH$(1,1)$ is given by:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{2.5}$$

The parameters satisfy $\omega > 0$, $\alpha \geq 0$, and $\beta \geq 0$. The parameter α measures the short term impact of shocks on volatility, like a spike in volatility after a surprise interest rate announcement. While β on the other hand captures the persistence of volatility over time, so a high β essentially means the market "remembers" the shock for a long time. For stationarity, a condition is that $\alpha + \beta < 1$, which ensures that shocks to volatility decay over time.

2.2 Dynamic Conditional Correlation

In this study, the Dynamic Conditional Correlation (DCC)-GARCH, introduced by Engle [9] is used to model the covariance structure of multiple assets. Unlike univariate GARCH, the DCC model has fewer parameters to estimate since they are jointly estimated for the full matrix of market indices.

Let r_t be an $N \times 1$ vector of returns with conditional mean μ_t. It can be written as:

$$r_t = \begin{pmatrix} r_{1t} \\ r_{2t} \\ \vdots \\ r_{Nt} \end{pmatrix}, \quad \mathbb{E}[r_t] = \mu_t = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{Nt} \end{pmatrix}. \tag{2.6}$$

where the innovation term ε_t, used in $r_t = \mu_t + \varepsilon_t$, is assumed to follow a conditional multivariate normal distribution and $\mathcal{F}_{t-1}$ denotes the information set available up to time t-1, containing all past returns and other information used for forecasting.:

$$\varepsilon_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(0, H_t), \tag{2.7}$$

In the DCC model, the conditional covariance matrix H_t is decomposed as,

$$H_t = D_t R_t D_t, \tag{2.8}$$

where $D_t = \text{diag}(\sqrt{h_{1,t}}, \ldots, \sqrt{h_{N,t}})$ is a diagonal matrix of conditional standard deviations from the univariate GARCH models as mentioned before, and R_t denotes the time-varying conditional correlation matrix.

Each conditional variance $h_{i,t}$ is modeled using a GARCH(1,1) process:

$$h_{i,t} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1}. \tag{2.9}$$

Defining $\sigma_{i,t}^2 = h_{i,t}$, the diagonal matrix of conditional standard deviations is given by,

$$D_t = \text{diag}(\sigma_{1,t}, \ldots, \sigma_{N,t}). \tag{2.10}$$

The standardized residuals are defined as,

$$z_t = D_t^{-1} \varepsilon_t. \tag{2.11}$$

The dynamics of the intermediate covariance matrix Q_t are given by,

$$Q_t = (1 - a - b)\bar{Q} + a z_{t-1} z_{t-1}' + b Q_{t-1}, \tag{2.12}$$

where $\bar{Q}$ denotes the unconditional covariance matrix of z_t, and a and b are non-negative scalar parameters satisfying $a + b < 1$, which is required for stationarity. This makes sure that the correlation process remains stable and reverts toward the long run correlation matrix $\bar{Q}$ rather than diverging over time.

The conditional correlation matrix is obtained through normalization,

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}, \tag{2.13}$$

to ensure that R_t is a valid correlation matrix with unit diagonal elements.

2.3 Components

2.3.1 Parameter estimation

Engle and Sheppard [9] propose a two step estimation for the parameters of the DCC model. The estimation is divided into two separate parts, a volatility and a correlation part. The main advantage of this decomposition is that it substantially reduces the computational load compared to joint maximization of all parameters, while a disadvantage might be the potential loss of index-specific signal. In the empirical implementation, parameters are estimated using Gaussian quasi maximum likelihood (QML) [10]. Under this approach, the Gaussian distribution is used as a convenient estimation but does not require the returns to be truly normally distributed. However, to account for potential excess kurtosis in financial returns, a multivariate Student's t specification is also considered. In that case, an additional degrees of freedom parameter, ν is estimated. This estimate is a useful gauge of the excess kurtosis, where lower degrees of freedom correspond to more extreme outliers than in a standard gaussian distribution. For reference, the normal distribution could be seen as a constrained student's t distribution where $\nu \to \infty$. Higher kurtosis implies that extreme observations are to be expected more frequently [11].

Returning to the conditional covariance matrix,

$$H_t = D_t R_t D_t, \tag{2.14}$$

let θ denote the parameters in the variance part (i.e., those governing D_t) from equation (2.4), and let ϕ denote the parameters in the correlation part (i.e., those governing R_t). Under the Gaussian assumption, the full log-likelihood function becomes,

$$L(\theta, \phi) = -\frac{1}{2} \sum_{t=1}^{T} \left(n \log(2\pi) + \log |H_t| + r_t' H_t^{-1} r_t \right). \tag{2.15}$$

Substituting the conditional covariance matrix (equation (2.8))and using the determinant identity instead:

$$|H_t| = |D_t|^2 |R_t|, \tag{2.16}$$

This enables the log-likelihood function to be separated into one volatility component and one correlation component. As shown by Engle and Sheppard [9], the parameters can therefore be estimated in two steps first, the GARCH parameters governing D_t, and then the DCC parameters governing R_t.

$$L(\theta, \phi) = L_v(\theta) + L_c(\theta, \phi). \tag{2.17}$$

Where the volatility part, θ of the log-likelihood is given by:

$$L_v(\theta) = -\frac{1}{2} \sum_{t=1}^{T} \left(n \log(2\pi) + 2 \log |D_t| + r_t' D_t^{-2} r_t \right), \tag{2.18}$$

n is the number of assets and D_t contains the conditional standard deviations from the univariate GARCH models.

The correlation component, ϕ of the log-likelihood is given by:

$$L_c(\theta, \phi) = -\frac{1}{2} \sum_{t=1}^{T} \left(\log |R_t| + \varepsilon_t' R_t^{-1} \varepsilon_t - \varepsilon_t' \varepsilon_t \right), \tag{2.19}$$

Where $\varepsilon_t = D_t^{-1} r_t$ denotes the vector of standardized residuals.

The volatility part of the likelihood can be written as the sum of individual univariate GARCH log-likelihood functions:

$$L_v(\theta) = \sum_{i=1}^{n} \sum_{t=1}^{T} \ell_{it}(\theta_i), \tag{2.20}$$

Where $\ell_{it}(\theta_i)$ denotes the log-likelihood contribution of series i at time t.

The estimation follows a two step block coordinate descent proposed by Engle and Sheppard [9]. Since optimization is implemented using minimize in Python, the negative log-likelihood is minimized rather than maximizing the likelihood directly. The volatility component of the likelihood:

$$\hat{\theta} = \arg\min_{\theta} \; -L_v(\theta). \tag{2.21}$$

In the second step, the estimated variance parameters $\hat{\theta}$ are treated as given, and thus the correlation parameters are obtained by maximizing the correlation component of the likelihood:

$$\hat{\phi} = \arg\min_{\phi} \; -L_c(\hat{\theta}, \phi). \tag{2.22}$$

Since the second step depends on the first step, consistency of the first estimator implies consistency of the second step estimator, provided that the function is continuous at the true parameter values [9]. As for the multivariate Student's t distribution, the two step estimation is retained. Although as mentioned previously, the likelihood function is modified and an additional degrees of freedom, ν is estimated.

2.3.2 Exogenous inputs

Exogenous market signals may provide insight into financial stress and risk that the relative index returns don't necessarily portray on their own. There are volatility drivers, that together with the GARCH create a GARCH-X model, as well as correlation drivers that impose the DCC-X. This makes it possible to see patterns not only connected to historical price movements, but broad market indicators as well.

One such example is the credit spread. Simply put, it is the difference in yields between private debt instruments and government debt of similar maturities. These spreads are well-known indicators of upcoming market instability and systemic risk [12].

Also used in this study is a sentiment indicator called VIX. Often called the "fear-index", it provides real-time updates of the expected volatility in the market, based on the prices of options for the S&P 500. It is often interpreted to make informed trading decisions, and is incorporated here as a leading indicator that could improve the model fit of a resulting DCC-X-GARCH-X for out-of-sample data.

Another exogenous input is the Dark Index (DIX), introduced by SqueezeMetrics[13], which is a dollar-weighted average of the Dark Pool Index (DPI). It can be summarized as an off-market liquidity proxy that resembles institutional sentiment, meaning that a high DIX level indicate net institutional buying, and conversely a low DIX means net selling. In other words DIX can, much like VIX, be used to make an informed decision while trading, based on the liquidity flows of institutions.

Additive exogenous terms that are likely drivers of volatility are inserted straight into the previously stated GARCH(1,1) model, making it:

$$h_{i,t} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1} + \gamma_i x_{t-1} \tag{2.23}$$

where $x_{t-1} \in \mathbb{R}$ is an exogenous variable and $\gamma_i \in \mathbb{R}^k$ is the corresponding parameter vector.

If instead there are correlation drivers, the previously stated DCC matrix turns into

$$Q_t = (1 - a - b)\bar{Q} + a z_{t-1} z_{t-1}' + b Q_{t-1} + \Gamma x_{t-1} \tag{2.24}$$

where Γ is a matrix of parameters capturing the influence of the exogenous inputs on the correlation dynamics.

Significant estimates of γ and Γ for the market signals has the potential to make the model more efficient, since it won't be based solely on the internal shock persistence of α and variance persistence of β. Instead, explanatory power will be provided by the external drivers, potentially yielding a faster model reaction.

Two prerequisites for proper usage of these external drivers are stationarity and non-negativity. For examples such as VIX and the credit spread used, positiveness is a given per default. But this might not be the case for all market signals.

2.3.3 Forecasting

An important objective is to not only achieve good in-sample fit, but to forecast volatility and correlation well out-of-sample. To utilize what is already known to best estimate the expected value of what is to come.

$$\hat{r}_{t+k|t} = \mathbb{E}[r_{t+k} \mid \mathcal{F}_t]. \tag{2.25}$$

where k is an arbitrary step forward in the series and $\hat{r}_{t+k|t}$ is the expected value of the returns conditioned on observed returns.

Since GARCH is linear in past squared innovations and past variances, the one step ahead forecast has a closed form expression. And can be calculated as,

$$h_{t+1|t} = \mathbb{E}[\sigma^2_{t+1} \mid \mathcal{F}_t] = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2. \tag{2.26}$$

Since both ε_t^2 and h are known at time t, this forecast can be computed directly.

For forecasts more than one step ahead ($k > 1$):

$$\mathbb{E}[\varepsilon^2_{t+k-1} \mid \mathcal{F}_t] = \mathbb{E}[\sigma^2_{t+k-1} \mid \mathcal{F}_t]. \tag{2.27}$$

The k-step-ahead conditional variance forecast can then be written recursively as

$$\hat{\sigma}^2_{t+k|t} = \omega + (\alpha + \beta)\hat{\sigma}^2_{t+k-1|t}. \tag{2.28}$$

This implies that the variance forecast converges toward the long-run (unconditional) variance,

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}, \tag{2.29}$$

provided that $\alpha + \beta < 1$, which ensures covariance stationarity.

For DCC, unlike the univariate GARCH model, the forecasting procedure is nonlinear. While the auxiliary matrix Q_t follows a GARCH type linear recursion, the conditional correlation matrix R_t is obtained through a nonlinear normalization of Q_t. Consequently, h-step ahead forecasts of R_{t+k} cannot be computed directly in the same closed form manner as the variance forecasts in the GARCH model.

The nonlinearity arises from the normalization step

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}, \tag{2.30}$$

Since the expectation of a nonlinear function does not equal the function of the expectation, i.e., $\mathbb{E}[f(Q_{t+k})] \neq f(\mathbb{E}[Q_{t+k}])$, k-step forecasts of R_{t+k} do not create a simple closed form. Therefore, Engle and Sheppard [9] proposed the following approximation:

$$\bar{R} \approx \bar{Q}, \tag{2.31}$$

and the conditional expectation is approximated as

$$\mathbb{E}[R_{t+k} \mid \mathcal{F}_t] \approx \mathbb{E}[Q_{t+k} \mid \mathcal{F}_t], \tag{2.32}$$

Since the dynamics of Q_t follow a GARCH-like structure,

$$Q_t = (1 - a - b)\bar{Q} + a z_{t-1} z'_{t-1} + b Q_{t-1}, \tag{2.33}$$

The one step forecast is given by

$$Q_{t+1|t} = (1 - a - b)\bar{Q} + a z_t z'_t + b Q_t. \tag{2.34}$$

For $k \geq 2$, using the property

$$\mathbb{E}[\varepsilon_{t+k-1} \varepsilon'_{t+k-1} \mid \mathcal{F}_t] = \mathbb{E}[Q_{t+k-1} \mid \mathcal{F}_t], \tag{2.35}$$

the k-step-ahead forecast satisfies the recursion

$$Q_{t+k|t} = (1 - a - b)\bar{Q} + (a + b) Q_{t+k-1|t}. \tag{2.36}$$

From this approximation, the k-step forecast converges toward the unconditional correlation matrix at a rate determined by the persistence parameter $a + b$.

Iterating forward, the closed form solution becomes

$$Q_{t+k|t} = \bar{Q} + (a + b)^k \left(Q_t - \bar{Q} \right). \tag{2.37}$$

Thus, the forecast converges toward the unconditional correlation matrix $\bar{Q}$ as $k \to \infty$, once again provided that $a + b < 1$.

2.3.4 Evaluation Metrics

To determine the in sample fit of the estimated DCC-GARCH models, comparison is done by using information criterions and regression loss functions. These are Mean Square Error (MSE), Mean Absolute Error (MAE), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Since the models are estimated by (quasi) maximum likelihood, information criteria provide a good framework for evaluating goodness of fit while penalizing overfitted models.

The log-likelihood evaluates how well the model explains the observed data, with higher values showing better in sample fit. Overfitting occurs because larger models (more parameters) will always increase the likelihood. To discourage overfitting, information criteria adjust the likelihood by introducing a penalty term for the number of estimated parameters.

Mean Absolute Error is simply the average of the raw errors throughout the series, where everything is treated equal.

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t| \tag{2.38}$$

where T denotes the sample size, y_t is the actual value and $\hat{y}_t$ is the forecasted value.

Mean Square Error differs in the way it squares errors before calculating the average, which underpenalizes small deviations.

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 \tag{2.39}$$

where T denotes the sample size, y_t is the actual value and $\hat{y}_t$ is the forecasted value.

The two measurements together give a solid suggestion of the magnitude of the errors, and whether or not there are few culprits that severely hurt the model, or smaller cumulative errors.

The Akaike Information Criterion is defined as

$$AIC = -2\ell(\hat{\theta}) + 2k, \tag{2.40}$$

where $\ell(\hat{\theta})$ denotes the maximized log-likelihood and k is the total number of estimated parameters.

The Bayesian Information Criterion is given by

$$BIC = -2\ell(\hat{\theta}) + k \log T, \tag{2.41}$$

where T denotes the sample size. Compared to AIC, BIC imposes a stronger penalty on model complexity, particularly in large samples. Both criteria indicate that a lower AIC/BIC means a better model, making them good for comparative iterations of similar models with slight variations.

Suppose that a model A is deemed much better than a model B over a random test data set. How can one assure that random variations in the data do not unjustly favor A and yield deceptive MSE? To properly infer whether model A is statistically significantly better than model B, the Diebold-Mariano test is used[14]. Unlike straight comparisons of MAE/MSE values, a DM test will eliminate the risk of drawing a conclusion based on noise.

The test defines a loss differential at each time step:

$$d_t = L(e_{A,t}) - L(e_{B,t}) \tag{2.42}$$

where L is the loss function and $e_{i,t}$ is the forecast error of model i at time t. The null hypothesis of equal predictive accuracy is:

$$H_0 : \mathbb{E}[d_t] = 0 \tag{2.43}$$

The test statistic is computed as:

$$DM = \frac{\bar{d}}{\sqrt{\widehat{\mathrm{Var}}(\bar{d})}}, \tag{2.44}$$

where $\bar{d}$ denotes the sample mean of the loss differential series and $\widehat{\mathrm{Var}}(\bar{d})$ is an estimate of its variance.

A positive and statistically significant DM statistic indicates that model B gets lower forecast loss than model A, whereas a negative and significant statistic favors the A model. If the null hypothesis cannot be rejected, no model can be said to have better forecasting accuracy.

3 Modeling

3.1 Data Description

3.1.1 Assets and Markets

Broad market indices that are used in the study, like the S&P 500, aggregates many time-series into one. This allows for market-wide comparisons without the extreme dimensionality of covariance matrices that using the individual assets would entail. For the S&P 500 example, instead of creating 500 univariate GARCH models and feeding these into one huge DCC covariance matrix, one GARCH model is created based on the whole index. This is then paired with at least one other index, and the resulting DCC model will be a baseline without any signal from exogenous inputs.

The dataset is obtained from AP3 and consists of daily observations of major equity indices and market signals. The selected indices are used to assess the correlation dynamics between the S&P 500 and other global markets of interest to AP3.

3.1.2 Equity Indices

The following indices are included in the analysis, chosen to reflect geographically diverse markets without excessive implicit correlation between that could come from including indices from the same market:

Index	Full Name	Frequency
SPX	S&P 500	Daily
TPX	TOPIX	Daily
SX5E	Euro Stoxx 50	Daily
UKX	FTSE 100	Daily
OMX	OMX Stockholm 30	Daily
Sample period: 02/05/2011 - 07/01/2026		

Table 3.1: Equity indices used in the analysis

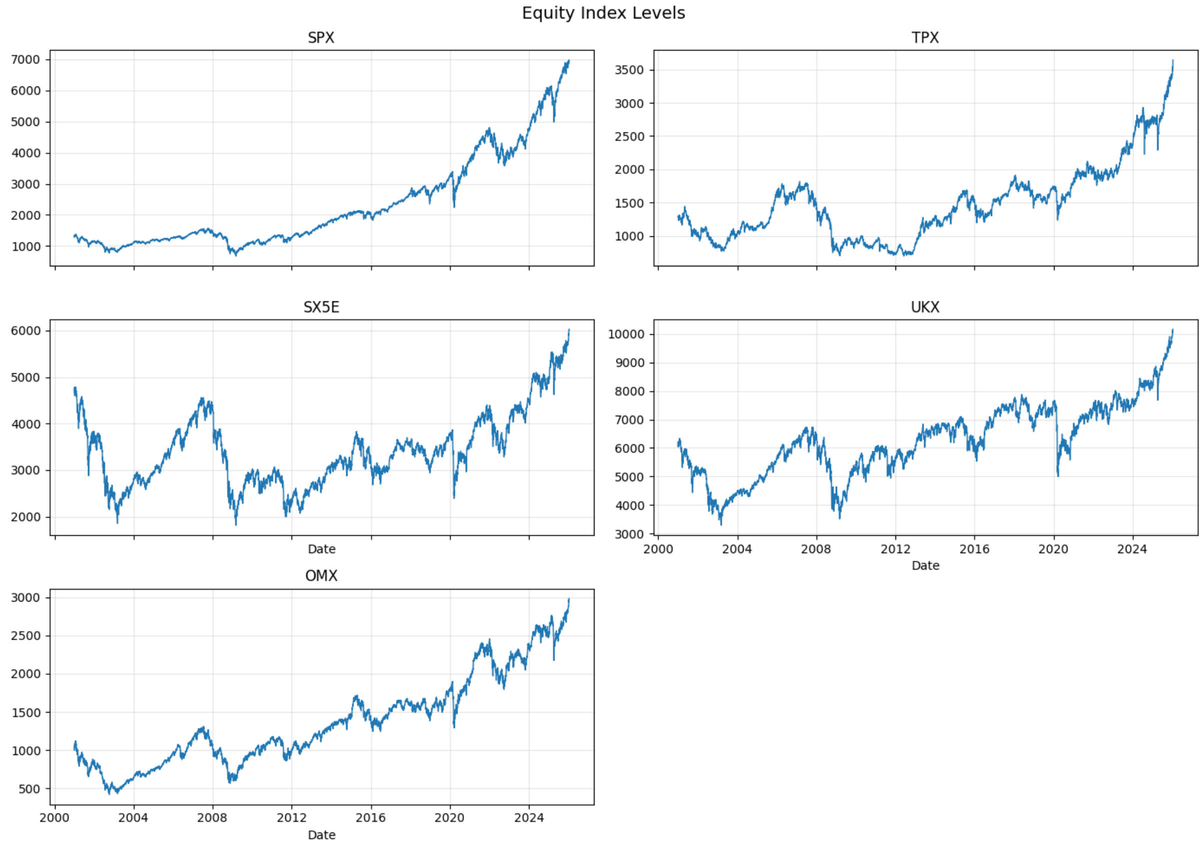


Figure 3.1: All equities used

Since the TPX index trades ahead of the U.S. market due to time zone differences, a one-day shift is applied to align it with the S&P 500 trading day.

To motivate the use of conditional volatility models, the mean-corrected returns are examined as in figure 3.3 of the selected indices. A rolling window is also estimated of the conditional standard deviation, $\hat{\sigma}_t = \sqrt{\widehat{\mathbb{V}}(r_t)}$, computed over a 21 day window (approximately one trading month) of SPX.

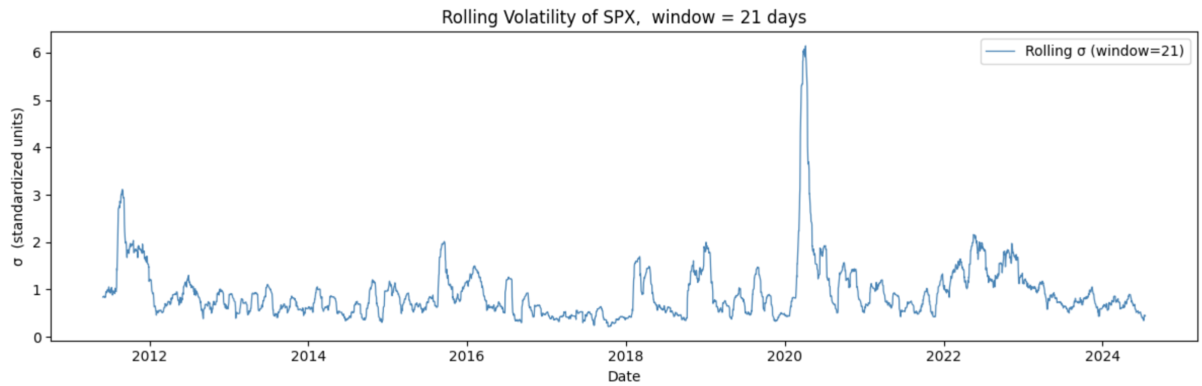


Figure 3.2: Rolling Volatility of SPX index with a 21 day window

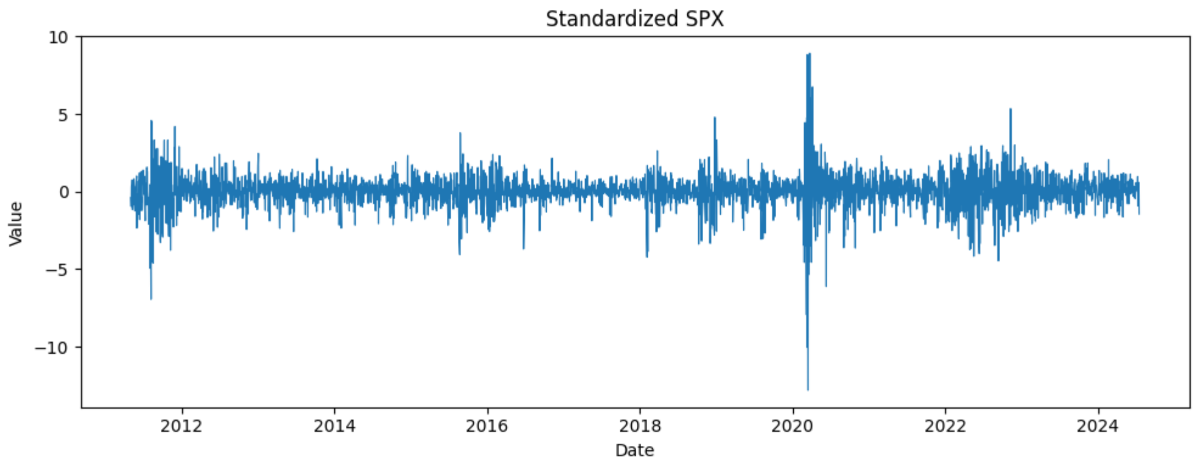


Figure 3.3: Mean-corrected daily returns of the SPX index

The graph shows clear periods of volatility clustering, as such the assumption of constant variance is violated, and therefore motivates the use of a GARCH model to capture the time-varying conditional volatility.

(Note that the data has been time shifted to adjust for exogenous signals, it now starts in 2011)

3.1.3 Preprocessing

All returns are transformed to logarithmic return series before they are fed into their respective GARCH model. The model itself is formulated on return series, not price levels [3]. This is largely done to make entirely different dimensions in price levels comparable. Say that the german DAX, trading at $\sim 25{,}000$ points, is to be compared with the S&P 500 at 7,000. To compare such series despite huge differences in price levels, there needs to be a price-agnostic representation of relative change. This is managed through the division of today's price P_t by yesterday's P_{t-1}, which leaves the price movement rather than the absolute levels themselves, as seen in the equation below:

$$r_t = \log(P_t) - \log(P_{t-1}) \tag{3.1}$$

3.1.4 Achieving a stable DCC model

Starting after pre-processing the returns, a univariate GARCH model was estimated for each individual asset. For simplicity, and since financial data tend to exhibit fatter tails, a Student's t-distribution was used when estimating the parameters. Both this step and the subsequent DCC estimation proved difficult in achieving a stable optimizer across all assets. As this was implemented ourselves, careful tuning of the bounds for the gradient descent optimizer was required, specifically using the `minimize` function in Python. Once a good balance between the parameters was achieved, ensuring that both persistent news and shock components had a well defined relationship, these were piped into the DCC model to estimate the correlation parameters. Initially, the model was fit on the full dataset provided by AP3, spanning from 2002 to 2026, in order to achieve as robust a model as possible and to capture broad market crises such as the 2008 financial crash and the COVID-19 pandemic. Since the model was implemented from scratch in Python, it was desirable to verify that our results were comparable to established libraries such as `rmgarch` in R. [15] Cross validating on the full dataset, both the likelihood of the estimated parameters and the visual output were nearly identical.

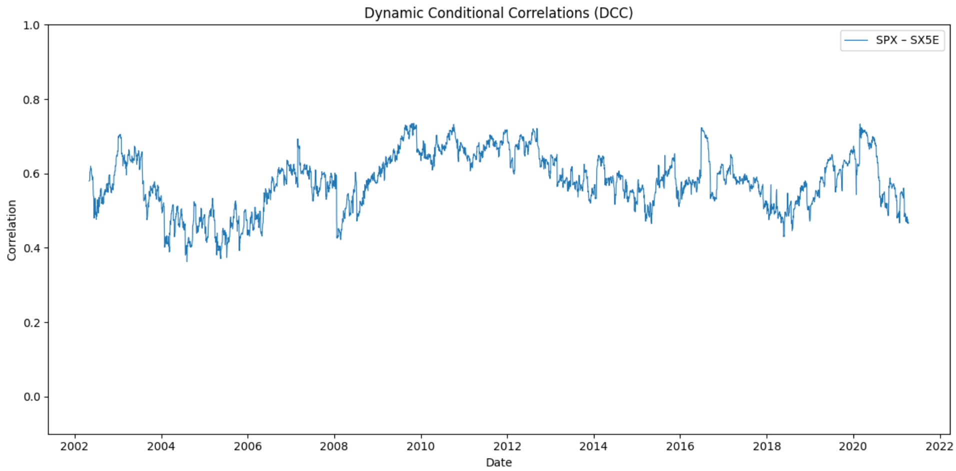


Figure 3.4: Single DCC model between SPX and SX5E

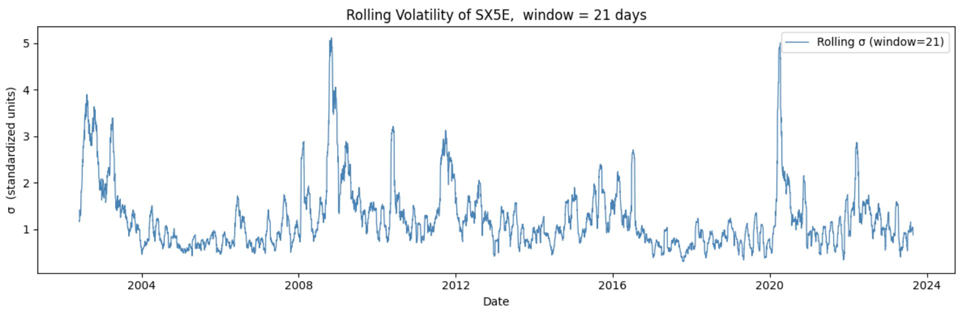


Figure 3.5: Rolling Volatility of SX5E index with a 21 day window

Clear volatility spikes in known market crises is shown for SX5E.

3.1.5 Market Signals

In addition to equity indices, several market signals are incorporated to capture broader financial conditions, investor sentiment and to explore the effect it has on the model:

Market Signal
Credit Spread
VIX Index
DIX Index

Table 3.2: Market signals kept in the analysis

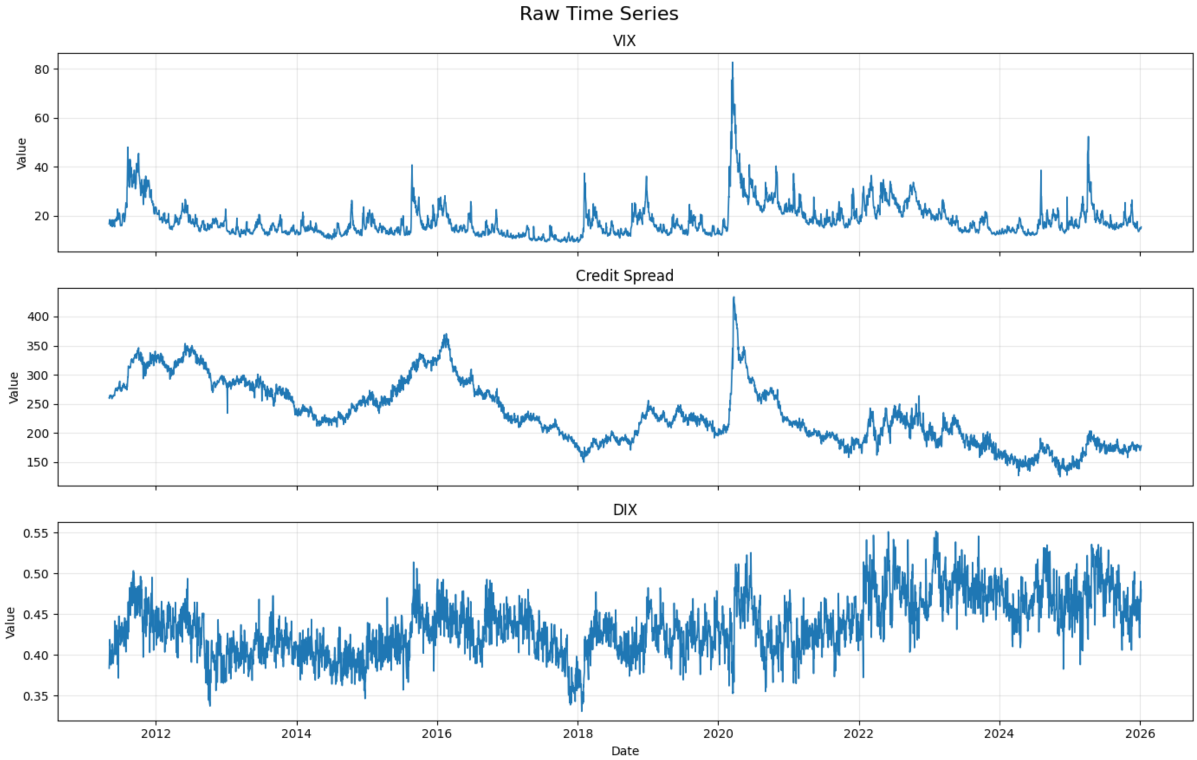


Figure 3.6: Timeseries of the exogenous signals

Out of all the trialed signals, these are the top three candidates. A credit spread index between a corporate bond yield and the US 10-year treasury is included as an exogenous variable in the GARCH-X model to account for variations in credit market conditions. The VIX index is incorporated into the GARCH-X specification to capture volatility expectations. The DIX index is used as an additional indicator, reflecting dark pool activity and institutional investor sentiment. Increased institutional buying can hypothetically dampen the conditional variance, hence why it enters into the GARCH-X specification alongside the other signals.

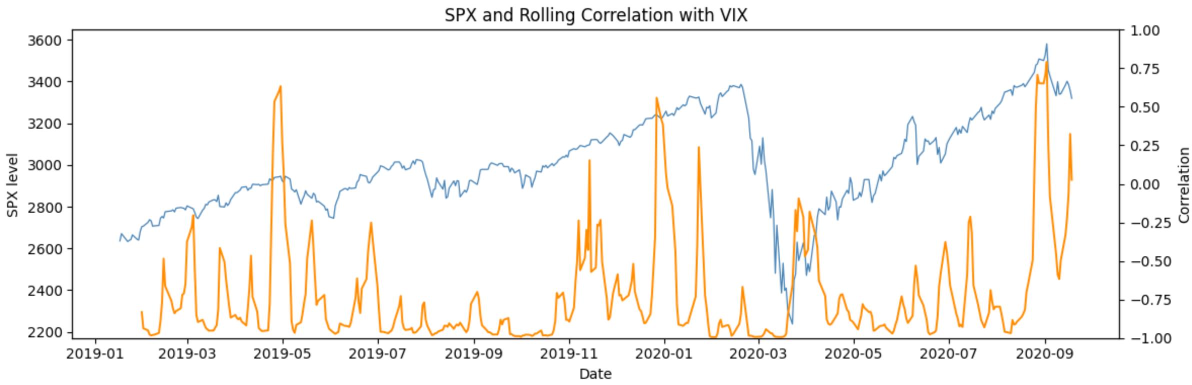


Figure 3.7: SPX and 10 Day Rolling Correlation with VIX index

The rolling correlation for VIX is mostly negative for the sample, but with sharp spikes during periods of market stress. This is a documented phenomenon where falling stock prices often coincide with a rising VIX [16]. The relationship also appears highly nonlinear and episodic rather than stable over time.

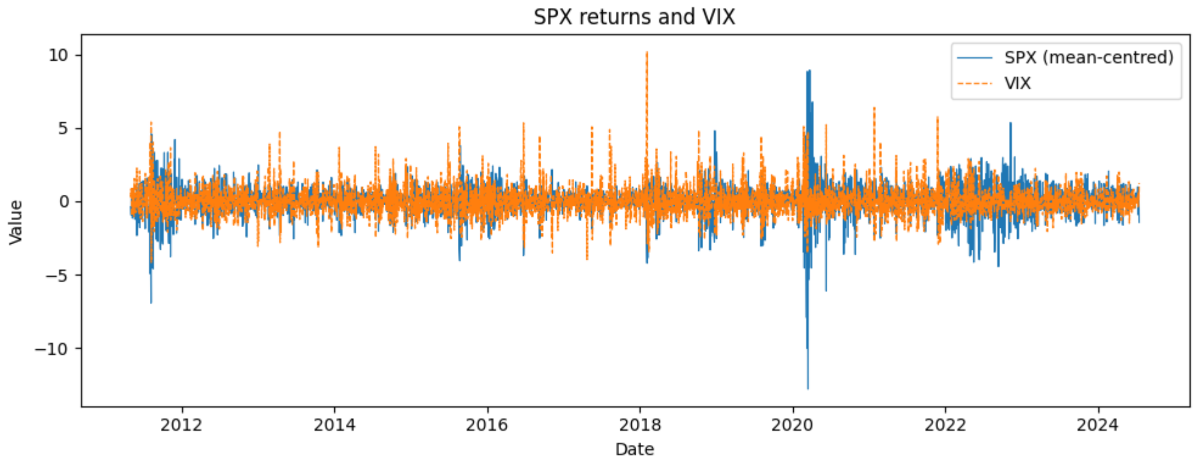


Figure 3.8: VIX index overlaid SPX

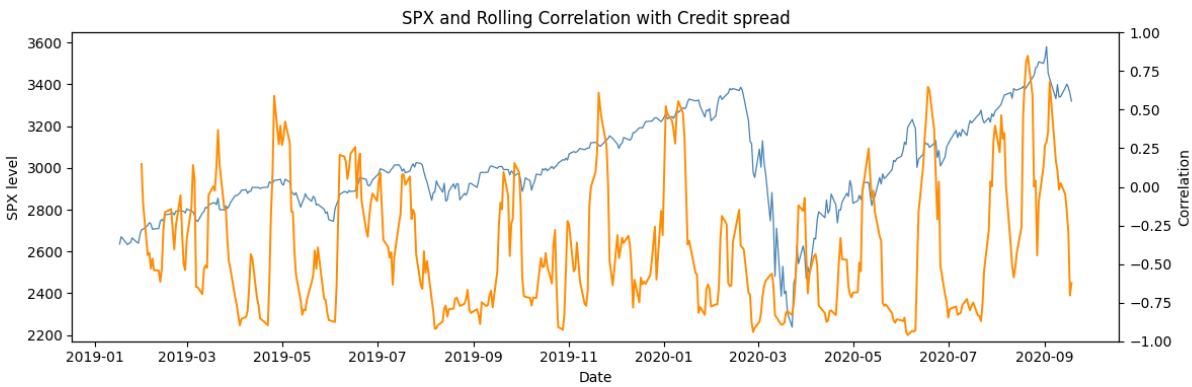


Figure 3.9: SPX and 10 Day Rolling Correlation with Credit Spread

For credit spreads the rolling correlation is predominantly negative and becomes more pronounced during stressed periods, particularly around the 2020 market decline, much like VIX but with greater variation over time.

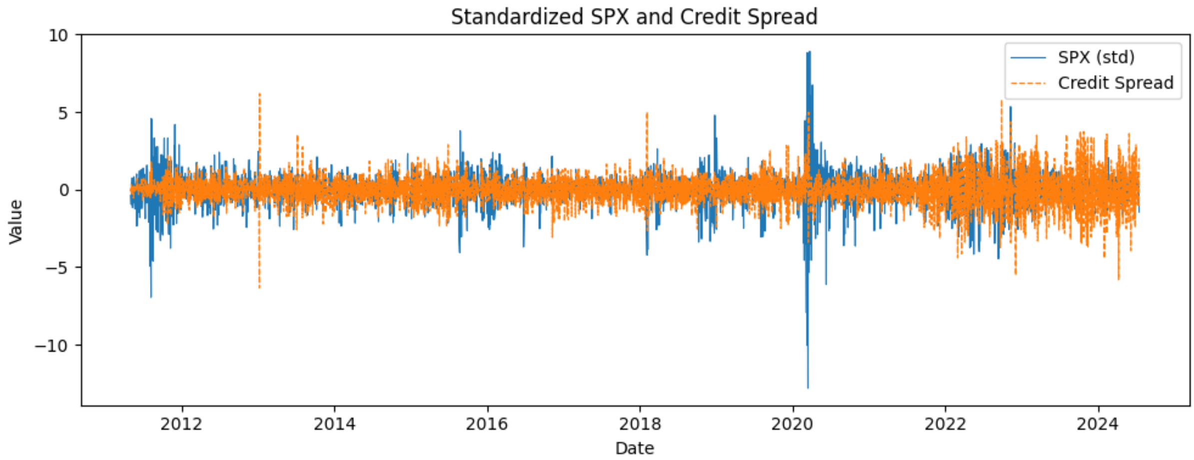


Figure 3.10: Credit Spread index overlaid SPX

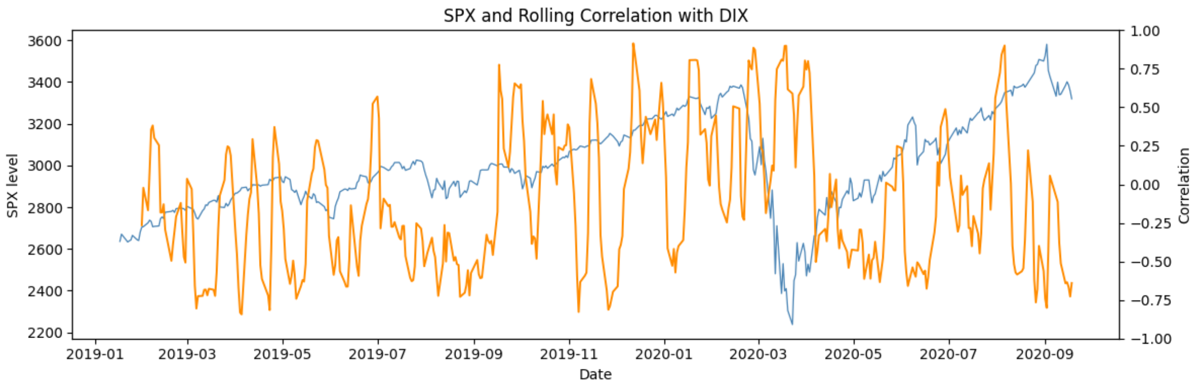


Figure 3.11: SPX and 10 Day Rolling Correlation with DIX index

The rolling correlation between SPX and DIX fluctuates a lot between positive and negative values throughout the sample. Compared to VIX and credit spreads, the relationship appears less structured and more unstable over time. Interestingly, while often choppy, all three signals appear to be uncorrelated overall. It's either negatively or positively correlated throughout.

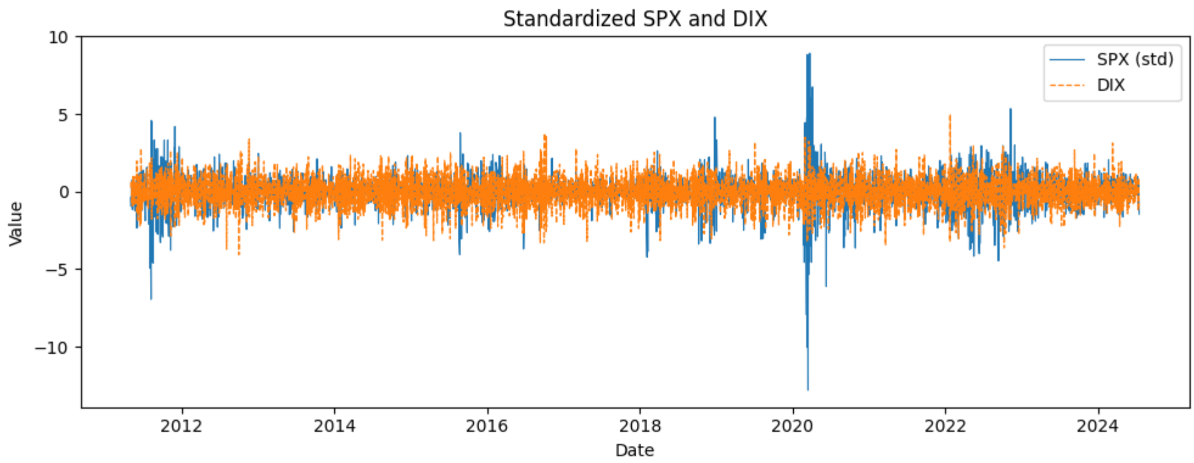


Figure 3.12: DIX index overlaid SPX

3.1.6 Expanding current model

Now further expanding the model to incorporate the remaining assets and to include the exogenous inputs required restricting the time window, as some of the exogenous time series were shorter than others. Skipping this alignment would have forced a choice between either filling in the missing dates or removing them entirely, both resulting in more approximate parameter estimation for the exogenous signal. With the data correctly aligned and an additional parameter introduced to the model, the parameters were re-estimated. The number of iterations reported reflects how many steps the optimizer took before converging to a minimum, or equivalently, a maximum when the log-likelihood is negated, as is standard practice in optimization. [17]

Index	ω	α	β	$\alpha + \beta$	ν	**LL**
SPX	0.0221	0.1557	0.8314	0.9871	6.19	-4215.87
SX5E	0.0296	0.1158	0.8727	0.9885	5.09	-5037.20
OMX	0.0206	0.0777	0.9065	0.9842	7.22	-4885.99
TPX	0.0427	0.0968	0.8724	0.9692	5.64	-4927.60
UKX	0.0340	0.1351	0.8359	0.9710	5.22	-4185.09

Table 3.3: GARCH(1,1) parameter estimates for each equity index (11 iterations)

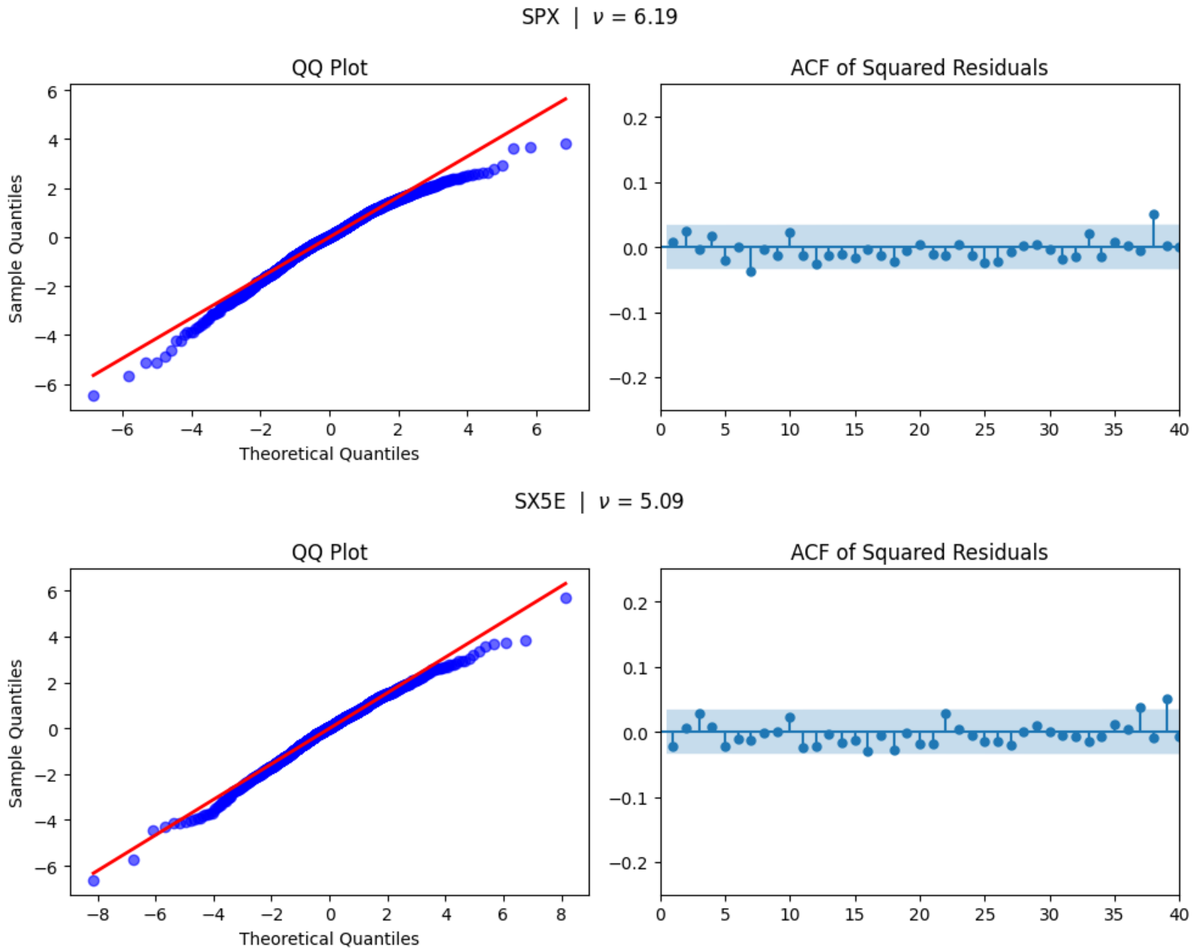


Figure 3.13: QQ plots and ACF of squared residuals (1/2)

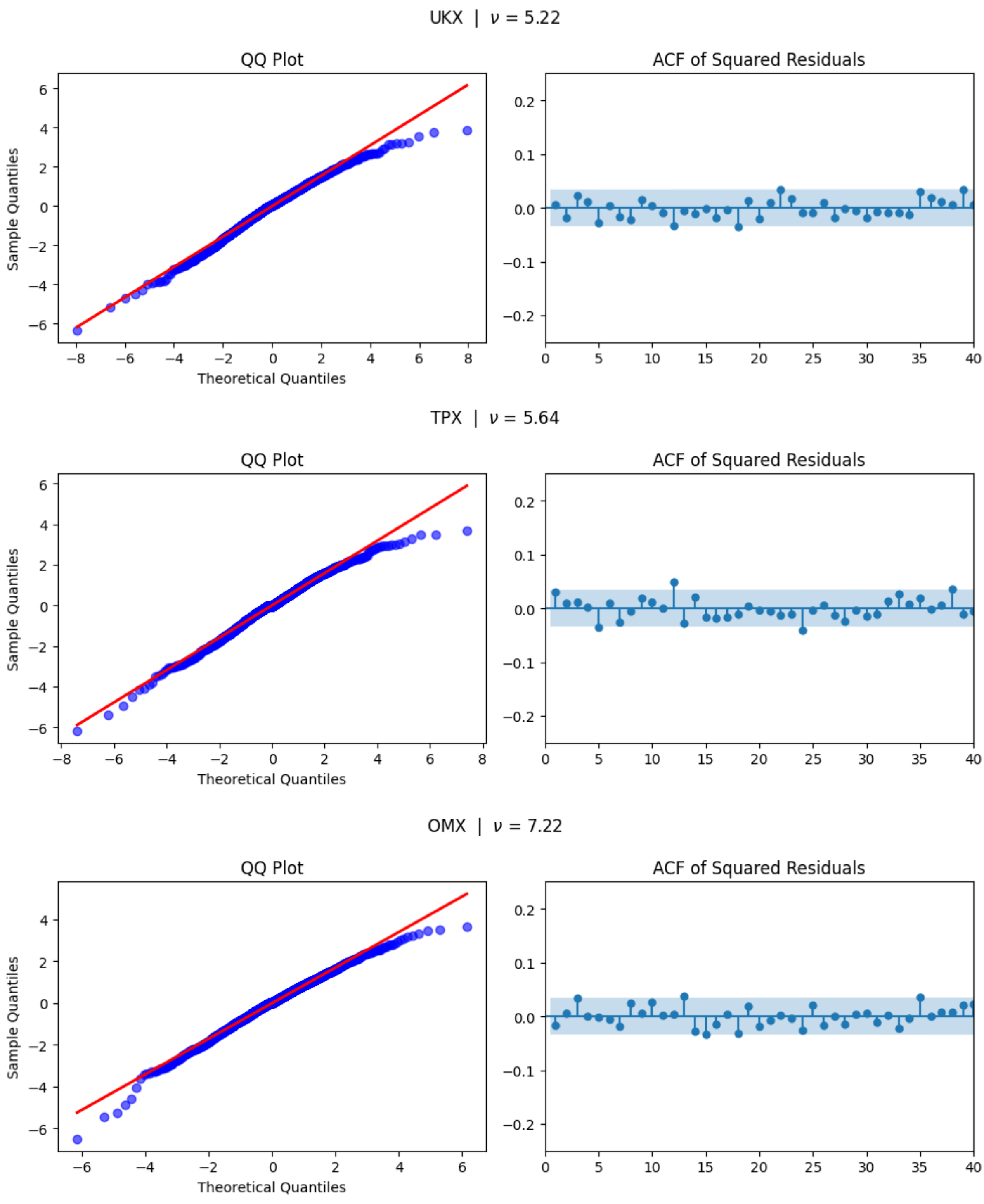


Figure 3.14: QQ plots and ACF of squared residuals (2/2)

The QQ plots in figures 3.13 - 3.14 shows that the standardized residuals follows the student-t distribution, however some slight deviation can be seen in the tails, suggesting that the estimated degrees of freedom ν could be slightly conservative. The ACF plots show no significant remaining correlation, indicating that the GARCH specification captures the volatility clustering in the data enough.

A noticeable amount of choppiness can be seen in the DCC correlation plots below, now that the number of dimensions has increased. As the amount of correlations rise, so does the cross product of standardized residuals, which may contribute to obscuring pairwise correlation structure in estimations.

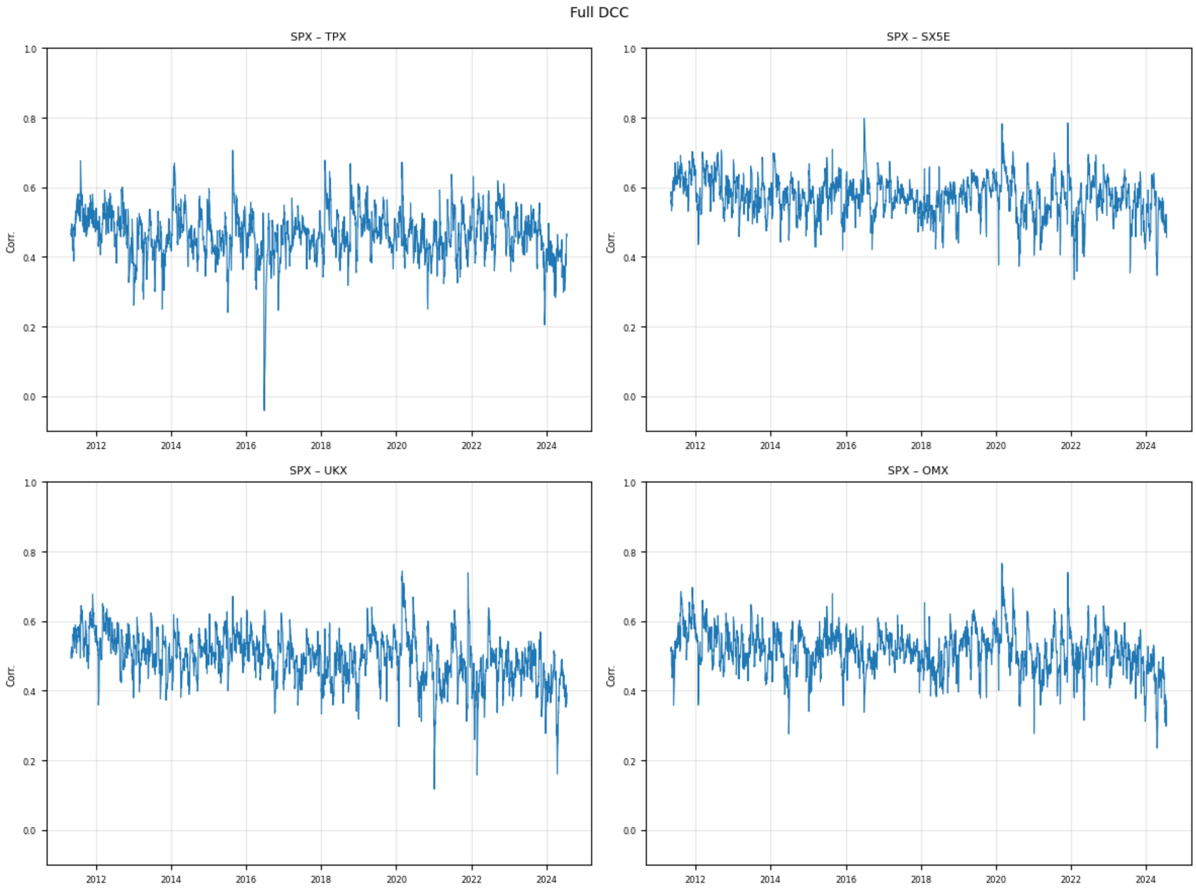


Figure 3.15: Full correlation matrix of equity indices paired with SPX

Parameter	**Value**
a	0.033273
b	0.888250
$a+b$	0.921523

Table 3.4: DCC Model Parameters

LogL	**AIC**	**BIC**
-3368.90	6741.81	6754.10

Table 3.5: Benchmark GARCH+DCC model fit

Signal	Pair	MSE Base.	MSE Exo.	MAE Base.	MAE Exo.	DM(MSE)	DM(MAE)
VIX	SPX–TPX	2.18322	2.76225	0.76890	0.77609	−0.673	1.647*
	SPX–SX5E	2.43440	2.08657	0.82269	0.79567	2.106**	3.146***
	SPX–UKX	1.97303	1.73385	0.72190	0.69350	2.471**	4.221***
	SPX–OMX	2.14843	1.86145	0.79612	0.76701	1.054	4.697***
				Fit: LogL = -3584.80,		AIC = 7173.59,	BIC = 7185.88
DIX	SPX–TPX	2.18322	2.18081	0.76890	0.76918	−0.869	−0.624
	SPX–SX5E	2.43440	2.31686	0.82269	0.81934	0.442	1.061
	SPX–UKX	1.97303	1.88000	0.72190	0.71995	2.136**	2.881***
	SPX–OMX	2.14843	2.09372	0.79612	0.79435	1.197	−0.111
				Fit: LogL = -3373.86,		AIC = 6751.72,	BIC = 6764.01
Credit Spread	SPX–TPX	2.18322	2.14668	0.76890	0.76391	0.943	1.008
	SPX–SX5E	2.43440	2.37698	0.82269	0.81867	2.385**	2.797***
	SPX–UKX	1.97303	1.77808	0.72190	0.70871	3.179***	4.463***
	SPX–OMX	2.14843	2.02513	0.79612	0.78801	1.892*	1.808*
				Fit: LogL = -3345.38,		AIC = 6694.75,	BIC = 6707.04

Base. refers to benchmark GARCH+DCC; Exo. refers to the exogenous model.
Positive Diebold–Mariano statistics favour the exogenous model.
*, **, *** denote significance at the 10%, 5%, and 1% levels.

Table 3.6: Forecast performance of exogenous models relative to the base model

3.1.7 Pairwise model approach

One potential issue with having a larger covariance matrix that covers many pairs is their inherent heterogeneity. Drawing conclusions about the effectiveness of e.g. the US-derived VIX as an external regressor on the full model is one thing, but does that translate to say Japan or even smaller markets on different continents entirely? It is likely that some pairs see strong effects while others have zero or even negating properties that average out the potential forecasting edge of the model.
In short, the suspicion that the exogenous signal effects are not uniform across the globe is investigated. So, rather than fitting a single multivariate DCC model across all assets at the same time, the model is now estimated pairwise one DCC for each pair relative to SPX. Each pair shares the same univariate GARCH(1,1) base as before, but the dynamic correlation is estimated separately for every SPX asset.

Worth noting is how the forecaster operates. At each point in the test set, the model produces a 7-step-ahead forecast, that is, it predicts the correlation for each of the next 7 trading days using only information available up to that point. The window then advances by one day and the process repeats. Averaging the forecast errors across all such windows gives the reported MSE and MAE figures, which also applies for the table above (table 3.6). So, at no point does the model have access to future data when generating a forecast, ensuring that the evaluation is free of lookahead bias and reflects genuine out-of-sample predictive performance.

3.2 Pairwise Evaluation

3.2.1 GARCH-VIX + DCC

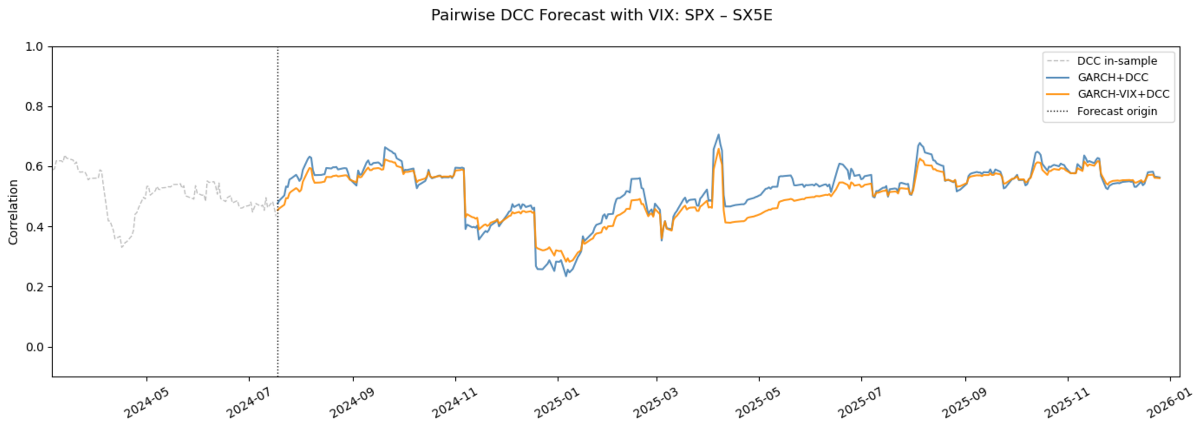


Figure 3.16: DCC forecast with VIX: SPX - SX5E

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	2.48274	0.81867	-2612.81	5229.61	5241.83
GARCH-VIX+DCC	**2.35850**	**0.77962**	**-2394.97**	**4793.93**	**4806.15**

Table 3.7: SPX-SX5E forecast errors and model fit, VIX variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-SX5E	1.0071	0.3139	4.6674	0.0000***

Table 3.8: DM-test GARCH+DCC vs GARCH-VIX+DCC, SPX-SX5E.

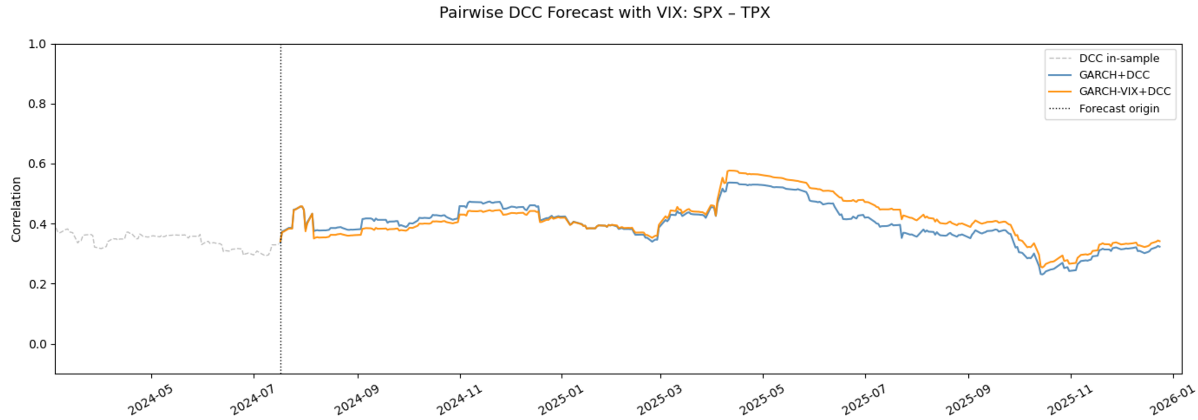


Figure 3.17: DCC forecast with VIX: SPX - TPX

Model	**Avg MSE**	**Avg MAE**	**LogL**	**AIC**	**BIC**
GARCH+DCC	**2.29595**	**0.76922**	**-2889.89**	**5783.79**	**5796.00**
GARCH-VIX+DCC	4.14454	0.84445	-2897.56	5799.11	5811.33

Table 3.9: SPX-TPX forecast errors and model fit, VIX variant

Pair	**DM (MSE)**	**p-val**	**DM (MAE)**	**p-val**
SPX-TPX	-2.5911	0.0096***	-3.1017	0.0019***

Table 3.10: DM-test GARCH+DCC vs GARCH-VIX+DCC, SPX-TPX.

3.2.2 GARCH-DIX + DCC

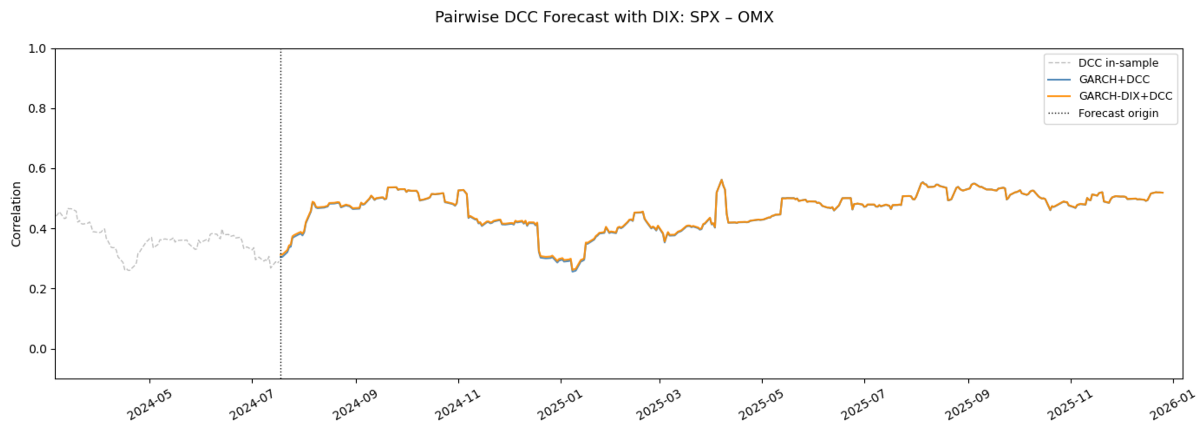


Figure 3.18: DCC forecast with DIX: SPX - OMX

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	2.17062	0.78123	-2800.97	5605.95	5618.16
GARCH-DIX+DCC	**2.16875**	**0.78059**	**-2798.97**	**5601.94**	**5614.15**

Table 3.11: SPX-OMX forecast errors and model fit, DIX variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-OMX	1.4258	0.1539	-4.6440	0.0000***

Table 3.12: DM-test GARCH+DCC vs GARCH-DIX+DCC, SPX-OMX.

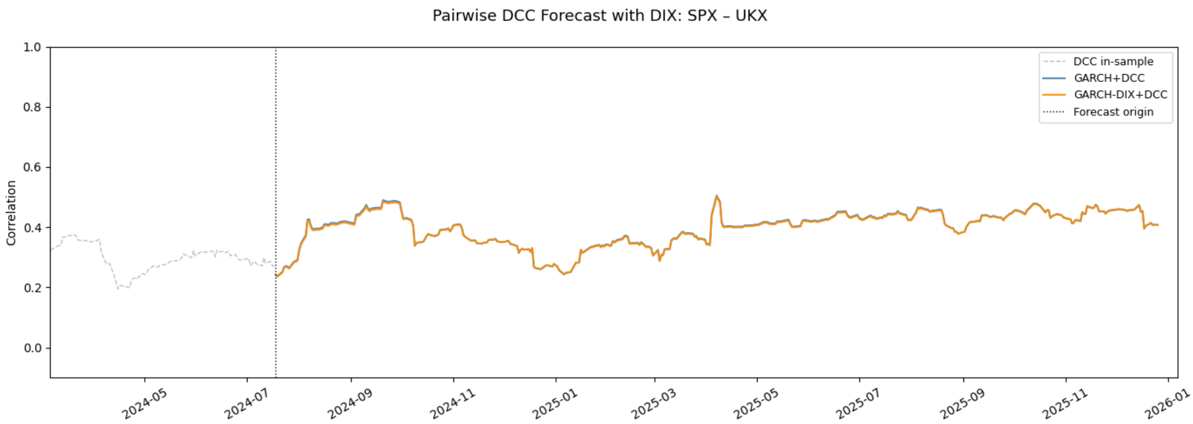


Figure 3.19: DCC forecast with DIX: SPX - UKX

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	**1.99573**	0.69192	-2779.66	5563.31	5575.53
GARCH-DIX+DCC	2.01087	**0.69114**	**-2779.58**	**5563.17**	**5575.38**

Table 3.13: SPX-UKX forecast errors and model fit, DIX variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-UKX	1.0863	0.2774	7.9659	0.0000***

Table 3.14: DM-test GARCH+DCC vs GARCH-DIX+DCC, SPX-UKX.

3.2.3 GARCH-Credit Spread + DCC

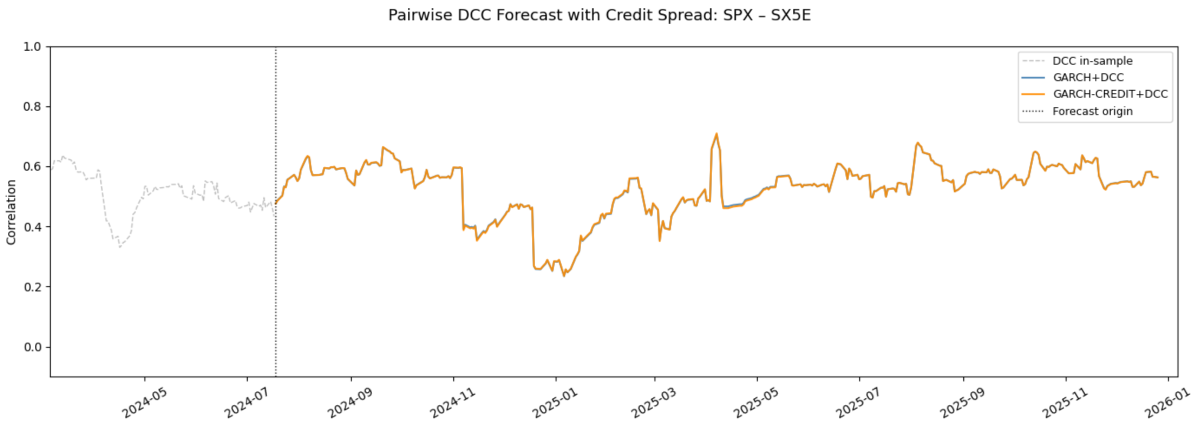


Figure 3.20: DCC forecast with Credit Spread: SPX - SX5E

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	2.48274	0.81867	-2612.81	5229.61	5241.83
GARCH-CREDIT+DCC	**2.43667**	**0.81297**	**-2599.82**	**5203.64**	**5215.86**

Table 3.15: SPX-SX5E forecast errors and model fit, Credit Spread variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-SX5E	0.8429	0.3993	0.5181	0.6044

Table 3.16: DM-test GARCH+DCC vs GARCH-CREDIT+DCC, SPX-SX5E.

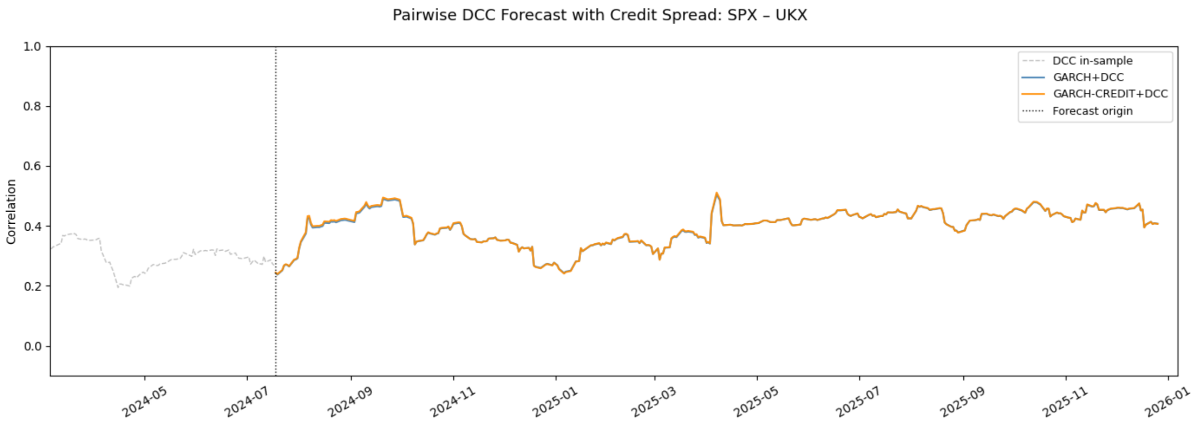


Figure 3.21: DCC forecast with Credit Spread: SPX - UKX

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	**1.99573**	0.69192	-2779.66	5563.31	5575.53
GARCH-CREDIT+DCC	2.01147	**0.68932**	**-2769.82**	**5543.65**	**5555.86**

Table 3.17: SPX-UKX forecast errors and model fit, Credit Spread variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-UKX	-0.8885	0.3743	-4.8899	0.0000***

Table 3.18: DM-test GARCH+DCC vs GARCH-CREDIT+DCC, SPX-UKX.

The remaining test figures and tables can be found in the appendix, since they are well represented in those already presented above.

3.3 Forecasting performance and exogenous effect

This thesis explores the significance of exogenous market signals for volatility and correlation forecasting using DCC-GARCH. Results show that no combination is a blueprint for every combination of equity markets. Some of the market signals improve forecasting performance, such as VIX for the SPX-SX5E pair, see figure 3.16, while others have negligible or even worsening effects, as observed from the figure and tables of 3.17.

Credit spreads show consistent improvements among the trialed signals. As shown in Table 3.6 both the information criteria and out-of-sample forecast metrics improve relative to the baseline DCC-GARCH model. The multivariate specification got the highest log-likelihood of all tested models (LogL $= -3345.38$ compared to -3368.90 for the baseline model), while at the same time lowering both MSE and MAE for its forecast evaluations, although statistical significance is concentrated primarily in the European market pairs.

For the pairwise models the gains are weaker and uneven. It improves the loglikelihood for all, although marginally, however its only for SPX-OMX, SPX-UKX that a small statistical significant difference in MAE can be seen. So as a result, the visual differences between the baseline and exogenous forecasts remain relatively small overall.

DIX show weaker and less consistent effects on the models. As seen in Table 3.6, the full multivariate model with DIX as an exogenous signal only produces marginal changes relative to the baseline. The small improvements observed are, for the most part, statistically insignificant, and for the one case that is significant, SPX-UKX, the MSE and MAE drop by 0.11754 and 0.00195 respectively. Similar results are noted for the pairwise models. In Tables 3.11-3.12, SPX-OMX shows a slight improvement in MAE however, the DM test still favors the baseline over the full forecasting period. As the results are modest, this may suggest that dark-pool activity contains limited systemic information for forecasting correlation dynamics at such a broad level.

VIX produces interesting results as its effect is strongly dependent on which pairs are evaluated. In the pairwise setup it yields both positive and negative impacts on residual analysis, whilst bettering model fit in all but the US-Japan pair. In the full multivariate model, VIX worsens the model fit once again and the MSE & MAE for US-Japan. The other market pair forecasting performances are improved quite substantially, both squared and absolute errors. This supports the claim that the signals aren't uniformly applicable on a global scale and might instead require tailoring to some degree. It stands out how abnormally large the MSE became for SPX-TPX and SPX-UKX. This could be due to divergences in those two markets compared to the US-centric VIX index, but it very well may be due to poor exogenous model addition.

It is apparent that there are seemingly inconsistent combinations where a clearly worse average error is still deemed better in the DM-test, and often significantly so. This could be due to MSE & MAE being aggregates while the DM-test value comes from a time series constructed off the sequence of loss differentials through the set like this series:

$$d_t = L(e_1, t) - L(e_2, t) \tag{3.2}$$

There could potentially be few but extreme values that hike up the average, even though the majority of subtractions throughout the DM-series points to the other conclusion. For example:

Time	Baseline Model Error	Exogenous Model Error	Loss Differential (d_t)
1	1.0	0.8	0.2
2	1.0	0.7	0.3
3	1.0	0.8	0.2
4	1.0	3.5	-2.5

Table 3.19: Example of how Diebold-Mariano and MAE/MSE can lead to different conclusions

In this example, the exogenous model performs better in three out of four periods, producing positive loss differentials in most observations. However, the single large forecasting error in time sample 4 substantially increases the average MAE/MSE of the exogenous model, illustrating how the aggregates may disagree with the favoring of the Diebold-Mariano test.

3.4 Heterogeneity across markets

When correlations across heterogeneous global markets are estimated jointly, a US options derived fear index may genuinely be relevant for certain pairs, especially the ones with strong ties to the US. But these also get diluted by pairs where it carries little to no signal. So although several model specifications show some improvements the magnitude of these differences is generally small. So the question arises of whether the observed improvements are economically or statistically significant, or merely the result of random variation in the data. In many cases, the performance remains very close to the baseline DCC-GARCH model, suggesting that the added complexity from some exogenous variables does not consistently translate into meaningful predictive gains. This may also partially stem from the Efficient Market Hypothesis [18], meaning public information is incorporated in prices and market dynamics very quickly, leaving little room for forecasting betterment.

3.5 Statistical Significance

Conflicting takeaways sometimes arise from studying the MSE and MAE. It can at times be seen that the MSE becomes worse with exogenous input, while the MAE lowers at the same time. This means that the average prediction accuracy improves all the while becoming more sensitive to larger shocks. Cases can be made for both residual metrics being the more important one. The average will arguably matter more often but conversely, modeling tail events well is where most of the practical value lies from a risk management perspective. For many institutions preserving capital under stressed market conditions is mandated with higher priority than generation of higher returns.

3.6 Limitations

Parameter estimation and optimizer selection proved to be a bit troublesome to implement. A recurring issue was the gradient descent optimizer converging early to a parameter bound rather than to the true minimum, a common issue in constrained numerical optimization [19]. Achieving a stable solution required tuning of initial values, parameter bounds, and optimizer type, something that also increased as the covariance matrix grew when including additional assets, increasing estimation complexity and parameter interactions [9]. The model with multiple assets parameters behaved differently from the pairwise case, and the final estimates likely reflect a compromise between the imposed bounds and the true optimum, at some cost to accuracy. Another challenge was aligning time-series that had time-zone differences between them. It was decided that the simplest and probably most effective solution is to use one period lagged returns for e.g. Asian daily prices so that American events that transpire after Asian market close are correctly accounted for in the alignment.

3.7 Minimum-variance Portfolios

Although the primary objective of this study is correlation forecasting and comparing internal models rather than portfolio optimization, an illustrative portfolio experiment was conducted. This was done in order to evaluate potential improvements of the whole DCC-GARCH framework which was implemented from scratch.
Using the forecasted conditional covariance matrices from the pairwise models, a rolling two-asset minimum-variance portfolio was constructed. Portfolio weights were recalculated at each forecast date using the one-step forecasts from the DCC-GARCH. For comparison, a corresponding minimum-variance portfolio based on a constant correlation assumption was also created, together with a naive equal-weight benchmark.

To avoid unrealistic concentration between the assets, portfolio weights were clipped to the interval $[0.2, 0.8]$ and smoothed using an exponential moving average. This can be seen in some graphs where it flatlines across either of those axis for a period of time. It still felt appropriate since the purpose of the experiment isn't to identify an optimal investment strategy, but rather to find out how dynamic correlation forecasts can influence portfolio diversification. In the following segment the Sharpe ratio is defined as how much return is achieved for each unit of portfolio risk, with the formula being:

$$SR = \frac{R_p - R_f}{\sigma_p} \tag{3.3}$$

Where, R_p = portfolio return, R_f = risk free rate, σ_p = portfolio standard deviation

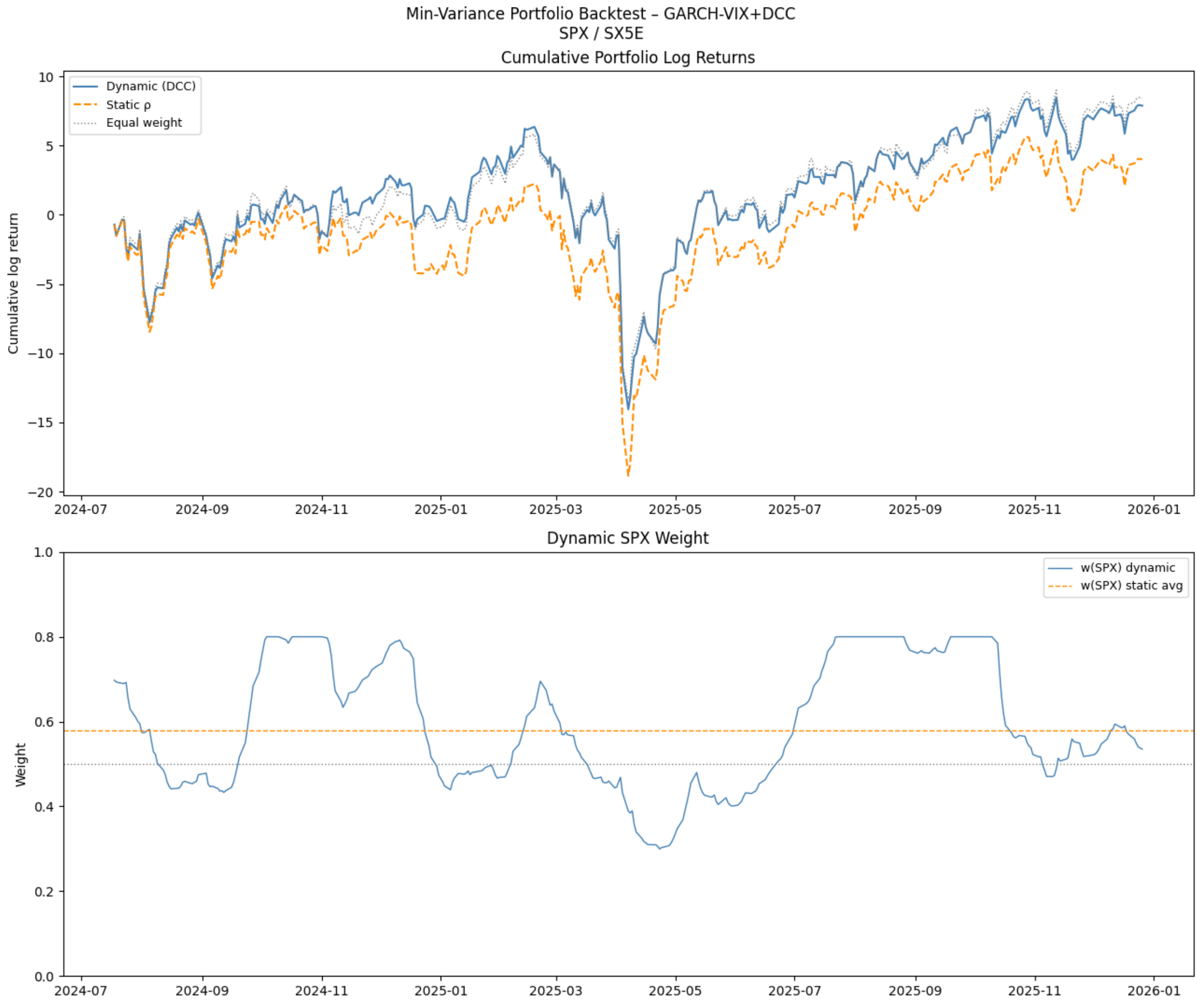
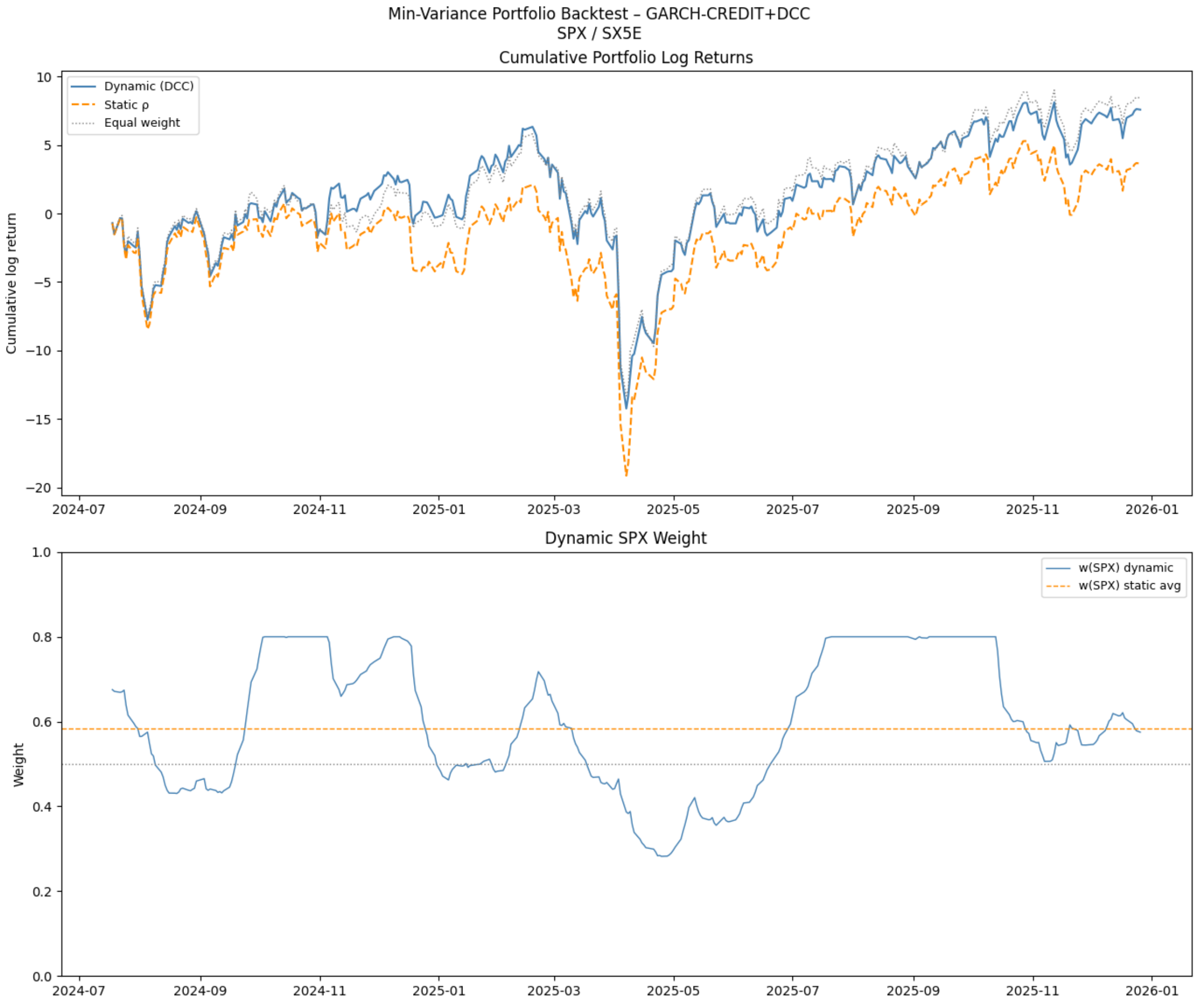


Figure 3.22: Dynamic portfolio weights with VIX: SPX-SX5E

Metric	Base	VIX	Δ
Dynamic variance	0.7340	0.7332	-0.0008
Variance reduction (%)	7.69	8.11	+0.42
Sharpe ratio (dynamic)	0.399	0.403	+0.003
Sharpe ratio (static)	0.179	0.196	+0.017
Sharpe ratio (equal weight)	0.431		

Table 3.20: Portfolio comparison for SPX-SX5E with VIX

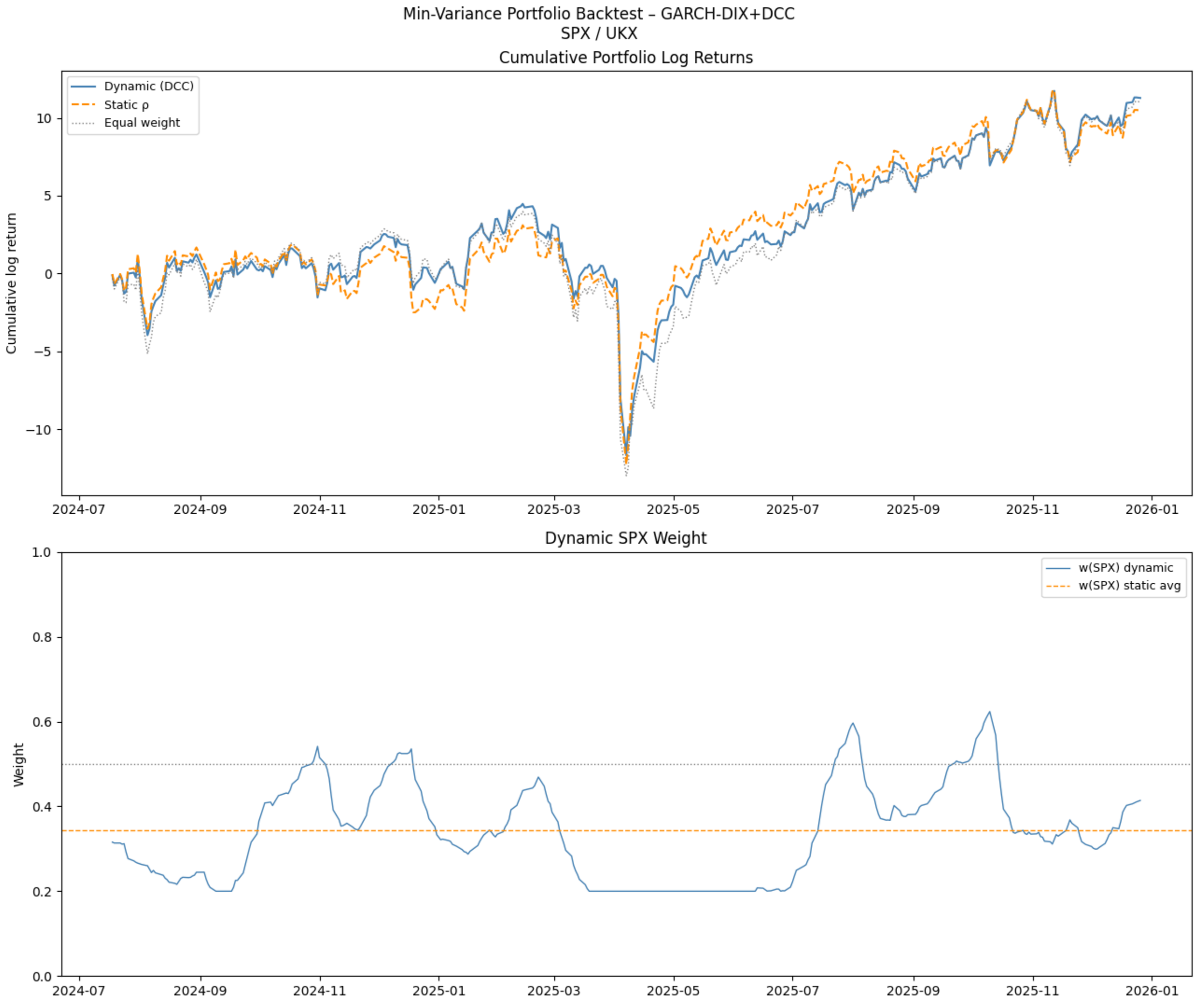
Both the base and VIX model show a reduction in variance, where VIX slightly outperforms the base model. The VIX model also achieves somewhat higher dynamic and static Sharpe ratios.



Metric	Base	Credit	Δ
Dynamic variance	0.7340	0.7338	-0.0002
Variance reduction (%)	7.69	7.80	+0.11
Sharpe ratio (dynamic)	0.399	0.387	-0.012
Sharpe ratio (static)	0.179	0.178	-0.001
Sharpe ratio (equal weight)	0.431		

Table 3.21: Portfolio comparison for SPX-SX5E with Credit Spread

Similar results are observed for the magnitude of variance reduction as in the previous case, although the difference between the base and Credit Spread models is somewhat smaller.



Metric	Base	DIX	Δ
Dynamic variance	0.4602	0.4599	-0.0003
Variance reduction (%)	2.22	2.24	+0.01
Sharpe ratio (dynamic)	0.725	0.727	+0.003
Sharpe ratio (static)	0.666	0.668	+0.002
Sharpe ratio (equal weight)	0.665		

Table 3.22: Portfolio comparison for SPX-UKX with DIX

As seen in this third pair, the variance reduction is more modest in both the base model and the built out DIX one, compared to the two previous pairs. The Sharpe ratios are also higher. One could guess that the less severe downturn from the tariff news in early April corresponds to less variance reduction as well as higher Sharpe ratios.

The rest of the backtested portfolios and their graphs can be found in the appendix at the bottom of the study.

4 Conclusion

This study examined whether exogenous market signals can improve the forecasting accuracy of a DCC-GARCH model applied to global equity index correlations. The results suggest that while statistically detectable, they are modest in magnitude. They are also highly dependent on both signal choice and the dimensionality of the model.

Credit spreads came out as the most consistently beneficial input, producing improvements in both in sample fit and out of sample forecast errors across several index pairs. This is consistent with their established role as leading indicators of systemic stress. VIX produced the most nuanced results, it improves pairwise correlation forecasts for US-adjacent markets such as SPX–SX5E, yet worsens the full multivariate model, where the full model log-likelihood goes from -3369 to -3584 when VIX is included. DIX showed the weakest and least consistent effects throughout, suggesting that dark pool activity carries limited information for broad correlation dynamics at a global scale.

A recurring theme across all signals is that the extent of improvement is small. Even the most favorable results rarely produce large reductions in MSE or MAE, and in many cases the differences are not statistically significant under the DM test. This can be pointed to a known property of financial return data, the DCC-GARCH baseline, driven by its own lagged correlation structure with high persistence, already captures much of the dynamics that exogenous signals might otherwise explain.

Another finding is the divergence between the full multivariate and pairwise DCC model. Exogenous signals that show negligible or negative effects in the full model can produce improvements in pairwise specifications, and vice-versa. This suggests that model dimensionality itself can tone down exogenous information, as heterogeneous pair level effects might be averaged out within the larger covariance matrix.

In summary, exogenous market signals can provide marginal improvements when applied selectively and in the right model structure. No one signal constitutes a reliable universal improvement, and the added model complexity is only justified for specific pairs and signals.

5 Future Work

There are several ways this study could be extended. One interesting approach would be introducing regime-switching or other Copula-based DCC models. This may lead to better modeling of structural changes in the markets, stress, contagion and identification of inflection points.

The set of exogenous signals could be bigger and more refined at a market level. Instead of comparing the pairwise models to one big global benchmark, the dimensionality could be kept high but across similar markets, tied together either geographically or economically. Perhaps an Asian DCC-GARCH, a different one for Europe, a Nordic one. The US could have its own across different indices.

Furthermore, a more expansive study of actually useful DCC-X signals could be interesting as they proved difficult to find in this study.

One could perform the parameter estimation recursively not only through the correlation equation but all the way back through the variance equation iteratively.

Bibliography

[1] John C. Hull. *Options, Futures, and Other Derivatives.* Pearson, 11 edition, 2022.

[2] Kristin J. Forbes and Roberto Rigobon. No contagion, only interdependence: Measuring stock market comovements. *Journal of Finance*, 57(5), 2002.

[3] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 1986.

[4] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[5] James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 1989.

[6] Peter R. Hansen and Asger Lunde. A forecast comparison of volatility models: Does anything beat a garch(1,1)? *Journal of Applied Econometrics*, 20(7), 2005.

[7] Zefang Song, Lingjun Chen, and Wenlin Huang. Estimation and application of garch-x model based on high-frequency data. *American Journal of Industrial and Business Management*, 15(2), 2025.

[8] Robert Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 1982.

[9] Robert Engle. Dynamic conditional correlation: A simple class of multivariate garch models. *Journal of Business & Economic Statistics*, 2002.

[10] Tim Bollerslev and Jeffrey M. Wooldridge. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2):143–172, 1992.

[11] NIST/SEMATECH. Measures of skewness and kurtosis. `https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm`, 2023. NIST/SEMATECH e-Handbook of Statistical Methods, accessed March 2026.

[12] Simon Gilchrist and Egbert Zakrajšek. Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720, 2012.

[13] SqueezeMetrics. The dark index (dix). `https://squeezemetrics.com/monitor/docs`, 2024. Accessed: 2026-04-17.

[14] Francis X. Diebold and Roberto S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995.

[15] Alexios Ghalanos. *rmgarch: Multivariate GARCH Models*, 2023. R package version 1.3-9.

[16] Fischer Black. Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, pages 177–181, 1976.

[17] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[18] Eugene F. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

[19] Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer, 2006.

A Additional Tables and Figures

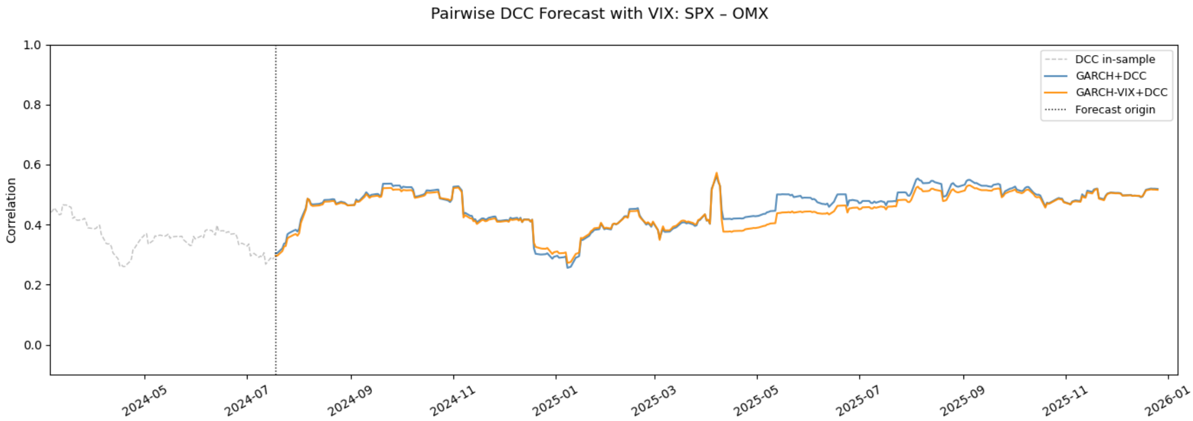


Figure A.1: DCC forecast with VIX: SPX - OMX

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	**2.17062**	**0.78123**	-2800.97	5605.95	5618.16
GARCH-VIX+DCC	2.45909	0.78153	**-2792.09**	**5588.17**	**5600.39**

Table A.1: SPX-OMX forecast errors and model fit, VIX variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-OMX	1.1098	0.2671	4.1672	0.0000***

Table A.2: DM-test GARCH+DCC vs GARCH-VIX+DCC, SPX-OMX.

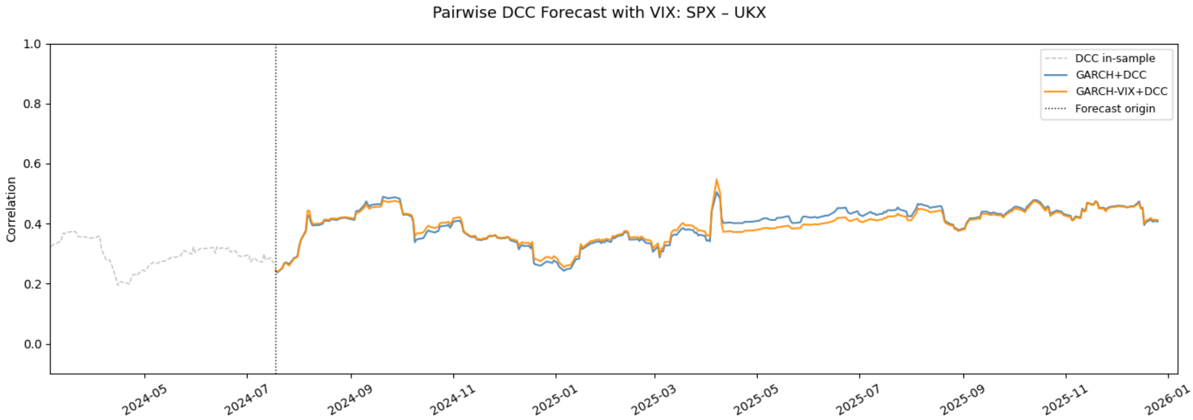


Figure A.2: DCC forecast with VIX: SPX - UKX

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	**1.99573**	**0.69192**	-2779.66	5563.31	5575.53
GARCH-VIX+DCC	3.13556	0.71541	**-2777.44**	**5558.89**	**5571.10**

Table A.3: SPX-UKX forecast errors and model fit, VIX variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-UKX	2.0215	0.0432**	2.1884	0.0286**

Table A.4: DM-test GARCH+DCC vs GARCH-VIX+DCC, SPX-UKX.

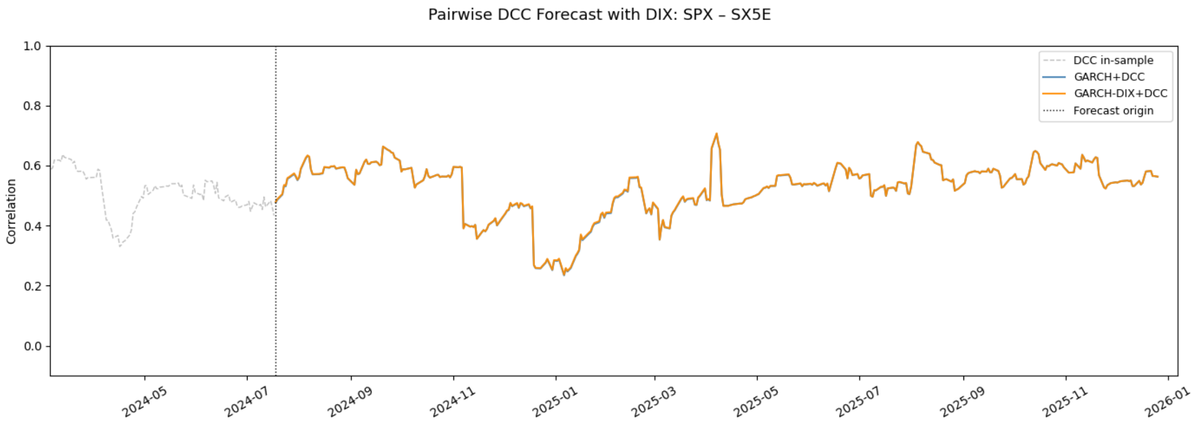


Figure A.3: DCC forecast with DIX: SPX - SX5E

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	2.48274	0.81867	-2612.81	5229.61	5241.83
GARCH-DIX+DCC	**2.47547**	**0.81814**	**-2609.63**	**5223.26**	**5235.48**

Table A.5: SPX-SX5E forecast errors and model fit, DIX variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-SX5E	0.1956	0.8449	-5.4744	0.0000***

Table A.6: DM-test GARCH+DCC vs GARCH-DIX+DCC, SPX-SX5E.

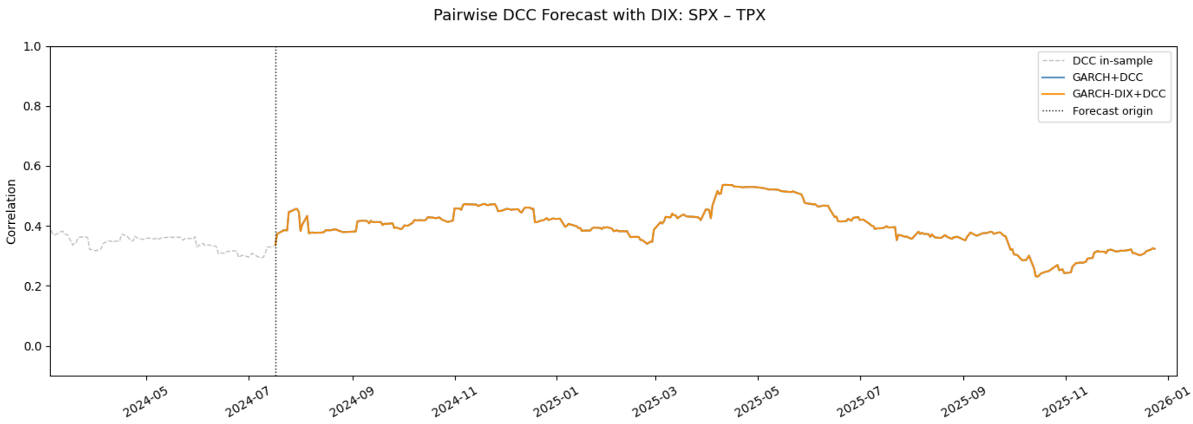


Figure A.4: DCC forecast with DIX: SPX - TPX

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	**2.29595**	0.76922	**-2889.89**	**5783.79**	**5796.00**
GARCH-DIX+DCC	2.29622	**0.76907**	-2889.97	5783.95	5796.17

Table A.7: SPX-TPX forecast errors and model fit, DIX variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-TPX	-1.9868	0.0469**	0.3483	0.7276

Table A.8: DM-test GARCH+DCC vs GARCH-DIX+DCC, SPX-TPX.

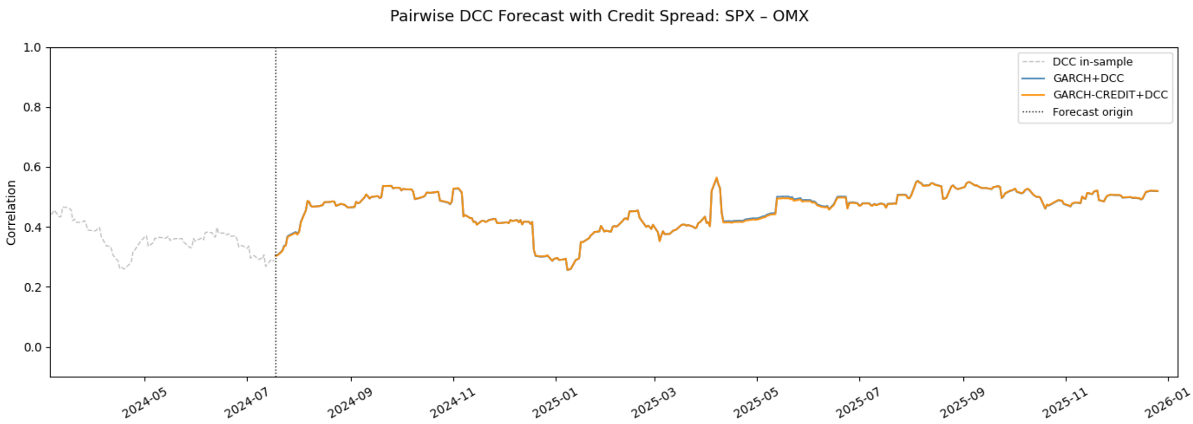


Figure A.5: DCC forecast with Credit Spread: SPX - OMX

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	2.17062	0.78123	-2800.97	5605.95	5618.16
GARCH-CREDIT+DCC	**2.14600**	**0.77422**	**-2788.06**	**5580.12**	**5592.34**

Table A.9: SPX-OMX forecast errors and model fit, Credit Spread variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-OMX	-0.4388	0.6608	2.1404	0.0323**

Table A.10: DM-test GARCH+DCC vs GARCH-CREDIT+DCC, SPX-OMX.

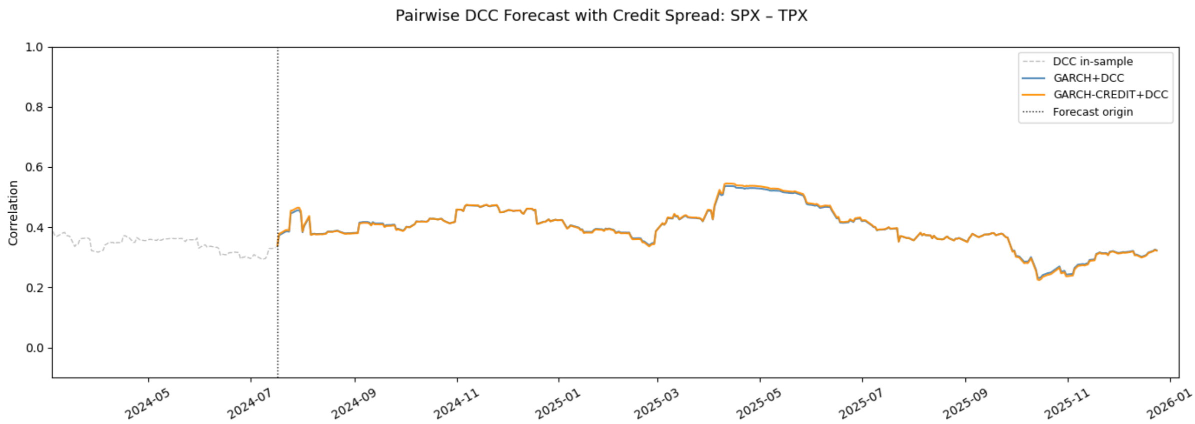


Figure A.6: DCC forecast with Credit Spread: SPX - TPX

Model	Avg MSE	Avg MAE	LogL	AIC	BIC
GARCH+DCC	**2.29595**	**0.76922**	-2889.89	5783.79	5796.00
GARCH-CREDIT+DCC	2.38227	0.77078	**-2882.26**	**5768.52**	**5780.73**

Table A.11: SPX-TPX forecast errors and model fit, Credit Spread variant

Pair	DM (MSE)	p-val	DM (MAE)	p-val
SPX-TPX	-0.0558	0.9555	-1.2503	0.2112

Table A.12: DM-test GARCH+DCC vs GARCH-CREDIT+DCC, SPX-TPX.

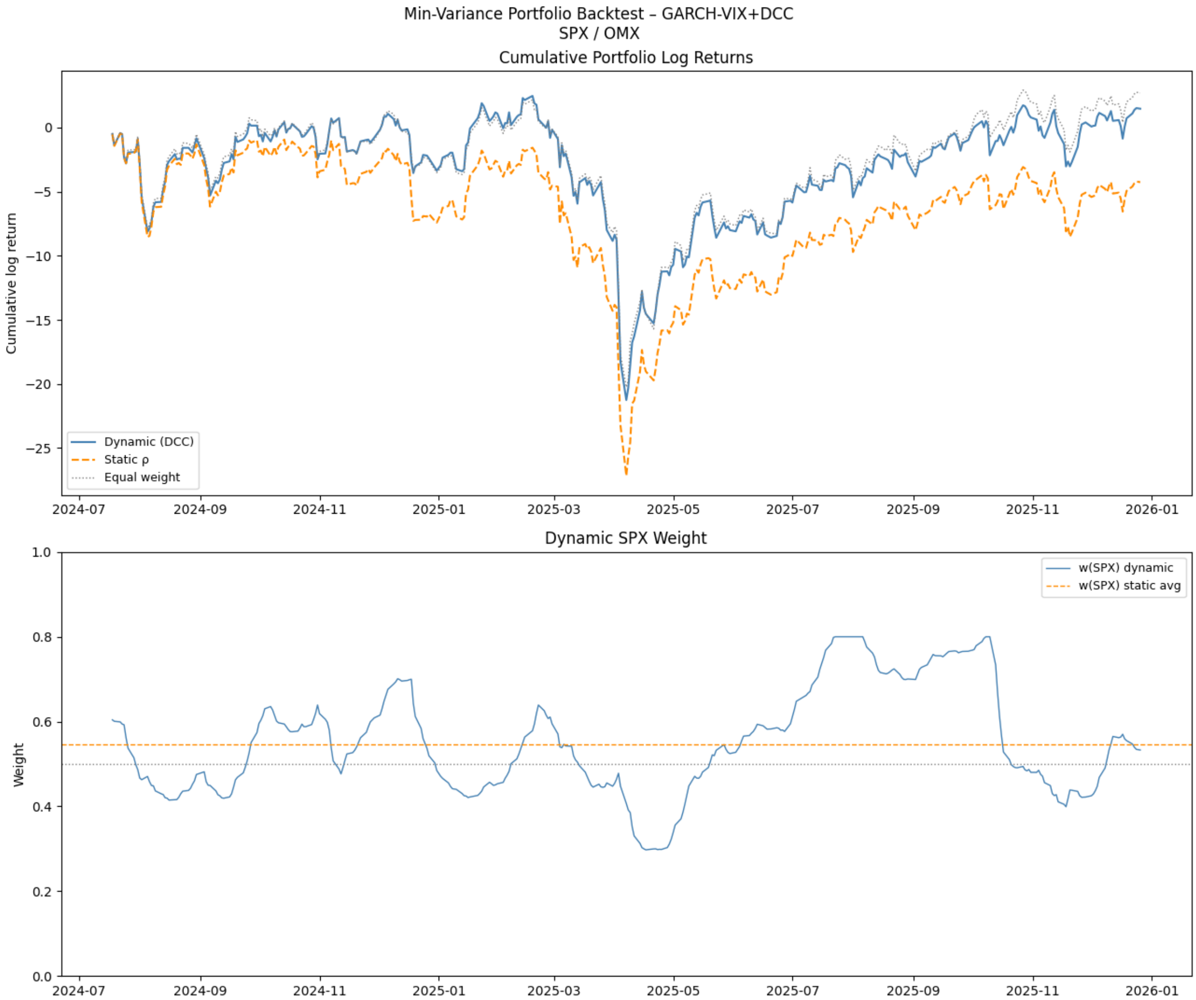


Figure A.7: Dynamic portfolio weights with VIX: SPX-OMX

Metric	Base	VIX	Δ
Dynamic variance	0.7263	0.7257	-0.0006
Variance reduction (%)	8.07	8.19	+0.12
Sharpe ratio (dynamic)	0.076	0.076	-0.001
Sharpe ratio (static)	-0.222	-0.209	+0.013
Sharpe ratio (equal weight)	0.139		

Table A.13: Portfolio comparison for SPX-OMX with VIX

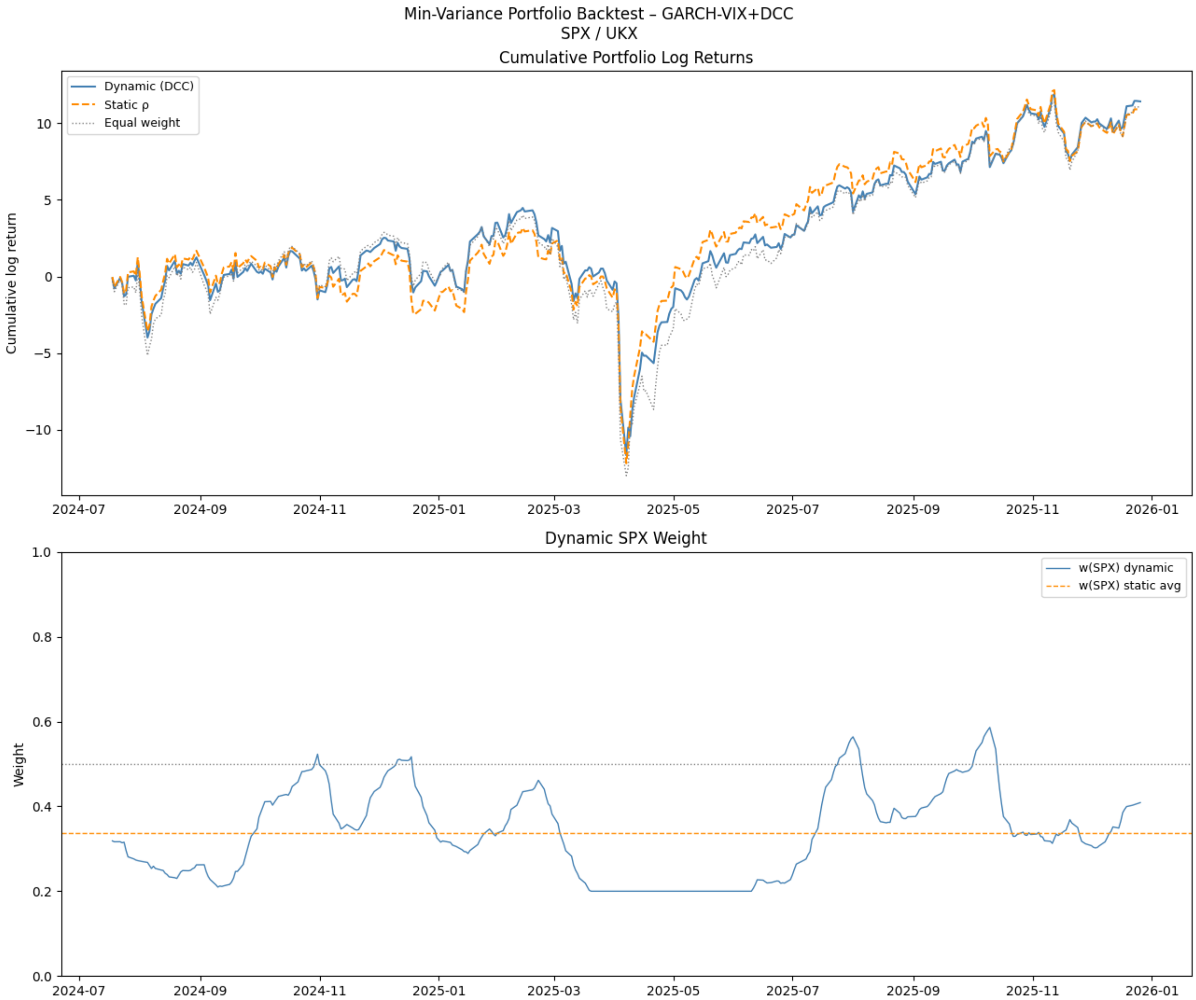


Figure A.8: Dynamic portfolio weights with VIX: SPX-UKX

Metric	Base	VIX	Δ
Dynamic variance	0.4602	0.4581	-0.0021
Variance reduction (%)	2.22	2.09	-0.13
Sharpe ratio (dynamic)	0.725	0.737	+0.012
Sharpe ratio (static)	0.666	0.694	+0.028
Sharpe ratio (equal weight)	0.665		

Table A.14: Portfolio comparison for SPX-UKX with VIX

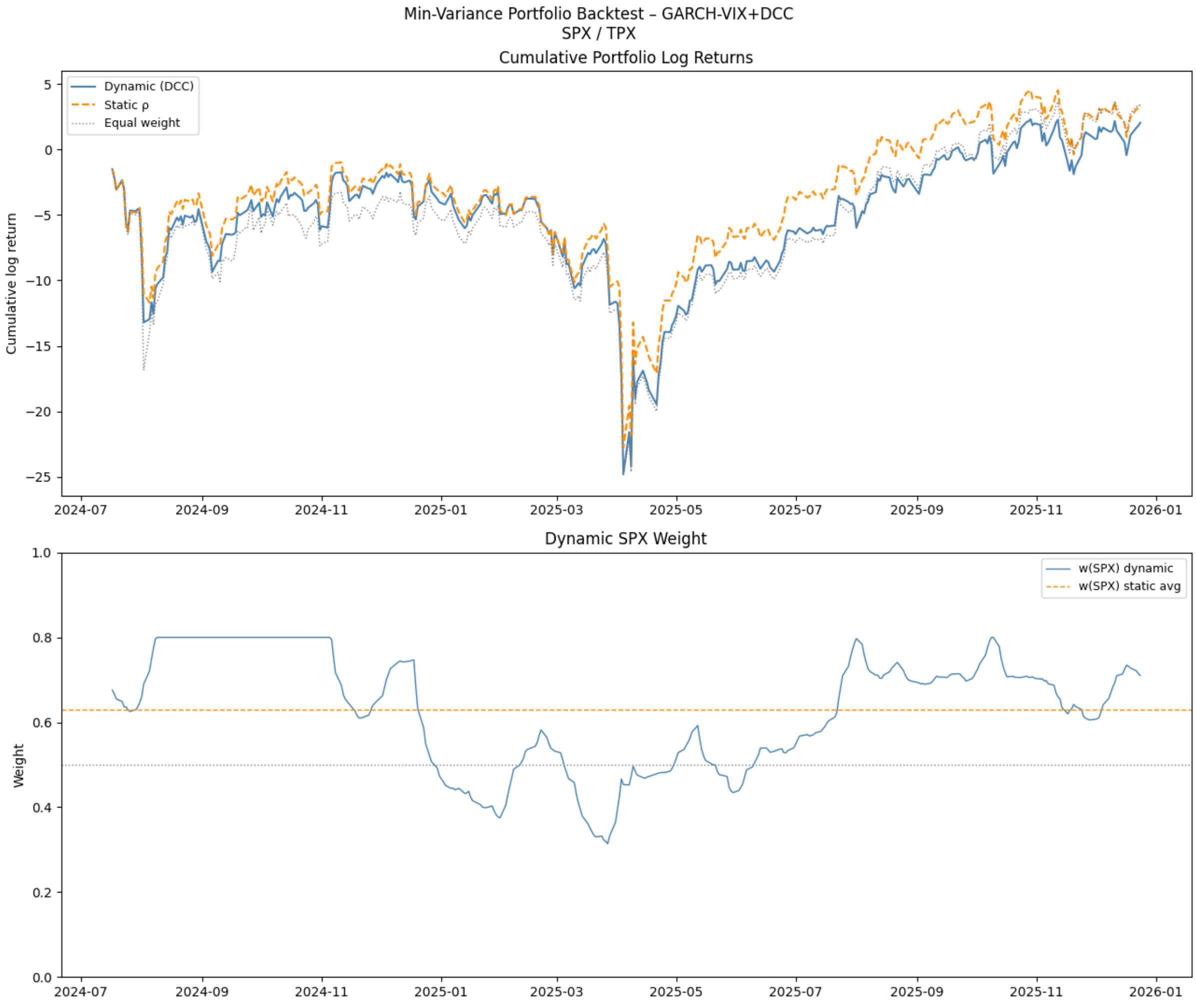
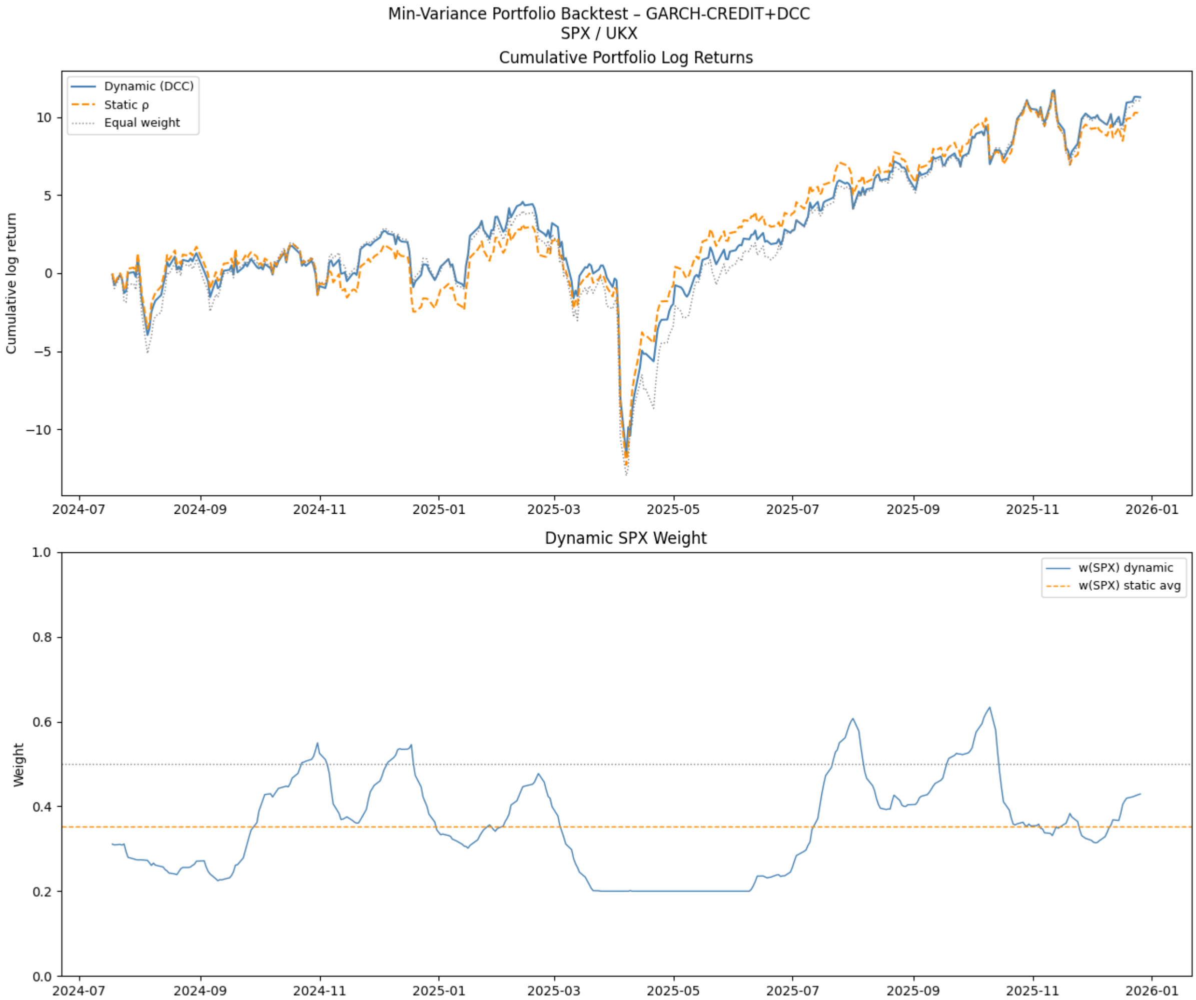


Figure A.9: Dynamic portfolio weights with VIX: SPX-TPX

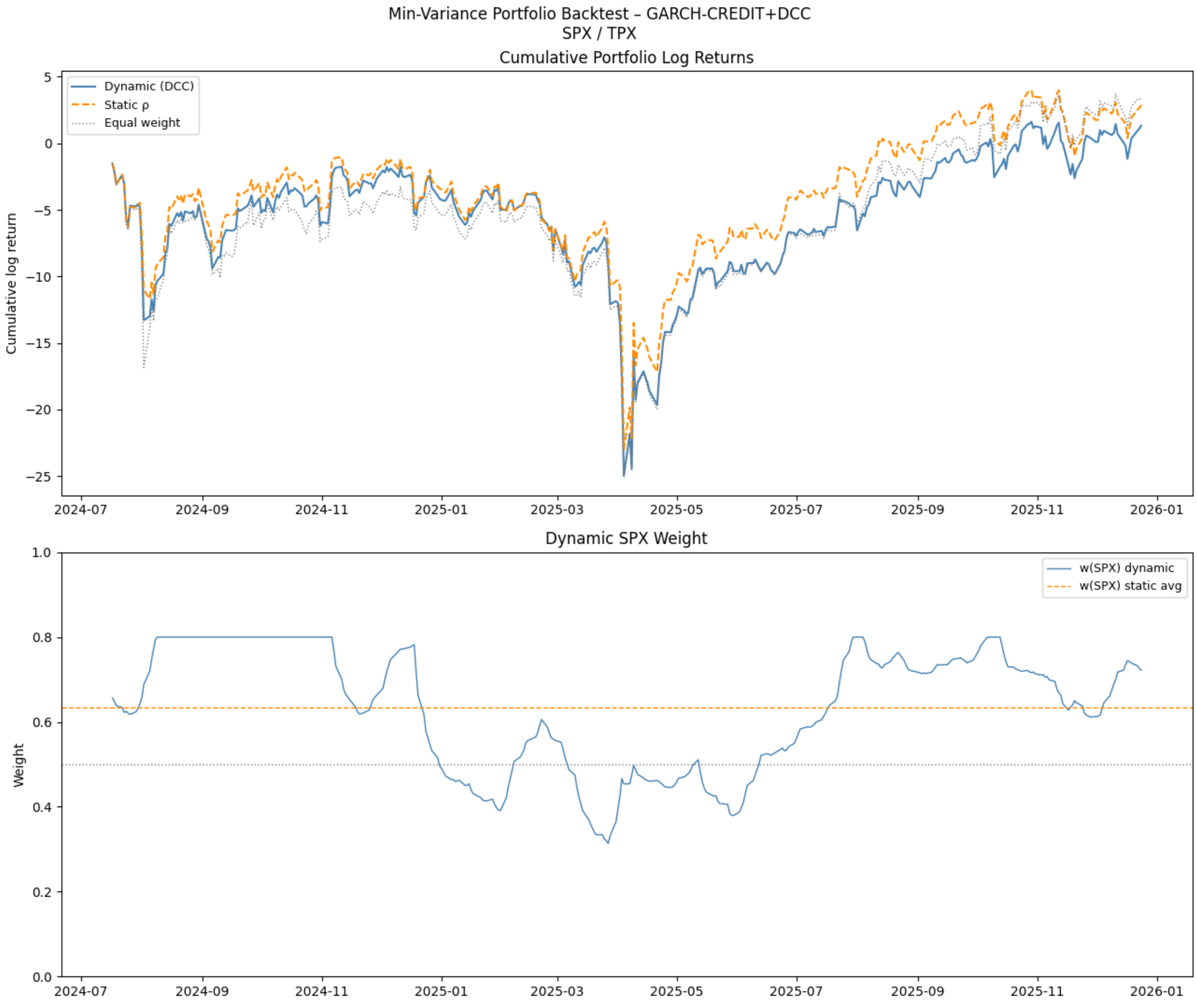
Metric	Base	VIX	Δ
Dynamic variance	1.1145	1.1182	+0.0037
Variance reduction (%)	-0.30	-0.53	-0.23
Sharpe ratio (dynamic)	0.074	0.084	+0.010
Sharpe ratio (static)	0.132	0.142	+0.010
Sharpe ratio (equal weight)	0.134		

Table A.15: Portfolio comparison for SPX-TPX with VIX



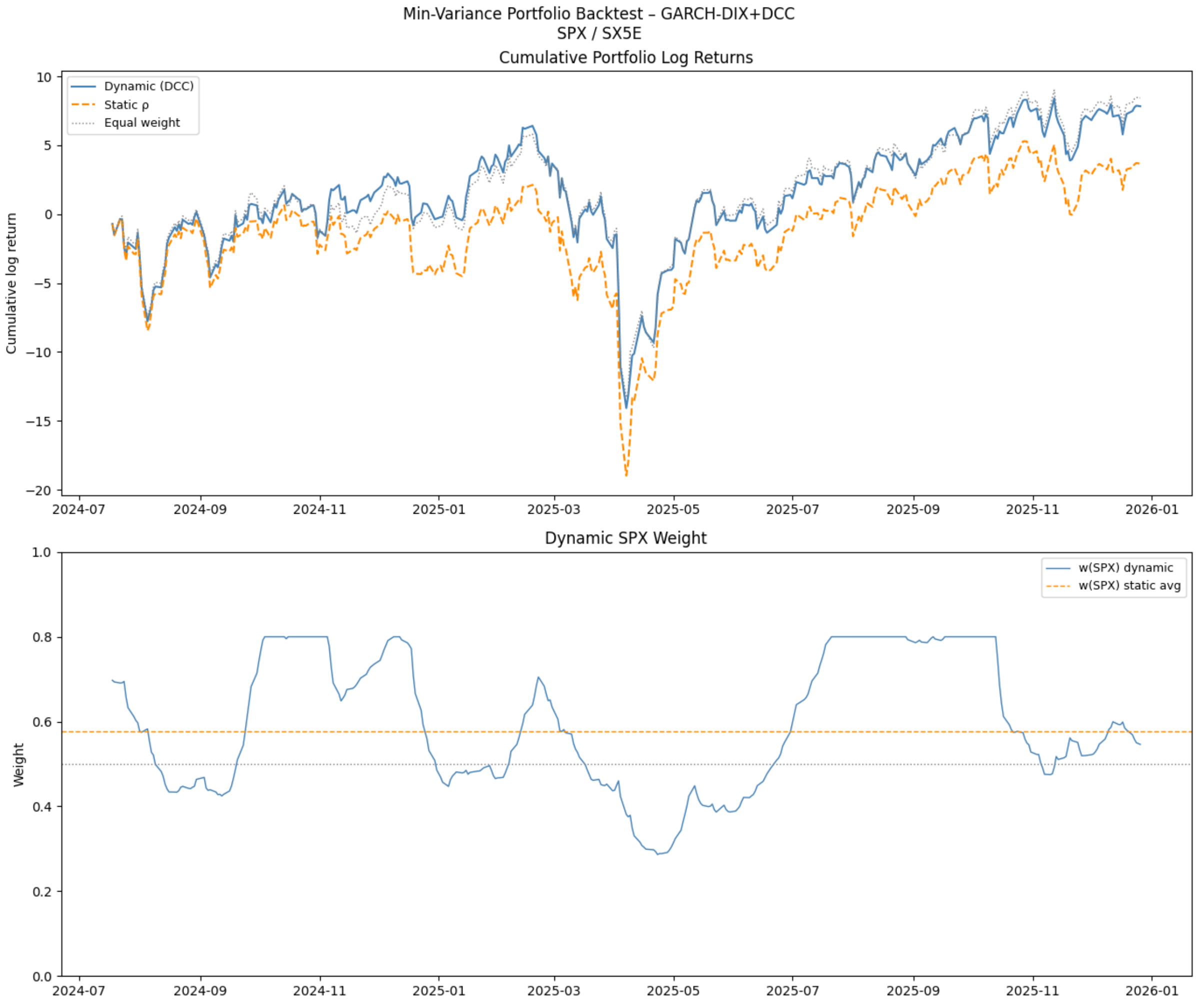
Metric	Base	Credit	Δ
Dynamic variance	0.4602	0.4619	+0.0017
Variance reduction (%)	2.22	2.09	-0.13
Sharpe ratio (dynamic)	0.725	0.725	+0.000
Sharpe ratio (static)	0.666	0.652	-0.014
Sharpe ratio (equal weight)	0.665		

Table A.16: Portfolio comparison for SPX-UKX with Credit Spread



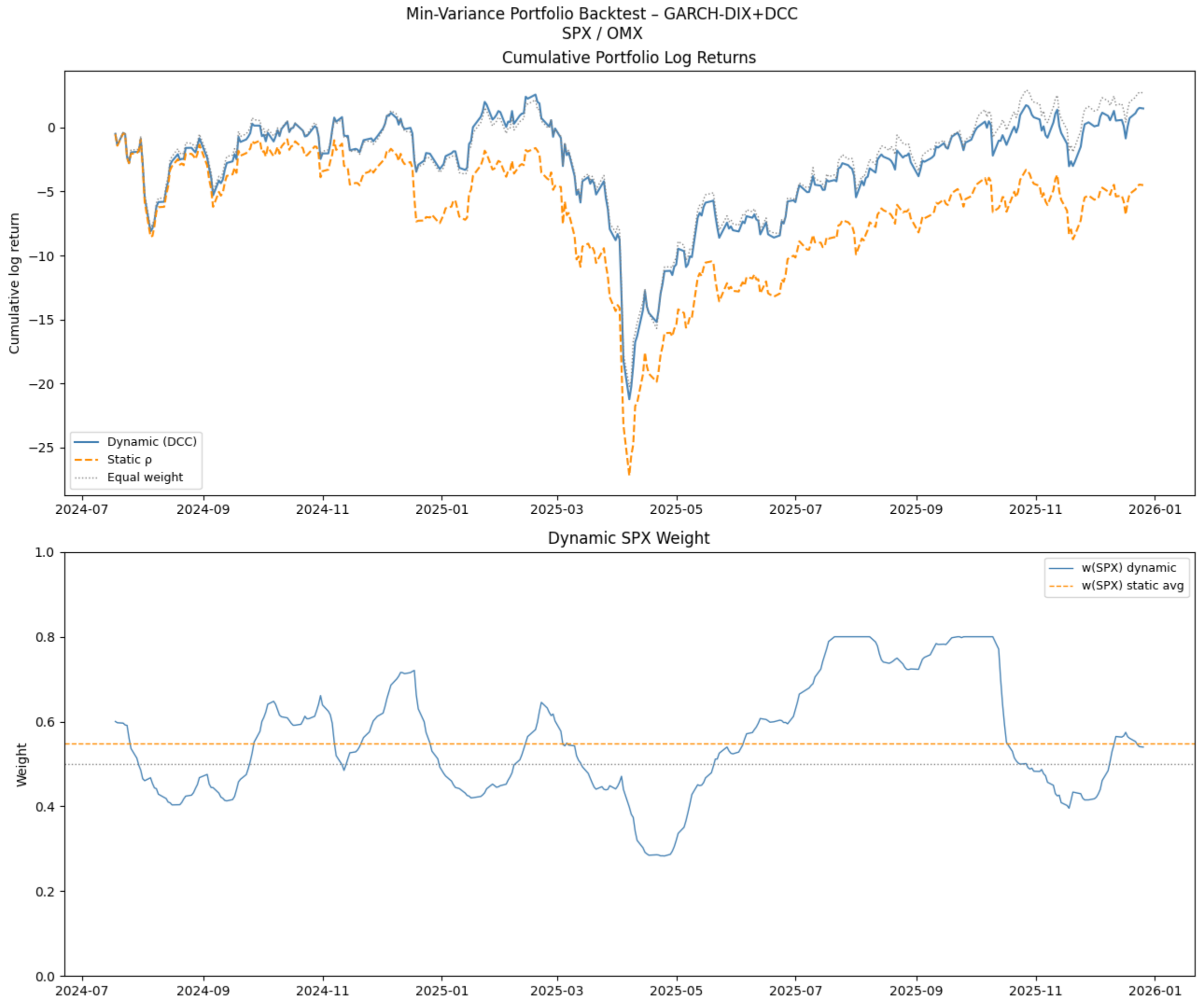
Metric	Base	Credit	Δ
Dynamic variance	1.1145	1.1169	+0.0024
Variance reduction (%)	-0.30	-0.90	-0.60
Sharpe ratio (dynamic)	0.074	0.055	-0.020
Sharpe ratio (static)	0.132	0.120	-0.012
Sharpe ratio (equal weight)	0.134		

Table A.17: Portfolio comparison for SPX-TPX with Credit Spread



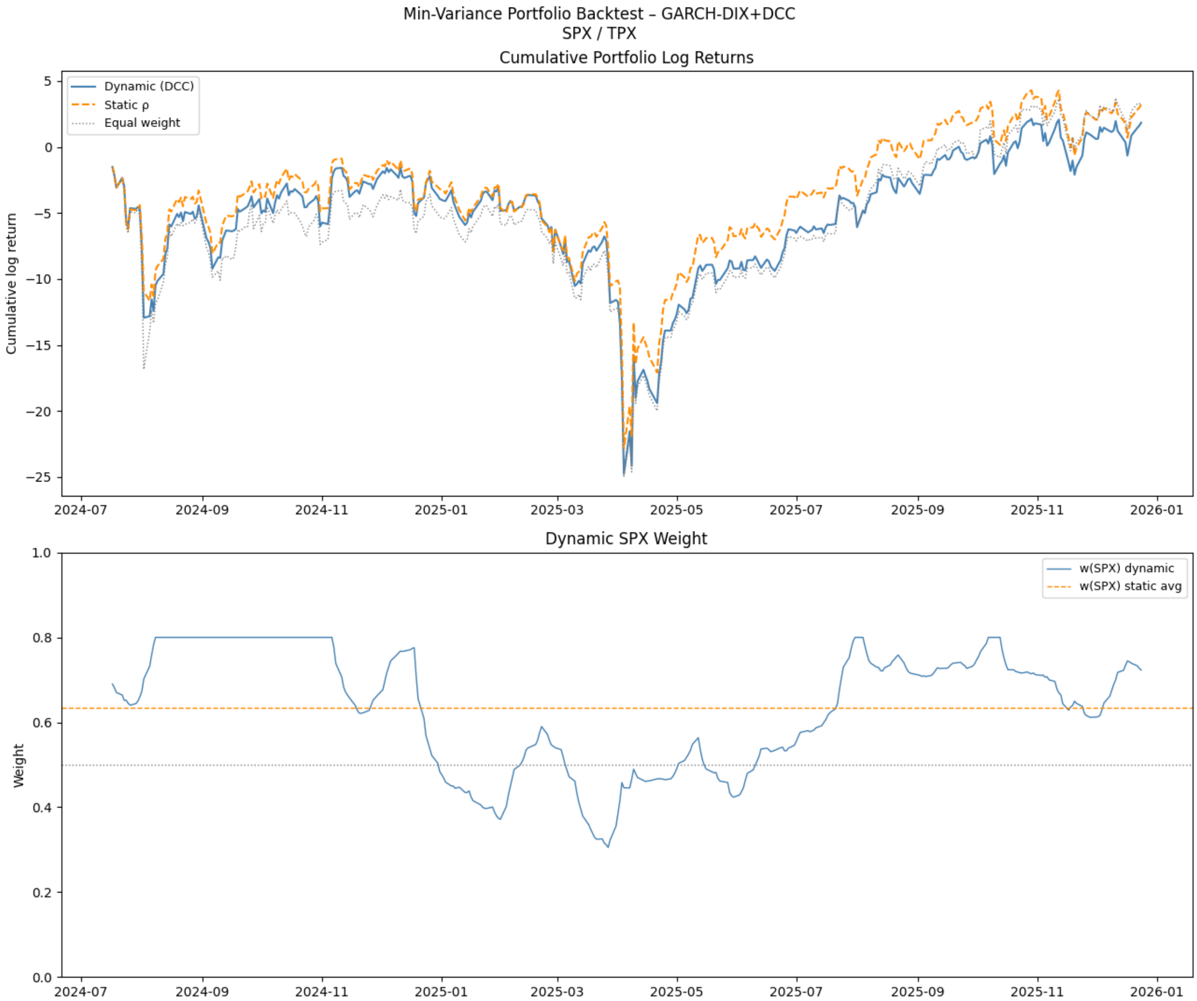
Metric	Base	DIX	Δ
Dynamic variance	0.7340	0.7339	-0.0001
Variance reduction (%)	7.69	7.71	+0.01
Sharpe ratio (dynamic)	0.399	0.400	+0.001
Sharpe ratio (static)	0.179	0.180	+0.001
Sharpe ratio (equal weight)	0.431		

Table A.18: Portfolio comparison for SPX-SX5E with DIX



Metric	Base	DIX	Δ
Dynamic variance	0.7263	0.7262	-0.0001
Variance reduction (%)	8.07	8.07	+0.00
Sharpe ratio (dynamic)	0.076	0.076	+0.000
Sharpe ratio (static)	-0.222	-0.222	+0.000
Sharpe ratio (equal weight)	0.139		

Table A.19: Portfolio comparison for SPX-OMX with DIX



Metric	Base	DIX	Δ
Dynamic variance	1.1145	1.1133	-0.0011
Variance reduction (%)	-0.30	-0.19	+0.11
Sharpe ratio (dynamic)	0.074	0.076	+0.002
Sharpe ratio (static)	0.132	0.133	+0.001
Sharpe ratio (equal weight)	0.134		

Table A.20: Portfolio performance comparison for SPX-TPX with DIX