Multivariate Volatility Measures and Models with Applications

Zhenglyu Huang

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Statistics
University of Sydney

June 2025

Abstract

This thesis explores methodologies for modelling and estimating correlation and covariance dynamics, presenting advancements in statistical approaches and their applications across multiple domains. We provide a comprehensive literature review of existing methodologies for modelling covariance matrices, focusing on their advantages, limitations, and practical implications, which highlights the need for efficient estimators and dynamic modelling techniques to address challenges such as heteroskedasticity, non-positive definiteness, and dynamic correlation structures. This review establishes the foundation for the proposed approaches developed in subsequent chapters.

The development begins with an extension of the methodology to measure covariance matrices in multi-dimensional settings, introducing efficient range-based correlation estimators. The proposed estimators leverage simple and weighted averages of range measures, with performance evaluated using simulation studies. Results demonstrate that the simple average mode estimator is optimal for moderate correlations, while the MSE-weighted mode estimator is superior for higher correlations. The impact of heteroskedasticity and non-positive definite (NPD) transformations are also addressed, showing that the NPD-transformed estimators improve performance for higher dimensions. Statistical tests confirm the superiority of range-based covariance matrices, achieving higher power and lower type I errors compared to sample covariance matrices. The methodologies proposed enable time series analysis of covariance matrices, facilitating dynamic correlation modelling and portfolio optimisation while mitigating random noise from data.

With our proposed range-based correlation matrix measures, we extend the two-stage multivariate Conditional Autoregressive Range Model (MCARR)-return models to directly model covariance matrix series using the Wishart distribution. Through simulation studies, we compare two approaches: modelling the covariance matrices and modelling the variances and correlation matrices. Correlation matrix modelling demonstrates better performance, guided by specific priors and stationary conditions.

We apply the proposed methodologies to real-world financial datasets. The first application focuses on U.S. stock market indices. Through comparative analyses, the range-based correlation and covariance estimators demonstrate superior modelling efficiency and predictive accuracy, particularly for the one-step-ahead Value-at-Risk forecasts. The second application identifies the impact of the COVID-19 pandemic on the cryptocurrency and commodity markets. Dividing the analysis into three periods of the pandemic development, we reveal significant coefficient variations across periods, highlighting dynamic market behaviour during the COVID-19 era. Additionally, we examine the integration of external features into the model, concluding that incorporating such features alongside persistence terms may lead to overfitting. These applications validate the robustness and adaptability of the proposed methodologies in capturing complex financial dynamics.

The findings of this thesis contribute to the development of more robust and adaptable tools for dynamic covariance and correlation modelling, offering practical applications in financial risk management and beyond. Future research directions include investigating the impact of drift and

non-normality on estimator performance, refining existing methodologies by incorporating additional information to address current limitations, and further analysing the dynamics of correlations.

Acknowlegement

I would like to express my heartfelt gratitude to everyone who has supported and encouraged me throughout this journey.

Firstly, I am deeply thankful for Australia's welcoming and supportive society. The values of democracy and liberty that underpin this society are vital for fostering academic environments where ideas can flourish freely. These principles have provided me with the opportunity to pursue my goals with confidence and creativity.

I would also like to extend my sincere thanks to my supervisors, Jennifer Chan and Gareth Peters. Their patience, guidance, and continuous support have been instrumental in shaping my work. Their encouragement and belief in my potential have inspired me to strive for excellence. I am especially grateful to my parents for their unconditional love, support, and belief in me. Their care and encouragement have been a cornerstone of my journey, inspiring me to persevere and grow.

Lastly, I would like to thank my partner, Xiang Li. Her unwavering encouragement, understanding, and belief in me have been a source of strength and motivation throughout this endeavour.

This achievement would not have been possible without the support of these individuals and the values of democracy and liberty that create an environment where learning and innovation can thrive.

Notations

- $\mathbf{\Sigma}_t$: The covariance matrix at time t
- $\sigma_{ij,t}$: The (i,j) entry of $\mathbf{\Sigma}_t$

We have the following simplified notation of the univariate case.

- σ_t : $\sqrt{\sigma_{11,t}}$
- N : Numbers of days of interest
- τ : τ-th intraday time
- n, n_t : Number of intraday time
- $P_{i,t,\tau}$: the price of asset i on day t at time τ
- τ : intraday time
- $O_{i,t}$: $P_{i,t,\min \tau}$
- $H_{i,t}$: $\max P_{i,t,\tau}$
- $L_{i,t}$: $\min P_{i,t,\tau}$
- $C_{i,t}$: $P_{i,t,\max \tau}$
- $O_{i,t,\tau}$: $P_{i,t,\tau}$
- $H_{i,t,\tau}$: $\max\limits_{\tau \leq \tau_2 < \tau+1} P_{i,t,\tau_2}$
- $L_{i,t,\tau}$: $\min\limits_{\tau \leq \tau_2 < \tau+1} P_{i,t,\tau_2}$
- $C_{i,t,\tau}$: $P_{i,t,\tau+1}$
- $o_{i,t,\tau}$: $\log O_{i,t}/C_{i,t-1}$, the opening jump.
- $h_{i,t,\tau}$: $\log H_{i,t}/O_{i,t}$
- $l_{i,t,\tau}$: $\log L_{i,t}/O_{i,t}$
- $c_{i,t,\tau}$: $\log C_{i,t}/O_{i,t}$
- $\boldsymbol{o}_{i,t}$: $(o_{i,1}, ..., o_{i,t,\tau})^\top$
- $\boldsymbol{h}_{i,t}$: $(h_{i,1}, ..., h_{i,t,\tau})^\top$
- $\boldsymbol{l}_{i,t}$: $(l_{i,1}, ..., l_{i,t,\tau})^\top$
- $\boldsymbol{c}_{i,t}$: $(c_{i,1}, ..., c_{i,t,\tau})^\top$

We have the following simplified notation of univariate case.

- h : $h_{1,1}$
- l : $l_{1,1}$
- c : $c_{1,1}$
- o_τ : $o_{1,t,\tau}$
- h_τ : $h_{1,t,\tau}$
- l_τ : $l_{1,t,\tau}$
- c_τ : $c_{1,1,\tau}$

- $\boldsymbol{J_f}$: Jacobian matrix of function $\boldsymbol{f}$

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1. Motivation

Price movement in financial markets reflects the dynamic interplay of supply and demand, investor sentiment, and external factors such as macroeconomic conditions, geopolitical events, and market news. These movements, which occur over various time horizons, can range from gradual trends to sharp fluctuations, and they embody the market's collective response to new information and shifting expectations. Volatility, a key measure of the variability in price movement, plays a central role in understanding and managing financial markets. It serves as a barometer of risk, providing insights into the uncertainty surrounding an asset's future returns. High volatility often signals periods of heightened market stress or uncertainty, while low volatility may indicate more stable conditions. Volatility is crucial for pricing derivatives, assessing portfolio risk, and designing effective trading strategies. Moreover, it impacts investor behaviour and decision-making, influencing market liquidity and capital allocation. Higher volatility indicates greater price swings, which can mean higher potential returns but also higher risk, while lower volatility suggests more stable price movements. As such, understanding and accurately modelling volatility is fundamental to financial analysis, risk management, and the broader functioning of financial markets.

Other risk metrics, such as Value at Risk (VaR; Jorion, 1996) and Conditional Value at Risk (CVaR; Artzner et al., 1999) have become widely used statistical measures in finance and risk management to estimate the potential loss in value of a portfolio or investment over a specified period, given normal market conditions, at a certain confidence level. The $\alpha\%$ VaR represents the probability of the price, or the return falls below the VaR value is $\alpha\%$. U.S. regulators, particularly the Federal Reserve, the Office of the Comptroller of the Currency (OCC), and the Securities and Exchange Commission (SEC) mandate the use of VaR in risk management frameworks, especially for large financial institutions. Under the Basel III regulations, which are applied in the U.S. through the Dodd-Frank Wall Street Reform and Consumer Protection Act, institutions are required to maintain capital reserves based on their VaR calculations to absorb potential losses during periods of market stress. Additionally, U.S. regulations enforce periodic stress testing of VaR models to ensure their accuracy and robustness, helping regulators monitor systemic risks and prevent financial instability. If the VAR models substantially underestimate the potential losses during stress scenarios compared to actual or plausible worst-case outcomes, they may be deemed inadequate. While the regulators do not penalize the VaR models if risks are overestimated, additional opportunity costs would occur due to the extra amount of capital reserves. VaR and volatility are intrinsically linked concepts in risk management and financial analysis. VaR quantifies the potential loss of a portfolio or asset over a specified time horizon and confidence level, providing a measure of downside risk. Volatility, as a measure of the dispersion of returns, plays a critical role in determining the magnitude of VaR. Higher volatility typically leads to greater potential losses, resulting in a larger VaR. Conversely, lower volatility suggests more stable returns, leading to a smaller VaR. The relationship between VaR and volatility highlights the importance of accurate volatility estimation, as misjudging volatility can result in underestimating or overestimating the

risk exposure of a portfolio. This connection underscores the value of robust volatility models in financial decision-making, allowing practitioners to better predict and manage risk under various market conditions as well as meet regulator requirements with a minimal cost. Furthermore, in option pricing, under the risk-neutral probability measure, the stochastic process governing asset prices becomes independent of the drift. This means that option prices are solely driven by volatility and the risk-free rate, as the expected drift is neutralized. Consequently, volatility plays a pivotal role in determining the fair value of options, with higher volatility leading to greater option premiums. In the risk-neutral framework, traders and analysts can focus on estimating volatility rather than drift, which simplifies the modelling process and provides a more robust method for pricing derivatives. Thus, understanding and accurately estimating volatility is essential for both portfolio management and option pricing under various market conditions and is thus a critical process in finance.

Volatility can be measured from various sources, for instance, historical volatility and implied volatility. Historical volatility reflects past price movements and is calculated based on historical data, whereas implied volatility represents market expectations of future fluctuations, derived from the pricing of financial derivatives. Both measures play critical roles in financial modelling, portfolio management, and derivative pricing, but each comes with inherent limitations. For instance, historical volatility is highly sensitive to the time period used in its calculation. Selecting a longer time horizon may violate the assumption of homoscedasticity, as market conditions and volatility dynamics are likely to change over extended periods. Conversely, using a shorter time period can lead to significant estimation errors due to the limited data available, thereby reducing the reliability of the measure. Implied volatility, on the other hand, depends on the availability of underlying derivatives and assumes that these instruments are traded in an efficient market. This assumption may not hold in less liquid markets, where pricing discrepancies and reduced trading activity can lead to inaccurate estimates. Furthermore, in markets with limited derivative products, the practical application of implied volatility may be constrained, diminishing its effectiveness as a predictive tool. These challenges underscore the need for robust methods that can address the limitations of both historical and implied volatility, ensuring accurate and reliable volatility estimation across various market conditions.

Volatility estimation can be approached using either non-parametric or parametric methodologies. The first approach, volatility measurement, aims to minimise the loss of estimation across the data set under the assumption of constant volatility. This method simplifies the dynamics of return, making it computationally efficient and easy to apply in many financial models. However, the assumption of constant volatility can be unrealistic, especially during periods of market stress, when price movements tend to be more volatile. A practical way to overcome heteroscedasticity is to estimate the local volatility using local data. Although this approach may not fully capture the dynamic nature of financial markets and lack of predictability, it remains popular for its simplicity and traceability in the estimation of local volatility.

Although the non-parametric approach is straightforward to implement, the assumption of homoskedasticity becomes impractical when the data spans a longer time horizon. The volatility modelling, on the other hand, aims to capture such time-varying nature of volatility, in particular, the volatility clustering, that is, periods of high volatility tend to be followed by high volatility, whereas periods of low volatility are also followed by low volatility.

These models, such as Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (Bollerslev, 1986) and stochastic volatility (SV) models (Taylor, 1986), account for volatility clustering and other patterns observed in financial time series. However, these models model volatility

indirectly through modelling the return process. Volatility models such as the Conditional Autoregressive Range Model (CARR) (Chou, 2005) using volatility measures as input provide a more direct and accurate reflection of market risk by allowing volatility to evolve, making them more suitable for applications where dynamic risk assessment is crucial, such as in derivative pricing and risk management.

Moreover, these traditional single-asset volatility models, while useful, often fall short when applied to portfolios with multiple securities, where the covariance between assets plays a significant role in determining overall risk. Higher-dimensional volatility models, which account for the dynamics of multiple assets, provide a more comprehensive view of portfolio risk, asset interactions, and market behaviour. In addition, correlations between assets under the multivariate models are not assumed to be constant but change during different market conditions, such as crises or bull markets. Generalizing volatility measures and models to higher dimensions can capture time-varying correlations to reflect the dynamic relationships between assets in response to changing market conditions. This provides a more dynamic and realistic assessment of portfolio risk. Furthermore, high-dimensional volatility modelling is also important in derivative pricing and hedging, where the value of an option or derivative is often dependent on the volatility and correlation of multiple underlying assets. By incorporating high-dimensional volatility estimates, derivative pricing models can better account for the complex interactions between assets, leading to more accurate pricing and more effective hedging strategies.

The remaining chapter is organized as follows. Section 1.2 provides a literature review of the model and methodology development as well as highlights the research gap this thesis aims to contribute. Sections 1.3 and 1.4 lay out the technical details of these related volatility and covariance measures and models in the literature review, illustrating the important advancements of our proposed covariance measures and models. Lastly, Section 1.5 summarizes the aims and contributions of this thesis.

1.2. Liturature review

Since volatility refers to the variance of log-returns, a straightforward approach to volatility estimation involves using the return time series. One of the most commonly used estimators in financial markets is historical volatility, which is computed as the sample variance of the return series. While historical volatility is widely applied for risk assessment and portfolio management, it has been criticized for its limited predictive power, particularly in dynamic market environments where volatility exhibits clustering and time-varying characteristics.

Under the assumption of log-normality of asset returns, research on return-based volatility estimation reached a theoretical limit, as the maximum-likelihood estimator achieved the Cramér–Rao bound, indicating that no unbiased estimator with lower variance could be derived within this framework. This limitation prompted researchers to explore alternative methods for estimating volatility that could improve efficiency and capture additional market dynamics.

A major breakthrough came with the work of Parkinson (1980), who derived the distribution of the difference between the maximum and minimum values of a random walk, leading to the development of the Parkinson (PK) measure. This innovation shifted the focus of volatility estimation away from simple return-based approaches, such as close to open (CO) volatility estimation, toward methods that leverage open, high, low and close (OHLC) data to obtain more accurate volatility estimates. By incorporating the extreme value of intraperiod price movement, range-based estimators are able to account for more information about market fluctuations, thereby improving efficiency compared to traditional return-based methods. Following Parkinson's work,

researchers have continued to refine and extend range-based volatility estimators. Garman and Klass (1980) identified an analytical minimum-variance estimator for volatility based on high-low-close prices, which further improved efficiency over the PK measure. Later, Meilijson (2011) extended this approach into the space of non-analytical estimators, broadening the scope of range-based volatility estimation. Another important line of research addresses the impact of drift in price movements, which can introduce biases in volatility estimation. To account for this issue, Rogers and Satchell (1991) proposed an unbiased estimator that remains independent of the drift component, ensuring more robust volatility measurement in the presence of price trends. Similarly, Yang and Zhang (2000) introduced a realized measure specifically designed to account for the effect of drift, improving the accuracy of volatility estimates in non-stationary market conditions.

The development of range-based volatility estimators has significantly improved the accuracy and reliability of volatility measurement in financial markets. By utilizing additional information beyond simple return calculations, these methodologies have enhanced risk assessment, trading strategies, and derivative pricing, making them valuable tools for financial practitioners and researchers. However, a key limitation of these estimators is their implicit assumption of homogeneous volatility, which is often unrealistic when applied over extended time horizons. In practice, financial markets exhibit heteroskedasticity, where volatility fluctuates over time due to structural changes, economic cycles, or external shocks. A practical approach to volatility estimation involves utilizing high-frequency daily data. By segmenting each trading day into multiple subperiods, one can compute a realized volatility (RV) measure, which provides a more refined assessment of market fluctuations. Merton (1980) demonstrated that the realized CO volatility measure converges under the assumption that asset prices follow a geometric Brownian motion. Subsequently, Andersen et al. (2010) extended this result to a broader class of semimartingales, thereby generalizing its applicability to a wider range of stochastic processes governing asset prices. However, the implementation of these methods relies on the availability of intraday trading data, which may not always be accessible, particularly for less liquid assets or markets with limited data reporting. The requirement for high-frequency observations poses challenges in data acquisition and computational efficiency, especially when applying these methods to large-scale datasets.

To address this issue, Engle (1982) introduced the Autoregressive Conditional Heteroskedasticity (ARCH) model, which treats volatility as a latent process that evolves dynamically based on past squared returns. This model allows for time-varying volatility, making it more applicable in financial time series analysis. Bollerslev (1986) extended the ARCH framework by developing the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, which incorporates a persistence term in the volatility process. The GARCH model captures long-memory properties in volatility, reflecting the tendency of large price fluctuations to cluster over time. Mikosch and Stărică (2000) derived the asymptotic distribution of the sample autocorrelation for the GARCH(1,1) process, demonstrating its convergence under the assumption that the process is stationary.

Further advancements in volatility modeling have focused on capturing asymmetries in volatility responses to positive and negative returns. Nelson (1991) proposed the Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) model, which explicitly accounts for the asymmetric impact of returns on volatility. Unlike the standard GARCH model, EGARCH allows for leverage effects, where negative shocks to returns tend to increase volatility more than positive shocks of the same magnitude. Similarly, the Glosten-Jagannathan-Runkle (GJR) GARCH model, introduced by Glosten et al. (1993), provides an alternative approach to modeling asymmetric volatility. This model incorporates a discrete volatility jump indicator, which is activated

by negative shocks, offering a more intuitive representation of how market downturns influence volatility. By explicitly distinguishing between positive and negative return shocks, the GJR-GARCH model enhances the ability to model risk asymmetries observed in financial markets.

Since the AR term in the GARCH model, which consists of the squared residuals, can be interpreted as the estimated volatility for a given day, researchers have explored methods to incorporate external volatility estimations into volatility modeling frameworks. By integrating the PK measure with these heteroskedastic models Chou (2005) proposed the Conditional Autoregressive Range (CARR) model and the Conditional Autoregressive Range model with exogenous variables (CARRX) model, which captures the asymmetric impact of return on volatility. Chou and Liu (2008) showed that the CARR model can significantly improve the predictability of volatility and outperform the return-based models in forecasting. The Heterogeneous Autoregressive model of Realized Volatility (HAR-RV), introduced by Corsi (2009), captures the long-memory property of volatility by incorporating multiple realized volatility measures computed over different time horizons.

Apart from these tool sets for understanding the volatility dynamics of the market, cross-dependency is another important feature for investors to consider in setting portfolio strategies for stock market indices. In portfolio management, we are required not only to estimate the volatility of each asset in a portfolio but also the correlation between these assets. This research motivates the extension of volatility models to the multivariate case in which we model the covariance and correlation matrices instead of the variances. The fitted mean and covariance matrices can be applied to the Mean-Variance (MV) portfolio optimisation, introduced by Markowitz (1952), which remains a cornerstone in modern finance for constructing portfolios in the investment universe. However, the success of MV optimisation critically depends on the accuracy of these input estimates particularly the mean returns and the variance-covariance matrix, highlighting the importance of multivariate volatility and return models in the finance and fund management industry.

In the multivariate setting, the estimation of the variance-covariance matrix is the sample covariance matrix, which is a natural extension of sample volatility and is widely utilized due to its computational simplicity and the inherent property of positive definiteness. The estimation of covariance matrices is a fundamental topic in financial econometrics, statistical modeling, and portfolio optimization, as it plays a critical role in quantifying the relationships between multiple assets or variables. Research in this area is typically divided into covariance estimation and correlation estimation, each addressing different aspects of dependence structure modeling. Campbell et al. (1998) proposed an S-estimator of multivariate mean vector and scale matrix as a generalization, which minimizes the determinant of the covariance matrix. Brandt and Jones (2006) proposed a way to generalize variance estimators to covariance measures for financial data. As for correlation estimation, the Pearson correlation coefficient, induced from sample covariance and variance, serves as a widely accepted correlation estimator, particularly under the assumption that the data follows a multivariate normal distribution. However, in cases where this assumption does not hold, alternative non-parametric correlation estimators are often employed. For instance, the Spearman rank correlation and Kendall's tau correlation extend correlation estimation to arbitrary data distributions, including non-normally distributed data and even ordered categorical data. These rank-based measures are particularly useful in scenarios where the relationship between variables is nonlinear or where the data contains significant outliers that could distort traditional covariance-based estimations.

Apart from non-normality, nonlinearity is another consideration. Hoeffding (1994) proposed the Hoeffding's D correlation measure that can pick up the nonlinear relationship. Zheng et al.

(2012) proposed a generalized measure of correlation (GMC) that deals with asymmetries in variances, nesting Pearson correlation as a special case. Recently, Zhang et al. (2024) proposed the quantile GMC to describe the nonlinear quantile relationship between vectors using a conditional quantile function. Moreover, when correlation is applied to non-independent observations or aggregated data, it may produce biased and specious results due to differing patterns between groups versus within groups. In light of this, Bakdash and Marusich (2017) proposed repeated measures correlation and introduced the R package `rmcorr` for implementation. Lastly, the distance correlation matrix uses the Euclidean distance between random vectors (Székely et al., 2007). Drouet Mari and Kotz (2001) provides a summary of correlation measures.

Unlike volatility estimation, which has seen extensive advancements incorporating intraperiod price movements, there has been relatively little research on covariance or correlation measures that leverage intraperiod information. Most existing methods rely on closing prices or low-frequency data, overlooking the richer information available from high-frequency intraday price fluctuations. Brandt and Diebold (2006) proposed an arbitrage-free range covariance estimation. However, a key limitation of this method is its reliance on an external proxy variable, which must be observable for implementation. This requirement restricts its practical applicability, particularly outside the foreign exchange market. As the range-based volatility measures such as the PK measure was proved to be efficient in volatility estimation, extensions of the idea of PK measure to the covarance and correlation matrix measures should be promising but is lacking in the literature.

Similar to estimation approaches, the extension of volatility modeling to the multivariate setting can be achieved through either covariance modeling or correlation modeling. These two approaches offer distinct ways to capture the dynamics of volatility and dependencies among multiple assets or financial instruments. One of the earliest multivariate volatility models is the VEC-GARCH model, proposed by Bollerslev et al. (1988), which utilizes a vectorization operator to transform the covariance matrix into a vector. This formulation allows for the inclusion of all possible linear dependencies between elements of the covariance matrix, making it a theoretically comprehensive approach to modeling multivariate volatility. However, a significant drawback of the VEC-GARCH model is that it does not inherently guarantee the positive definiteness of the estimated covariance matrix. As a result, practical implementations often impose strong constraints on the coefficient matrix, such as restricting it to a diagonal form. While these restrictions help maintain mathematical feasibility, they significantly limit the model's flexibility and its ability to accurately capture complex dependency structures in high-dimensional settings. To address some of these limitations, Engle and Kroner (1995) introduced the BEKK-GARCH model, which can be regarded as a special case of the VEC-GARCH model. The BEKK model expresses the covariance process as a linear combination of quadratic forms, ensuring the positive definiteness of the covariance matrix while maintaining a more structured formulation. However, despite its theoretical advantages, the BEKK-GARCH model is computationally intensive, as it involves a large number of parameters that need to be estimated, especially in higher dimensions.

In contrast to covariance modeling, an alternative approach is correlation modeling, which focuses on directly estimating the time-varying correlation structure while separately modeling volatility. The simplest example of correlation modeling is the Constant Conditional Correlation (CCC-GARCH) model, introduced by Bollerslev (1990), which assumes that correlations between assets remain constant over time. While this assumption simplifies computation and ensures positive definiteness, it is often unrealistic, as empirical evidence suggests that correlations fluctuate over time, particularly during periods of financial stress. To account for the dynamic nature of correlations, Engle (2002) proposed the Dynamic Conditional Correlation (DCC-GARCH) model,

which extends the CCC-GARCH framework by allowing the correlation matrix to evolve over time. The DCC-GARCH model has become one of the most widely used approaches in multivariate volatility modeling due to its balance between flexibility and computational feasibility. Unlike fully parameterized models such as VEC-GARCH or BEKK-GARCH, DCC-GARCH estimates only a small number of parameters governing the correlation dynamics, as the coefficients in the correlation matrix process are simply scalers, making it more scalable for higher-dimensional applications.

Alternatively, rather than relying on a parsimonious representation of the covariance matrix based on return vectors, one can adopt the covariance regression model proposed by Hoff and Niu (2012). This approach allows the covariance matrix to be expressed as a function of explanatory covariates, enabling the direct modelling of how these factors influence the variability and co-movement of assets. Such methods are particularly effective in applications where the underlying drivers of the high dimension data are likely to be common. For example, they are well-suited for constructing portfolios consisting of stocks listed on a specific exchange, where asset price movements are influenced by shared market conditions and economic factors.

Further advancing the multivariate volatility modeling framework, Chou et al. (2009) introduced the Multivariate Conditional Autoregressive Range (MCARR) model, which extends the CARR model to a multivariate setting. The MCARR model improves volatility estimation by leveraging various MGARCH formulations, that is the DCC formulation, to incorporate intraperiod price ranges, offering a more information-rich alternative to traditional return-based volatility estimators. By utilizing range-based measures, the MCARR model enhances the efficiency of volatility estimation, capturing market dynamics that might be overlooked in conventional models relying solely on closing prices. Building on this foundation, Tan et al. (2022) proposed an alternative approach using the parity of volatility within linear combinations of assets to model the covariance structure. However, despite these advancements, prior research has largely specified the error distribution as a set of independent univariate distributions, which limits the ability to fully capture the dependence structure and interactions within multivariate financial data. A key research gap exists in the integration of matrix distributions, such as the Wishart distribution (Wishart, 1928) and the Matrix-F distribution (Goodman, 1963), as error distributions within multivariate volatility modeling frameworks. Matrix distributions provide a more natural and mathematically consistent way to model covariance and correlation dynamics, as they inherently preserve the positive semi-definiteness of covariance matrices and allow for more flexible dependency structures. Incorporating such distributions into multivariate volatility models could enhance estimation accuracy and robustness, particularly in capturing joint volatility shocks and systemic risk in financial markets.

These multivariate models are important tools to understand different market characteristics such as the cryptocurrency market, which is already known to be more volatile than the traditional asset markets. Hence, the co-movement of cryptocurrency prices will add even more risk to investors. Oscillating autocorrelation was observed in the volatility of cryptocurrency returns Fang et al. (2021), but it is unsure if the range-based covariance between a pair of cryptocurrencies also displays this oscillating autocorrelation behaviour.

Another potential area of application of these multivariate covariance and return models will be the study of the COVID-19 pandemic, which brought global stress to many stock markets. Key findings from some studies have already revealed the dramatic increase in volatility in early 2020 (Nitithumbundit and Chan, 2022). During this period, assets like gold and cryptocurrencies acted as hedges against uncertainty. Research also revealed that market reactions to COVID-19 news were rapid and non-linear. These findings provide valuable insights to policymakers and

economists in formulating strategies to stabilize financial markets during future global crises. Applying multivariate predictive models such as the MCARR that incorporate pandemic-related variables will help forecast market movements and improve contingency planning for extreme events.

1.3. Review of uncertainty measures

1.3.1. Volatility measures. Statistically, the volatility measure is equivalent to the variance estimator of returns $r_t,\ t = 1, \ldots, N$. The sample variance, defined as

$$s^2 = \frac{1}{N-1}\sum_{t=1}^{N}(r_t - \bar{r})^2, \tag{1.1}$$

is a fundamental unbiased estimator to estimate the variance from N i.i.d samples. While it may not be the best estimator that fits our interests, technically speaking, minimizing the loss function, the sample variance is still a widely adopted, consistent estimator that makes no assumptions about the underlying distribution of r_t other than the finiteness of the first and the second moments. The estimation theory based on i.i.d samples has been well developed. In terms of the financial time series, our sample would neither be identically distributed nor independent in general. Even if we assume independence, different time gaps between each trade result in different distributions of the price. The geometric Brownian motion is a widely used stochastic process for modelling the evolution of financial asset prices over time. The stochastic differential equation of the geometric Brownian motion is given by:

$$dP_t = \mu P_t dt + \sigma P_t dW_t, \tag{1.2}$$

where P_t is the asset price , μ is a constant percentage drift, σ is a percentage volatility and W_t is a Wiener process. It combines deterministic trends with random fluctuations, making it a popular tool for capturing the unpredictable nature of markets. The geometric Brownian motion is characterized by two primary components: a drift term that represents the expected rate of return and a volatility term that accounts for the randomness in price movements. Using Itô's lemma, the solution of (1.2) is given by

$$\log\frac{P_t}{P_0} = (\mu - \sigma^2/2)t + \sigma W_t. \tag{1.3}$$

In our real application, since the trading is happening so frequently that the sequence of prices would be extremely large, we usually use the summary statistics between a fixed interval instead of individual trading. Then by (1.3), we have

$$\log\frac{P_t}{P_{t-1}} = (\mu - \sigma^2/2) + \sigma(W_t - W_{t-1}). \tag{1.4}$$

Hence the log-return series $\{\log\frac{P_t}{P_{t-1}}\}$ is a sequence of i.i.d random variables. We can use the result from the estimation theory to get an estimator of the volatility.

However such constructions only utilize the information by the end of the interval $(t-1, t)$. While using all trading information between $(t-1, t)$ is not feasible due to the large sample size, research has shown that a more efficient estimator can be conducted by incorporating high and low prices in our estimation. Let the opening, high, low and closing price of the interval $(t, t+1)$ be

$$\{O_t, H_t, L_t, C_t\} \text{ and } h_t = \log\frac{H_t}{O_t},\ l_t = \log\frac{L_t}{O_t},\ c_t = \log\frac{C_t}{O_t},\ o_t = \log\frac{O_t}{C_t} = -c_t \tag{1.5}$$

be the normalised high, low, close and open. Assuming no drift in log-return, that is $\mu - \sigma^2/2 = 0$ in equation (1.4), Parkinson (1980) propose a Parkinson measure (PK), defined as

$$\hat{\sigma}^2_{\mathtt{PK},t} = \frac{(h_t - l_t)^2}{4\log 2}, \tag{1.6}$$

by working on the asymptotic distribution of $h_t - l_t$. Garman and Klass (1980) extend the idea to utilize all O_t, H_t, L_t, C_t information. By searching for an analytical function that minimises the variance, a precise GK measure is given by

$$\hat{\sigma}^2_{\mathtt{GK},t} = 0.511(h_t - l_t)^2 - 0.019(c_t(h_t + l_t) - 2h_t l_t) - 0.383c_t^2. \tag{1.7}$$

A practical close form of (1.7) is given by

$$\hat{\sigma}^2_{\mathtt{GK2},t} = v_{\mathtt{GK2},t} = 0.5(h_t - l_t)^2 - (2\log 2 - 1_t)c_t^2. \tag{1.8}$$

Meilijson (2011) found a non-analytical estimator by a linear combination of 4 fundamental unbiased estimators

$$\hat{\sigma}^2_{\mathtt{M},t} = 0.273520\,\hat{\sigma}^2_{1,t} + 0.160358\,\hat{\sigma}^2_{2,t} + 0.365212\,\hat{\sigma}^2_{3,t} + 0.200910\,\hat{\sigma}^2_{4,t}, \tag{1.9}$$

$$\hat{\sigma}^2_{1,t} = 2\big[(h'_t - c'_t)^2 + l'^2_t\big], \tag{1.10}$$

$$\hat{\sigma}^2_{2,t} = c'^2_t, \tag{1.11}$$

$$\hat{\sigma}^2_{3,t} = 2(h'_t - c'_t - l'_t)c_t, \tag{1.12}$$

$$\hat{\sigma}^2_{4,t} = -\frac{(h'_t - c'_t)l'_t}{2\log 2 - 5/4}, \tag{1.13}$$

where

$$c'_t = |c_t|,\ \ h'_t = \begin{cases} h_t, & \log c_t > 0 \\ -l_t, & \log c_t < 0 \end{cases},\ \ l'_t = \begin{cases} l_t, & \log c_t > 0 \\ -h_t, & \log c_t < 0 \end{cases}. \tag{1.14}$$

If the drift is not 0, the above estimators would be biased. To accommodate the drift, Rogers and Satchell (1991) showed that the estimator given by

$$\hat{\sigma}^2_{\mathtt{RS},t} = v_{\mathtt{RS},t} = h_t(h_t - c_t) + l_t(l_t - c_t). \tag{1.15}$$

is an unbiased estimator in the presence of drift. While the RS measure is independent of the drift, it would have a larger variance than other estimators. Meanwhile, by reformating the middle term of precise GK

$$c_t(h_t + l_t) - 2h_t l_t = -h_t^2 + c_t h_t + l_t c_t - l_t^2 + h_t^2 - 2h_t l_t + l_t^2 = -h_t(h_t - c_t) - l_t(l_t - c_t) + (h_t - l_t)^2, \tag{1.16}$$

we can find that indeed GK-measure is a linear combination of PK, RS and classical return-based volatility measure.

The above estimators are constructed based on time interval $(t, t+1)$. If we have a subsequence of $\{O_{t\tau}, H_{t\tau}, L_{t\tau}, C_{t\tau}\}$, $\tau = 1, \ldots, n$ within $(t, t+1)$, we can have a realized version of each estimator by taking the average over periods, i.e., the realized PK is given by

$$v_i^{\mathtt{PK}} = \sum_{\tau=1}^{n} \frac{(h_{i,t,\tau} - l_{i,t,\tau})^2}{4\ln 2}. \tag{1.17}$$

Furthermore, multiperiod data allows us to get an estimation of the drift of return, which can be used to calibrate the bias raised from the drift. Yang and Zhang (2000) proved that the minimal

variance estimator of n periods should take form of

$$\hat{\sigma}_{\texttt{YZ},t} = \frac{1}{n-1}\sum_{\tau=1}^{n}(o_{t\tau} - \bar{o}_t)^2 + k\frac{1}{n-1}\sum_{\tau=1}^{n}(c_{t\tau} - \bar{c}_t)^2 + (k-1)\frac{1}{n}\sum_{\tau=1}^{n}\hat{\sigma}^2_{\texttt{RS},t\tau} \tag{1.18}$$

where $\bar{o}_t = \frac{1}{n}\sum_{\tau=1}^{n} o_{t\tau},\ \bar{c}_t = \frac{1}{n}\sum_{\tau=1}^{n} c_{t\tau},\ \hat{\sigma}^2_{\texttt{RS},t\tau} = h_{t\tau}(h_{t\tau} - c_{t\tau}) + l_{t\tau}(l_{t\tau} - c_{t\tau})$. The parameter k is given by

$$k = \frac{\alpha - 1}{\alpha + \frac{n+1}{n-1}}, \tag{1.19}$$

where α is a constant related to the drift and in practice, we can set it to 1.34.

1.3.2. Covariance measures and matrices. The covariance estimation is much more complicated compared to variance estimation. We always need multiple periods of data to estimate covariance. There are two approaches in general: directly calculating covariance or estimating correlation.

To estimate the covariance directly for $c_{i,t}$ defined in (1.5) with subscript i to denote asset i, the sample covariance defined as

$$\text{cov}(c_i, c_j) = v_{ij} = \frac{1}{N-1}\sum_{t=1}^{N}(c_{i,t} - \bar{c}_i)(c_{j,t} - \bar{c}_j) \tag{1.20}$$

is commonly used and is guaranteed to be semi-positive definite. The sample covariance estimation can be easily extended to the sample covariance matrix.

As a generalization, Campbell et al. (1998) proposed an S-estimator of multivariate mean vector and scale matrix, which minimizes the determinant of the covariance matrix, subject to a constraint on the magnitudes of the corresponding Mahalanobis distances:

$$N^{-1}\sum_{t=1}^{N}\rho(z_t) = N^{-1}\sum_{t=1}^{N}\rho\big\{\big[(\boldsymbol{c}_t - \hat{\boldsymbol{\mu}})^\top\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{c}_t - \hat{\boldsymbol{\mu}})\big]^{1/2}\big\} = b_0 \tag{1.21}$$

where $\rho(\cdot)$ is symmetric with continuous derivative $\psi(\cdot)$, and $\rho(0) = 0$. The choice of $\rho(\cdot)$ can be the Tukey biweight function, and b_0 is the expected value of $\rho(z_t)$ assuming a multivariate Gaussian distribution. The solution to (1.21) is

$$\widehat{\boldsymbol{\Sigma}} = \frac{\sum_{t=1}^{N}\text{w}(z_t)(\boldsymbol{c}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{c}_t - \hat{\boldsymbol{\mu}})^\top}{\sum_{t=1}^{N} z_t\psi(z_t)},\ \text{where } \hat{\boldsymbol{\mu}} = \frac{\sum_{t=1}^{N}\text{w}(z_t)\boldsymbol{c}_t}{\sum_{t=1}^{N}\text{w}(z_t)},\ \text{and } \text{w}(z_t) = \frac{\psi(z_t)}{z_t}$$

with iterations to rescale $\widehat{\boldsymbol{\Sigma}}$ to ensure (1.21) and stop if the determinant increases.

However, similar to sample variance estimation, these measures based on closing prices do not utilize the information within a reference period. Brandt and Diebold (2006) proposed a way to generalize variance estimators to covariance measures for financial data. Let the asset k be the portfolio made of assets i and j only and w_i, w_j be the corresponding weight. Furthermore, assume the market is efficient, that is we have $r_{k,t} = \text{w}_i r_{i,t} + \text{w}_j r_{j,t}$. Hence we have

$$\sigma_k^2 = \text{w}_i^2\sigma_i^2 + \text{w}_j^2\sigma_j^2 + 2\text{w}_i\text{w}_j\text{cov}(c_i, c_j). \tag{1.22}$$

Thus we can construct a covariance estimator

$$\widehat{\text{cov}}(c_i, c_j) = \frac{\hat{\sigma}_k^2 - \text{w}_i^2\hat{\sigma}_i^2 - \text{w}_j^2\hat{\sigma}_j^2}{2\text{w}_i\text{w}_j}, \tag{1.23}$$

where $\hat{\sigma}$ can be any variance estimator. Using range variance estimators, such a covariance estimator is more efficient compared to a sample covariance matrix. The drawback of such a method will be the availability as well as the liquidity of the portfolio. The less liquidity, the higher the chance of arbitrage opportunities, which voids equation (1.22). Furthermore, while the weight of the portfolio will change over time, it is not practical to have frequent refinancing as it will induce transaction costs. Thus we can not perform such calculations over a long period. A special case will be $\mathrm{w}_i = 1, \mathrm{w}_j = -1$. In such case, asset k represents the rate between the value of 1 unit of asset i and asset j. In particular, if assets i and j are quoted currency on the same numeraire, then asset z would be the exchange rate between currency i and j. Thus, the covariance estimator constructed from triangular arbitrary relationships demonstrated a better performance in currency trading as it is much more likely to overcome the cons we discussed for the portfolio cases.

On the other hand, correlation estimation could be more feasible since as long as the correlation matrix is positive definite, the covariance matrix is also positive definite. Furthermore, a correlation estimation could utilize a variance estimator from the previous section to construct a robust covariance estimator. As for correlation measures, Pearson, Kendall and Spearman correlations are very common. The Pearson correlation

$$\mathrm{cor}(c_i, c_j) = \frac{\sum_{t=1}^{N} c_{i,t}c_{j,t} - \frac{1}{N}\Big(\sum_{t=1}^{N} c_{i,t}\Big)\Big(\sum_{t=1}^{n} c_{j,t}\Big)}{\sqrt{\Big(\sum_{t=1}^{N} c_{i,t}^2\Big)\Big(\sum_{t=1}^{N} c_{j,t}^2\Big)}}, \tag{1.24}$$

is directly induced by sample covariance, with normality assumptions for both variables. Other assumptions include linearity and homoscedasticity. When the normality assumption fails, Spearman correlation is a non-parametric alternative applying Pearson correlation to the ranks of the data. Another non-parametric correlation is the Kendall rank correlation considering the number of concordant and discordant pairs.

Apart from non-normality, nonlinearity is another consideration. Hoeffding's D is a correlation measure that can pick up the nonlinear relationship by comparing the joint cumulative distribution function (cdf) with the product of marginal cdfs (Hoeffding, 1994). Zheng et al. (2012) proposed a generalized measure of correlation (GMC)

$$[\mathrm{GMC}(\mathsf{C}_i|\mathsf{C}_j),\ \mathrm{GMC}(\mathsf{C}_j|\mathsf{C}_i)] = \left\{1 - \frac{\mathrm{E}\{[\mathsf{C}_j - \mathrm{E}(\mathsf{C}_j|\mathsf{C}_i)]^2\}}{\mathrm{var}(\mathsf{C}_j)}, 1 - \frac{\mathrm{E}\{[\mathsf{C}_i - \mathrm{E}(\mathsf{C}_i|\mathsf{C}_j)]^2\}}{\mathrm{var}(\mathsf{C}_i)}\right\} \tag{1.25}$$

where C_i is the the random variable of which the realization is c_i. The $\mathrm{GMC}(\mathsf{C}_j|\mathsf{C}_i)$ is identical to R^2 for the regression model $\mathsf{C}_j = \mathrm{E}(\mathsf{C}_j|\mathsf{C}_i) + \epsilon$. GMC deals with asymmetries in variances and captures linear as well as nonlinear relationships between random variables and nests Pearson correlation as a special case. A test using kernel-based estimators for GMC is derived to test whether or not two random variables are symmetric in explaining variances. Recently, Zhang et al. (2024) proposed the quantile GMC to describe the nonlinear quantile relationship between vectors using a conditional quantile function. Moreover, when correlation is applied to non-independent observations or aggregated data, it may produce biased and specious results due to differing patterns between groups versus within groups. In light of this, Bakdash and Marusich (2017) proposed repeated measures correlation and introduced the R package `rmcorr` for implementation. Lastly, the distance correlation matrix uses the Euclidean distance between random vectors (Székely et al., 2007). Drouet Mari and Kotz (2001) provides a summary of correlation measures.

Despite these pools of correlation and correlation matrix measures, there is not much research on utilizing high-frequency opening, high, low and closing prices of assets in the financial market to construct correlation measures as it is for the case of univariate volatility measures.

1.4. Review of return and volatility models

1.4.1. Univariate models. The financial time series usually features heteroscedasticity, which introduces bias in estimators. Most estimators evaluate efficiency under homoscedasticity. Heteroskedasticity leads to inefficiency because the estimator does not appropriately weight observations based on their varying variances. Thus, the development of volatility models has significantly evolved over the years, as it would have a better performance in modelling heteroscedasticity. One important volatility model is the autoregressive conditional heteroskedasticity (ARCH) model (Engle, 1982), which introduced the concept of volatility clustering, capturing the phenomenon of high volatility periods followed by relative calm. Let $\{r_t\}$ be the time series of return. Then the ARCH(p) is defined as

$$\begin{aligned} r_t &= \sigma_t e_t, && (1.26)\\ \sigma_t^2 &= a_0 + \sum_{k=1}^{p} a_k \sigma_{t-k}^2 \epsilon_{t-k}^2 && (1.27)\\ e_t &\sim \mathrm{N}(0,1). && (1.28) \end{aligned}$$

A fundamental concept of the financial market is the risk-return trade-off. Engle et al. (1987) incorporate the idea by adding an impact of volatility in the mean process to build ARCH in mean, or ARCH-M. The ARCH-M is formulated as

$$\begin{aligned} r_t &= \delta\sigma_t + \sigma_t e_t, && (1.29)\\ \sigma_t^2 &= a_0 + \sum_{k=1}^{p} a_k \sigma_{t-k}^2 \epsilon_{t-k}^2 && (1.30)\\ e_t &\sim \mathrm{N}(0,1). && (1.31) \end{aligned}$$

Building upon this, the GARCH model overcomes the limitation of requiring a high-order ARCH model to describe long memory in volatility by incorporating lagged values of both returns and volatility, offering improved forecasting capabilities (Bollerslev, 1986). The standard GARCH(p, q) is given by

$$\begin{aligned} r_t &= \sigma_t e_t, && (1.32)\\ \sigma_t^2 &= a_0 + \sum_{k=1}^{p} a_k \sigma_{t-k}^2 e_{t-k}^2 + \sum_{k=1}^{q} b_k \sigma_{t-k}^2 && (1.33)\\ e_t &\sim \mathrm{N}(0,1). && (1.34) \end{aligned}$$

The volatility model in (1.33) with order $p = q = 1$ implies

$$\sigma_t^2 = \frac{1-b_1^t}{1-b_1} a_0 + b_1^t \sigma_0^2 + a_1 \sum_{k=0}^{t-1} b_1^k r_{t-k-1}^2.$$

Over time, various extensions of the basic GARCH model have been developed to address different characteristics in time series data. Engle and Bollerslev (1986) proposed IGARCH given by

$$\sigma_t^2 = a_0 + e_{t-1}^2 + \sum_{k=2}^{p} a_k(e_{t-k}^2 - e_{t-1}^2) + \sum_{k=1}^{q} b_k(\sigma_{t-k}^2 - e_{t-1}^2), \tag{1.35}$$

a restricted GARCH model to capture the high persistence of volatility shocks. By setting

$$\sum_{k=1}^{p} a_k + \sum_{k=1}^{q} b_k = 1 \tag{1.36}$$

in integrated variance with trend a_0 instead of varying around the long term mean $a_0/(1-\sum_{k=1}^{p} a_k - \sum_{k=1}^{q} b_k)$, the EGARCH model of Nelson (1991) given by

$$\begin{aligned} r_t &= \sigma_t e_t, && (1.37)\\ \log\sigma_t^2 &= a_0 + \sum_{k=1}^{p} a_k g(e_{t-k}) + \sum_{k=1}^{q} b_k \log\sigma_{t-k}^2 && (1.38)\\ g(e_t) &= \theta e_t + \gamma\big[|e_t| - \mathrm{E}(|e_t|)\big], && (1.39)\\ \mathrm{E}(e_t) &= 0, && (1.40) \end{aligned}$$

is an extension of GARCH(p, q), which can capture the asymmetric volatility in the financial time series, where negative shocks have a larger impact on volatility than positive shocks. Another popular extension is the GJR-GARCH model from Glosten et al. (1993), given by

$$\begin{aligned} r_t &= \sigma_t e_t, && (1.41)\\ \sigma_t^2 &= a_0 + \sum_{k=1}^{p}(a_k + \gamma_k \mathbf{1}_{e_{t-k}<0})\sigma_{t-k}^2 e_{t-k}^2 + \sum_{k=1}^{q} b_k \sigma_{t-k}^2 && (1.42)\\ e_t &\sim \mathrm{N}(0,1), && (1.43) \end{aligned}$$

which can capture the asymmetry in volatility caused by positive and negative returns. Further advancements include stochastic volatility models that allow volatility itself to be a stochastic process (Bergomi, 2015). The Threshold GARCH (TGARCH), proposed by Zakoian (1994), use a similar structure to model standard derivation as absolute residuals turn out to be more efficient than squared residuals in variance estimation for nonnormal distributions by Davidian and Carroll (1987). The TGARCH is given by

$$\begin{aligned} r_t &= \sigma_t e_t, && (1.44)\\ \sigma_t &= a_0 + \sum_{k=1}^{p}(a_k^+ \mathbf{1}_{e_{t-k}>0} - a_k^- \mathbf{1}_{e_{t-k}<0})\sigma_{t-k} e_{t-k} + \sum_{k=1}^{q} b_k \sigma_{t-k} && (1.45)\\ e_t &\sim \mathrm{N}(0,1). && (1.46) \end{aligned}$$

Similarly to ARCH-M in (1.29) to (1.30), we can have a risk premia term in return to construct the GARCH-M model, which captures the significant impact of volatility in bank stock return by Elyasiani and Mansur (1998). See Francq and Zakoian (2019) for more details of GARCH models.

GARCH models have shown efficiency in modelling volatility. However, these GARCH models treat volatility as a hidden process of price and hence, do not incorporate market information on volatilities. Alternatively, one can observe or measure volatility using high-frequency

returns. These realised volatility measures can be added to the GARCH model to obtain a realised EGARCH model given by

$$\begin{aligned} r_t &= \sigma_t e_t, && (1.47)\\ \log\sigma_t^2 &= a_0 + a_1 g(e_{t-1}) + b_1 \log\sigma_{t-1}^2 + c\log v_{t-1}, && (1.48)\\ \log v_t &= \alpha_0 + \alpha_1 g(e_t) + \beta_1 \log\sigma_t^2 + \varepsilon_t, && (1.49)\\ \varepsilon_t &\sim \mathrm{N}(0,\sigma_v^2) && (1.50) \end{aligned}$$

where the order $p = q = 1$ and v_t is any realised volatility measure on day t. However, these volatility measures, like returns, have random noise as well. Hence, volatility models were also proposed to model the volatility series similar to the return series to identify patterns and smooth out the noise for the observed variance (also the covariance matrix in multivariate setting) series as realisations of the actual variance (covariance matrix) process of returns (multiple returns).

Starting from the univariate case, Hsieh (1991) introduced the autoregressive variance AV(q) model for the conditional variance measures $v_t = \hat{\sigma}_t^2$ in (1.17), (1.8) and (1.15) given by

$$\ln v_t = b_0 + \sum_{k=1}^{q} b_{2k} \ln v_{t-k} + e_t, \qquad e_t \overset{i.i.d.}{\sim} \mathrm{N}(0,\sigma_e^2). \tag{1.51}$$

To incorporate the asymmetric impact of historical returns on the conditional variance, Li and Hong (2011) extended (1.51) to the following model

$$\ln v_t = b_0 + \sum_{i=1}^{p} \left\{ b_{1i} \left| \frac{r_{t-i}}{v_{t-i}^{1/2}} \right| + b'_{1i} \frac{r_{t-i}}{v_{t-i}^{1/2}} \right\} + \sum_{j=1}^{q} b_{2j} \ln v_{t-j} + e_t, \qquad e_t \overset{i.i.d.}{\sim} \mathrm{N}(0,\sigma_e^2). \tag{1.52}$$

However, taking the logarithmic transformation on low volatility measures can result in negative extreme residuals and negatively skewed residual distribution.

Instead of taking a log transformation, the CARR of Chou (2005) uses GARCH structure to model the conditional Parkinson range volatility measures of a financial asset's price movement. Results show that the CARR model is more robust in capturing the behaviour of extreme prices, such as price spikes or crashes, and can be used to forecast the likelihood of future extreme events. Let $R_t = h_t - l_t$ where h_t, l_t are defined in (1.5). The CARR model is defined as

$$\begin{aligned} R_t &= \lambda_t \epsilon_t, && (1.53)\\ \lambda_t &= a_0 + \sum_{k=1}^{p} a_k R_{t-k} + \sum_{k=1}^{q} b_k \lambda_{t-k}, && (1.54)\\ \epsilon_t &\overset{i.i.d.}{\sim} \mathrm{F}_v(\cdot|\phi)(\cdot), && (1.55) \end{aligned}$$

where the coefficients α_0, α_{1k}, $\beta_{1k} > 0$ measure the inherent uncertainty in the volatility measure, the short-term impact, and the long-term (persistence) impact of shocks to the range-based volatility measure, respectively and the conditional expectation at time t is $\mathrm{E}(v_t|\mathcal{F}_{t-1}) = \lambda_t$, $\mathcal{F}_t$ is nature sigma-algebra generated by OHLC data, and $\mathrm{F}(\cdot|\phi)$ denotes a certain distribution with positive real support, unit mean and shape parameter ϕ, for example, $\mathrm{Weibull}\Big(\phi, \Gamma\big(1+\frac{1}{\phi}\big)^{-1}\Big)$. For the process in (1.54) to be stationary, we also have the similar constraint as GARCH model, that is

$$\sum_{k=1}^{p} a_k + \sum_{k=1}^{q} b_k < 1. \tag{1.56}$$

In order to capture the asymmetric impact of the return, Chou (2005) also proposed the CARRX model, which is given by

$$\lambda_t = \beta_0 + \sum_{k=1}^{p} \beta_{1k} R_{t-k} + \sum_{k=1}^{q} \beta_{2k} \lambda_{t-k} + \sum_{k=1}^{h} \beta_{3k} X_k + \beta_4 |r_{t-1}| + \beta_5 r_{t-1}, \tag{1.57}$$

where β_{3k} measures the exogenous variable effects, and β_4 and β_5 represent the leverage effect.

Compared to GARCH models, CARR models are equivalent to modelling the range volatility measures instead of a point estimation $\sigma^2\epsilon_t^2$, which is more robust. Furthermore, the dynamics of volatility measure is now directly captured by $f(\cdot)$, instead of induced from return.

The fitted volatilities can be applied to return models to capture the heteroskedasticity of returns. Hence, this extended model is called a two-stage CARR-return model and is different from the GARCH and SV models that use only the information of returns. Instead, it uses both return and volatility information and models them as two stochastic processes.

Moreover, CARR-return models also differ from the realised GARCH models (Gerlach and Wang, 2016) since the latent volatility equation is omitted, although the realised GARCH model also has two stochastic processes for returns and realised volatilities. The choice of error distribution $f(\cdot)$ also plays an important role in capturing the volatility dynamics. Chan et al. (2019) considered the flexible GB2 family, including Weibull (Wei), Log-t, generalised gamma (GG) and GB2 distributions.

1.4.2. Mulivariate models. A simple approach is the vector autoregressive GARCH (VAR-GARCH) model and its VEC-GARCH model extension, which estimates the covariance matrix of multiple time series by considering the autoregressive and moving average components of contemporaneous and lagged covariance matrices (Bollerslev et al., 1988). The model uses a half-vectorization operator vec : $\boldsymbol{A} \to \boldsymbol{a}$ defined as

$$\text{vech}(\boldsymbol{\Sigma}) = \boldsymbol{a} = (\sigma_{1,1}, ..., \sigma_{d,1}, \sigma_{2,2}, ..., a_{d,2}, a_{3,3}, ..., a_{d,3} ..., a_{d,d}), \tag{1.58}$$

where $\boldsymbol{A}$ is a $d \times d$ symmetric matrix, $\boldsymbol{a}$ is a vector with $d(d+1)/2$ entries, $a_{i,j}$ is the i, j entry of $d \times d$ matrix A, to convert the covariance matrix into a vector. Let $\{\boldsymbol{r}_t = (r_{1,t}, ..., r_{d,t})^\top\}$ be the time series of return vector of n assets. The specification of VAR-GARCH(p,q) is given by

$$\boldsymbol{r}_t | \mathcal{F}_t \sim \text{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t) \tag{1.59}$$

$$\text{vech}(\boldsymbol{\Sigma_t}) = \boldsymbol{a}_0 + \sum_{k=1}^{p} \boldsymbol{A}_k \text{vech}(\boldsymbol{r}_{t-k} \boldsymbol{r}_{t-k}^\top) + \sum_{k=1}^{q} \boldsymbol{B}_k \text{vech}(\boldsymbol{\Sigma}_{t-k}), \tag{1.60}$$

where $\boldsymbol{a}_0$ is a $d(d+1)/2 \times 1$ column vector, $\{\boldsymbol{A}_k\}$ and $\{\boldsymbol{B}_k\}$ are $d(d+1)/2 \times d(d+1)/2$ matrices. Such formulation captures all possible linear relationships on lag (p, q). However, the covariance matrix processes using scalar coefficients are not guaranteed to be positive definite. Furthermore, the numbers of parameters grow at $o(d^4)$. While many multivariate GARCH models can be expressed as VAR-GARCH, it is not practical to fit the model. Researchers usually set the coefficient matrix to be diagonal in practice, which is equivalent to having a GARCH structure in both variance and covariance processes. However, such a formulation is too restrictive. Furthermore, we still can not guarantee that the covariance matrix is positively definite.

Another popular multivariate GARCH model is the BEKK model proposed by Engle and Kroner (1995). The formulation of the BEKK model is given by

$$\begin{aligned} \boldsymbol{r}_t|\mathcal{F}_{t-1} &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_t) && (1.61)\\ \boldsymbol{\Sigma}_t &= \boldsymbol{A}_0 + \sum_{k=1}^{p} \boldsymbol{A}_k \boldsymbol{r}_{t-k}\boldsymbol{r}_{t-k}^\top \boldsymbol{A}_k^\top + \sum_{k=1}^{q} \boldsymbol{B}_k \boldsymbol{\Sigma}_{t-k} \boldsymbol{B}_k^\top, && (1.62) \end{aligned}$$

where $\boldsymbol{A}_0$ is a $d \times d$ positive definite matrix, $\{\boldsymbol{A}_k\}$ and $\{\boldsymbol{B}_k\}$ are $d \times d$ matrices. The advantage of the BEKK model is that quadratic forms of coefficients allow more flexibility and preserve the positive definite property of the covariance matrix. However, it is hard to interpret the effect of coefficients.

These multivariate GARCH models tried to generalize the univariate GARCH model by covariance modelling. Another approach would be correlation modelling. A simple MGARCH model of this type is the CCC-GARCH model, which assumes that the correlation between the different time series is constant over time. The CCC-GARCH model is formulated as

$$\begin{aligned} \boldsymbol{r}_t|\mathcal{F}_{t-1} &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_t) && (1.63)\\ \boldsymbol{\Sigma}_t &= \boldsymbol{D}_t \boldsymbol{\mathcal{R}} \boldsymbol{D}_t, && (1.64)\\ \sigma_{ii,t} &= a_{i0} + \sum_{k=1}^{p} a_{i,k} r_{t-k}^2 + \sum_{k=1}^{q} b_{i,k} \sigma_{ii,t-k}, && (1.65) \end{aligned}$$

where $\sigma_{ii,t}$ is the (i,i) entry of $\boldsymbol{\Sigma}_t$, $\boldsymbol{D}_t$ is the diagonal matrix with i-th diagonal being $\{\sqrt{\sigma_{ii,t}}\}$ and $\boldsymbol{\mathcal{R}}$ is a correlation matrix. This model is useful for modelling multivariate data where the correlation between different time series is not expected to change significantly over time (**?**). The DCC-GARCH model is given by

$$\begin{aligned} \boldsymbol{r}_t|\mathcal{F}_{t} &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_t) && (1.66)\\ \boldsymbol{\Sigma}_t &= \boldsymbol{D}_t \boldsymbol{\mathcal{R}}_t \boldsymbol{D}_t, && (1.67)\\ \sigma_{ii,t} &= a_{i0} + \sum_{k=1}^{p} a_{i,k} r_{t-k}^2 + \sum_{k=1}^{q} b_{i,k} \sigma_{ii,t-k}, && (1.68)\\ \boldsymbol{Q}_t &= \boldsymbol{\Psi} + \alpha(\boldsymbol{r}_{t-1}\boldsymbol{r}_{t-1}^\top - \boldsymbol{\Psi}) + \beta(\boldsymbol{Q}_{t-1} - \boldsymbol{\Psi}) && (1.69)\\ \boldsymbol{\mathcal{R}}_t &= (\boldsymbol{I} \otimes \boldsymbol{Q}_t)^{-1/2} \boldsymbol{Q}_t (\boldsymbol{I} \otimes \boldsymbol{Q}_t)^{-1/2}, && (1.70) \end{aligned}$$

where

$$\boldsymbol{\Psi} = \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{e}_t \boldsymbol{e}_t^\top \tag{1.71}$$

is the unconditional standardised residuals' covariance matrix, and $\otimes$ is the Kronecker product. The DCC-GARCH model (Engle, 2002) captures the dynamic correlation dynamic between time series across time by using a combination of past values of correlation and error terms. A disadvantage of DCC-GARCH is the lack of practical interpretation of latent process $\{\boldsymbol{Q}_t\}$. Furthermore, coefficients describing dynamics of correlations, α and β, are scalars, which is too restrictive in higher dimensional modelling. Bauwens et al. (2006) provided an overview of the development of these models. In order to model the correlation matrix directly, Hansen et al. (2014) proposed the Beta GARCH (BGARCH) model using Fisher transformation $\rho \to f_\rho(\rho) \triangleq \frac{1}{2}\log\frac{1+\rho}{1-\rho}$, a one-to-one mapping from $(-1,1)$ to $\mathbb{R}$. The correlation model is given by

$$f_\rho(\rho_{ij,t}) = a_{0,ij} + a_{1,ij} f_\rho(v_{ij,t-1}) + b_{1,ij} f_\rho(\rho_{ij,t-1}) \tag{1.72}$$

where $v_{ij,t}$ is a realised correlation measure of $\rho_{ij,t}$ and the variance model for each component i can be (1.49). For the realised BGARCH model, the additional model is given by

$$f_\rho(v_{ij,t}) = \alpha_{0,ij} + \alpha_{1,ij} f_\rho(\rho_{ij,t}) + \varepsilon_{ij,t}. \tag{1.73}$$

If the residuals $(\varepsilon_{i,t}, \varepsilon_{j,t}, \varepsilon_{ij,t})$ are correlated, they can also be modelled by a multivariate normal distribution.

Although the number of parameters in these MGARCH models may grow at a rate of $o(d)$, the computational time still increases exponentially with the dimensionality, posing significant challenges in high-dimensional settings. To account for the curse of dimensionality, Hoff and Niu (2012) proposed a covariance regression model based on the factor model of returns. Let $\{\boldsymbol{x}_t\}$ be the time series of $1 \times \mathfrak{l}$ column vector of $\mathfrak{l}$ explanatory variables. The covariance regression model aims to provide a model for $\text{cov}(\boldsymbol{r}_t|\boldsymbol{x}_t)$. Thus, the high-dimensional covariance matrix converts into a lower-dimensional time series modelling. The simplest $d \times d$ dimensional covariance regression is given by

$$\boldsymbol{\Sigma_t}|\boldsymbol{x}_t = \boldsymbol{C} + \boldsymbol{B}\boldsymbol{x}_t\boldsymbol{x}_t^\top\boldsymbol{B}^\top, \tag{1.74}$$

where $\boldsymbol{B}$ is a $d \times \mathfrak{l}$ dimensional coefficient matrix. One of the key advantages of the covariance regression approach lies in its ability to offer a clearer interpretation of the contributions of individual covariates. This clarity facilitates a deeper understanding of the mechanisms driving market behaviour, as it disentangles the effects of observable factors—such as macroeconomic indicators, sector-specific shocks, or asset characteristics—from the residual volatility that arises from unobserved or random influences. Such a decomposition is particularly valuable in dynamic and multi-dimensional settings, where understanding the drivers of variability and correlation is essential for risk management, portfolio optimisation, and asset pricing.

Moreover, the covariance regression model accommodates changes in the structure of the covariance matrix over time as well, making it a flexible tool for analysing evolving market conditions. By explicitly modelling the covariate contributions, it allows researchers and practitioners to assess how external factors, such as interest rates, geopolitical events, or market sentiment, impact asset volatilities and correlations. This capability is especially useful in periods of market stress or structural shifts, where traditional methods may fail to capture the nuanced interplay between covariates and covariance dynamics. Van Jaarsveldt et al. (2024) showed that by further extracting implicit factors from financial market factors, the performance of the portfolio built based upon the covariance regression model outperformed the portfolios built based upon DCC-GARCH in simulations. A survey of the MGARCH model can be found in Bauwens et al. (2006).

1.5. Aims and contributions

We have three important aims and contributions in this thesis, focusing on multivariate extension to current volatility measures and models.

By moving beyond single-asset volatility models, higher-dimensional approaches capture the complex interrelationships between multiple assets, providing a deeper understanding of financial risk and more accurate predictive capabilities. The ability to model these interactions has transformed how investors and institutions manage portfolios, assess market risks, and price complex financial products, leading to more robust and reliable financial strategies. This multivariate extension has an impact on not only portfolio construction but also portfolio risk management. In a multi-asset portfolio, the risk is not simply a sum of the individual volatilities of the assets but is influenced heavily by how the assets interact.

Our first aim and contribution is to generalise PK volatility measures to higher dimensions. The aim stems from the need for efficient covariance matrix measures as input measures to multivariate volatility models such as the MCARR models to better understand and manage the complex interactions among multiple financial assets in a portfolio. We are interested in identifying whether the clustering in covariance is driven by volatility or correlation.

Then our second aim and contribution is to propose a direct extension of the CARR model to the MCARR model using the Wishart distribution for the positive definite covariance matrices. We further propose two approaches for modelling the mean of the Wishart distribution, borrowing the ideas of BEKK-GARCH and DCC-GARCH models. Then, the fitted volatility can be applied to the multivariate return model to forecast multiple returns. This model is called the two-stage volatility-return or simply the MCARR-return model. Our proposed models enable more precise forecasting of portfolio volatility and risk metrics, such as Value at Risk (VaR) and Conditional Value at Risk (CVaR). Forecasts of VaR and CVaR are essential in portfolio management because they not only ensure compliance with regulatory requirements, such as stress testing and capital adequacy regulations, but also play a critical role in maximising the wealth of stakeholders. We conduct simulation experiments to study several issues to fine-tune the models. The models are implemented using a Bayesian approach via the user-friendly `RStan` package.

Another critical contribution of higher-dimensional volatility estimation is its ability to provide insights into systemic risk. Unlike idiosyncratic risk, which affects individual firms or sectors, systemic risk refers to non-diversifiable risk, that is, the potential for a shock in one part of the financial system to propagate across other markets or institutions, leading to widespread financial instability. Hence, our third aim and contribution is to apply the two-stage MCARR-return models to model and predict returns, volatilities and correlation to gain insights of different markets. Moreover, these models should be fine-tuned to adapt to diverse market environments such as the US stock and cryptocurrency markets.

We aim to provide two real applications of our proposed models. The first application focuses on how volatility in a major market like the U.S. can influence emerging markets or commodities. As global markets become more interconnected, higher-dimensional volatility models can contribute to a more comprehensive analysis of international risk factors. This capability is crucial for understanding how global events, such as geopolitical crises or economic policy changes, can ripple across markets and affect asset prices and volatility worldwide. Our high-dimensional volatility models will provide a broader view of global financial interactions and so enable more informed decisions in a rapidly globalising financial environment.

Secondly, as the global financial crisis of 2007–2008 underscored the importance of understanding how risks in one asset class, sector, or market can spill over into others, our second application will study the impact of COVID-19 pandemic. Our proposed volatility models can capture the interconnectedness between assets and markets across the stages of pandemic development, providing a clearer picture of how volatility shocks due to the pandemic can spread. For instance, during times of financial stress resulting from the pandemic, correlations between assets that typically behave independently can suddenly increase, as was seen during the financial crisis. Hence, considerations should be directed to incorporating factors from different sectors and asset classes into a unified framework to provide a clearer understanding of global financial risks due to the pandemic. It is particularly important for investors with internationally diversified portfolios, who must account for cross-border volatility spillovers and currency risks. These models allow

regulators and financial institutions to identify periods of heightened systemic risk and to take preemptive action, such as adjusting capital requirements or altering risk management practices, to mitigate the impact of such shocks.

In summary, the generalisation of volatility measures and models to higher dimensions will make profound contributions to modern finance. It will revolutionise portfolio risk management by offering more accurate insights into the relationships between assets, allowing for better diversification and risk mitigation strategies. In the realm of derivatives, it will enable more precise pricing and hedging of complex multi-asset products. Moreover, our real applications will provide critical insights into systemic risk and global market dynamics, enhancing the ability of financial institutions and regulators to manage and respond to volatility shocks. As financial markets continue to evolve, the contribution of our proposed high-dimensional volatility measures and models will remain fundamental to advancing risk management and pricing strategies in an increasingly interconnected and complex world.

The remainder of this thesis is organized as follows: Chapter 2 explores the properties of efficient range correlation measures and proposes a range correlation estimator. The chapter investigates the properties of the proposed estimator through extensive simulation studies. Chapter 3 generalises the MCARR model by introducing two formulations and examines their properties and performance using simulation analyses. Chapter 4 presents two applications of the MCARR-return models, utilising the proposed range correlation measure to demonstrate its utility in practical scenarios. Finally, Chapter 5 provides a summary of this thesis and outlines directions for future research.

CHAPTER 2

Measures of Variance-Covariance matrices

2.1. Introduction

In finance, volatility estimation is essential for investors, traders, and financial institutions for risk management, serving as a crucial metric for assessing and predicting the degree of price fluctuations in assets. Moreover, correlation measurement and estimation also play a pivotal role in portfolio management when a large number of stocks are invested, with some designated for hedging.

Since the seminal work of Parkinson (1980) introduced the range-based volatility measure, the estimation of volatility using OHLC data has become an increasingly prominent area of research in financial econometrics. Numerous extensions of the range volatility measure have been proposed by various researchers, each aimed at enhancing the accuracy and robustness of volatility estimation.

One of the key advantages of range-based volatility estimators lies in their ability to incorporate additional price information beyond simple return calculations. Unlike traditional return-based measures, which rely solely on closing prices, range-based estimators utilize the H and L prices within a given period, allowing them to capture a more comprehensive picture of price fluctuations within an observation window. This inclusion of extreme price movements enables range-based volatility estimators to provide more efficient estimates of market variability.

In terms of the multivariate volatility measure, despite the extensive research on return-based covariance matrix estimation, there has been relatively little focus on the development of range-based covariance matrix estimation methods. Some return-based covariance matrix estimation including the Covariance Estimation with Linear Shrinkage (Ledoit and Wolf, 2003, 2004), the Oracle Approximating Shrinkage Estimator (Chen et al., 2010b), etc, are documented in the `R CovTools` package `https://cran.r-project.org/web/packages/CovTools/CovTools.pdf`. Moreover, the Principal Orthogonal Complement Thresholding (POET) (Fan et al., 2013) estimator is included in `https://cran.r-project.org/web/packages/POET/index.html`. However, as range-based volatility estimators are shown to be more efficient, deriving range-based covariance matrix estimators and comparing them to these return-based estimators presents a promising research area.

Addressing this research gap, our contribution lies in proposing a range-based correlation measure, which serves as a foundation for constructing a corresponding range-based covariance measure when combined with range-based volatility measures. Our proposed range correlation measure is designed to capture the dynamic relationships between assets more effectively by leveraging the information contained in the range of high and low prices over a given time interval. This approach offers several advantages over traditional sample covariance methods, particularly in scenarios characterized by high volatility or irregular trading patterns. By utilizing range data, the measure inherently accounts for the full extent of price movements within a period, providing

a more robust representation of asset co-movements compared to measures based solely on closing prices. Furthermore, this range correlation measure can be seamlessly integrated with range volatility measures to construct a range-based covariance matrix. This induced covariance matrix retains the desirable properties of both range volatility and range correlation measures, making it a powerful tool for modelling and forecasting multi-asset volatility and correlation structures. The ability to combine these components ensures that the covariance matrix is both consistent with the observed data and capable of adapting to changing market conditions. The introduction of this range-based approach represents a significant contribution to the field of financial modelling, as it expands the toolkit available for covariance estimation, particularly in environments where traditional methods may be inadequate. By addressing the limitations of existing techniques, our range covariance measure provides a robust alternative for applications such as risk management, portfolio optimization, and financial forecasting, enabling practitioners to better capture and analyze the complex dynamics of financial markets.

Our second contribution is to apply the nearest positive definite (NPD) matrix transformation to tackle the challenge of non-positive definite covariance and correlation matrices. We evaluate the impact of such transformation again using simulation studies and find that the transformation improves the matrix performance. Our last contribution is to demonstrate applications to one and two sample tests. We also modify the two sample tests proposed in Cai et al. (2013) for match-pair samples using our proposed range-based covariance matrix estimators and prove the asymptotic convergence of the modified test statistic. Results for the range-based covariance matrix estimator are better than the sample covariance matrix estimator in terms of both the power of detecting inconsistency and type I error. In summary, the results confirm the efficiency of our proposed covariance matrix estimators.

The remaining chapter is organised as follows. Section 2.2 reviews the development of our proposed correlation measures. Section 2.3 discusses some necessary properties for an efficient correlation measure. Section 2.4 proposes the correlation matrix measures using Opening, High, Low and Closing price information. Section 2.5 performs two simulation studies to assess the performance of our proposed correlation matrix measures under heteroskedastic variance and transformation due to non-positive definite matrices. An application of our proposed range-based covariance matrices in statistical inference is provided in Section 2.7. Lastly, Section 2.8 summarises our findings and discusses further extensions to our proposed range-based correlation and covariance matrix measures.

2.2. Development of range correlation measures

In this chapter, we aim to develop an estimator of correlation on day t using all intraday periods τ ranging from 1 to n. Let asset 1 and asset 2 be arbitrary assets and asset 3 be the exchange rate between assets 1 and 2. Assume that the market is arbitrage-free, that is, $P_{3,t,\tau} = P_{1,t,\tau}/P_{2,t,\tau}$ for all τ and t. We have

$$c_{3,t,\tau} = \log \frac{C_{3,t,\tau}}{O_{3,t,\tau}} = \log \frac{C_{1,t,\tau}/C_{2,i,t,\tau}}{O_{1,t,\tau}/O_{2,t,\tau}} = c_{1,t,\tau} - c_{2,t,\tau}, \tag{2.1}$$

$$\text{Var}(\boldsymbol{c}_{3,t}) = \text{Var}(\boldsymbol{c}_{1,t} - \boldsymbol{c}_{2,t}) = \text{Var}(\boldsymbol{c}_{1,t}) + \text{Var}(\boldsymbol{c}_{2,t}) - 2\text{Cov}(\boldsymbol{c}_{1,t}, \boldsymbol{c}_{2,t}). \tag{2.2}$$

Recall that we set

- $o_{i,t,\tau}$: $\log O_{i,t,\tau}/C_{i,t,\tau-1}$, the opening jump,
- $h_{i,t,\tau}$: $\log H_{i,t,\tau}/O_{i,t,\tau}$,
- $l_{i,t,\tau}$: $\log L_{i,t,\tau}/O_{i,t,\tau}$,

- $c_{i,t,\tau} : \log C_{i,t,\tau}/O_{i,t,\tau}$,

where $(O_{i,t,\tau}, H_{i,t,\tau}, L_{i,t,\tau}, C_{i,t,\tau})$ are the open, high, low and close price of asset i on day t at time τ respectively. Applying the Parkinson measure to asset 3 in (1.23), a range-based covariance measure can be defined.

Definition 2.2.1. *Let asset 1 and asset 2 be arbitrary assets and asset 3 be the exchange rate between assets 1 and 2. Applying the Parkinson measure to asset 3 in* (1.6)*, a range-based sample covariance measure can be defined as*

$$\mathrm{cov}_{\mathrm{a}}(r_{1,t}, r_{2,t}) := \frac{\sum_{\tau=1}^{n}(h_{1,t,\tau} - l_{1,t,\tau})^2 + \sum_{\tau=1}^{n}(h_{2,t,\tau} - l_{2,t,\tau})^2 - \sum_{\tau=1}^{n}(h_{3,t,\tau} - l_{3,t,\tau})^2}{8n \log 2}. \quad (2.3)$$

This method has been demonstrated to be robust compared to the sample covariance matrix in the presence of microstructure noise, such as bid-ask bounce and asynchronous trading, which are commonly observed in real markets, as discussed in Brandt and Diebold (2006). The induced variance measure from (2.3) is exactly Parkinson's measure as $H_{12,t} = L_{12,t} = 1$ for the special case if assets 1 and 2 are the same.

Proposition 2.2.2. *Assume that $P_{3,t,\tau} = P_{1,t,\tau}/P_{2,t,\tau}$ for all t, τ. Let the covariance estimation be defined in* (2.3)*. Then the induced correlation measure is given by*

$$\begin{aligned}\mathrm{cor}_{\mathrm{a}}(r_{1,t}, r_{2,t}) &= \frac{\mathrm{cov}_{\mathrm{a}}(r_{1,t}, r_{2,t})}{\sqrt{\mathrm{cov}_{\mathrm{a}}(r_{1,t}, r_{1,t})\mathrm{cov}_{\mathrm{a}}(r_{2,t}, r_{2,t})}} \\ &= \frac{\sum_{\tau=1}^{n}(h_{1,t,\tau} - l_{1,t,\tau})^2 + \sum_{\tau=1}^{n}(h_{2,t,\tau} - l_{2,t,\tau})^2 - \sum_{\tau=1}^{n}(h_{3,t} - l_{3,t,\tau})^2}{2\sqrt{\sum_{\tau=1}^{n}(h_{1,t,\tau} - l_{1,t,\tau})^2 \times \sum_{\tau=1}^{n}(h_{2,t,\tau} - l_{2,t,\tau})^2}},\end{aligned}$$

is a valid correlation estimation, i.e

$$-1 \leq \mathrm{cor}_{\mathrm{a}}(r_{1,t}, r_{2,t}) \leq 1.$$

Proof. See Appendix 2.A. □

However, such an unbiased and arbitrary-free method may encounter significant limitations in real applications, where Asset 3, the exchange rate between Asset 1 and Asset 2, may not be actively traded or even listed in the market. The absence of market-traded data for Asset 3 restricts the availability of observed price information, which is crucial for accurate computation and validation of the proposed measure. Although it is possible to manually compute the exchange rate, all intermediate trading data for asset 1 and asset 2 are required to compute the high and low of the exchange rate, which is significantly more challenging to obtain from most data sources, in order to derive the OHLC data for the exchange rate. This limitation significantly undermines the method's applicability in real-world contexts where reliable and timely market information is critical for effective analysis. Furthermore, even in situations where direct trading exists for the exchange rate between Asset 1 and Asset 2, discrepancies in liquidity levels between these markets could pose additional challenges. Such differences would violate the efficient market hypothesis, which assumes uniform availability of information and market conditions across related assets. The violation of this assumption could introduce biases or inaccuracies into the estimation process, as the proposed measure relies on efficient price discovery to ensure validity. Moreover, geographical differences between markets, as well as variations in the structure of listed assets, can pose significant challenges. For example, consider the case where Asset 1 represents the share price of the Taiwan Semiconductor Manufacturing Company (TSMC) quoted in New Taiwan dollars (NTD),

Asset 2 corresponds to the exchange rate between the New Taiwan dollar and the US dollar, and Asset 3 represents the share price of TSMC quoted in US dollars (USD). TSMC is listed on the Taiwan Stock Exchange (code: 2330) and is also traded on the New York Stock Exchange (NYSE; code: TSM) as American Depositary Receipts (ADRs) with a ratio of 5:1, meaning that one unit of TSM corresponds to five units of ordinary shares on the original market. Although all three assets are actively traded, challenges arise when attempting to apply the arbitrage-free range covariance measure due to the lack of overlapping trading hours between the Taiwan Stock Exchange and the NYSE. This temporal misalignment prevents synchronous price observations, thereby failing the calculation of covariance measures based on overlapping trading sessions. Additionally, the ADR structure introduces further complexity. Investors can convert depositary receipts (TSM) into ordinary shares (2330) but not vice versa. This asymmetry adds a premium to the price of TSM, as the option to convert creates additional value. Consequently, after adjusting to the same numeraire, the price of TSM often exceeds five times the value of an ordinary share, reflecting this embedded premium. These factors highlight the practical challenges in implementing range-based covariance measures across geographically and structurally disparate markets. Addressing these issues may require adjustments to the method or the incorporation of additional mechanisms to account for variations in market liquidity and trading activity. In addition, an insufficient volume of trading would cause the scaling coefficient in the PK measure to deviate significantly from its asymptotic value, being $4\log 2$.

A natural question arises as to whether a covariance estimator can be constructed solely based on $(\boldsymbol{h}_1 - \boldsymbol{l}_1, \boldsymbol{h}_2 - \boldsymbol{l}_2)$. A trivial way to extend Parkinson measure would be

$$\mathrm{cov}_{\mathrm{b}}(r_{1,t}, r_{2,t}) = \sum_{\tau=1}^{n} \frac{(h_{1,t,\tau} - l_{1,t,\tau})(h_{2,t,\tau} - l_{2,t,\tau})}{4n\log 2}. \tag{2.4}$$

However, such an estimator is always positive. Later we will show that estimator consisting of $(\boldsymbol{h}_1 - \boldsymbol{l}_1, \boldsymbol{h}_2 - \boldsymbol{l}_2)$ only will not be efficient in general. In order to extend $\mathrm{cov}_{\mathrm{b}}(\cdot,\cdot)$ so that it can capture negative correlation, we can utilize signs of $c_{1,t,\tau}$ and $c_{2,t,\tau}$. We expect a positive contribution of covariance estimation if $c_{1,t,\tau}$ and $c_{2,t,\tau}$ have same the sign and vice versa. So the range-based covariance measure is defined as:

$$\mathrm{cov}_{\mathrm{c}}(r_{1,t}, r_{2,t}) = \sum_{\tau=1}^{n} \frac{\mathrm{Sign}(c_{1,t,\tau}c_{2,t,\tau})(h_{1,t,\tau} - l_{1,t,\tau})(h_{2,t,\tau} - l_{2,t,\tau})}{4n\log 2}. \tag{2.5}$$

The simulation study suggested that $\mathrm{cov}_{\mathrm{c}}(\cdot,\cdot)$ biases towards 0. To tackle the bias problem, one may find weights $\mathrm{w}_+, \mathrm{w}_-$ for the positive and negative parts of $\{c_{1,t,\tau}c_{2,t,\tau}\}$ by solving

$$\begin{cases} n_+\mathrm{w}_+ + n_-\mathrm{w}_- = n \\ \mathrm{w}_+/\mathrm{w}_- = f(n_+/n_-). \end{cases} \tag{2.6}$$

where $f(\cdot)$ is some function such as $f(x) = \sqrt{x}$, $n_+ = \sum_{\tau=1}^{n} \mathbf{1}(r_{1,t,\tau}r_{2,t,\tau} > 0)$ and $n_- = \sum_{\tau=1}^{n} \mathbf{1}(r_{1,t,\tau}r_{2,t,\tau} < 0)$. Using these weights, the proposed covariance estimator is given by

$$\mathrm{cov}_{\mathrm{d}}(r_{1,t}, r_{2,t}) = \sum_{\tau=1}^{n} \frac{\left[\mathbf{1}(r_{1,t,\tau}r_{2,t,\tau} > 0)\,\mathrm{w}_+ - \mathbf{1}(r_{1,t,\tau}r_{2,t,\tau} < 0)\,\mathrm{w}_-\right](h_{1,t,\tau} - l_{1,t,\tau})(h_{2,t} - l_{2,t})}{4n\ln 2}. \tag{2.7}$$

The simulation study showed that $f(x) = \sqrt{x}$ could provide a good estimator. However, it does not guarantee the correlation to be within $[-1, 1]$. The correlation is within $[-1, 1]$ only when $\max(|\mathrm{w}_+|, |\mathrm{w}_-|) \leq 1$. We recognize that it is more appropriate to propose a direct estimator

for correlation, given the inherent constraints on correlation values. Furthermore, a correlation estimator allows us to take advantage of any robust volatility estimator in estimating volatility. These experiences lead to proposing the eight pairs of correlation s_k in (2.25) in Section 2.4.1. We first investigate what properties an efficient correlation estimator should have.

2.3. Invariance property of a correlation estimator

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\boldsymbol{\kappa} : \Omega \to \mathbb{R}^d$ be a d-dimension random vector, whose joint density $f(\cdot; \boldsymbol{\theta})$ depends on an parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. We are particularly interested in the scenario where the objective is to estimate a projection of $\boldsymbol{\theta}$. We start with the following proposition.

Proposition 2.3.1. *Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \boldsymbol{\Theta} = \boldsymbol{\Theta}_1 \times \boldsymbol{\Theta}_2$. Let $D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1)$ be any estimator for $\boldsymbol{\theta}_2$ given $\boldsymbol{\theta}_1$ and $L(\boldsymbol{D(\kappa; \theta_1), \theta_2})$ be an arbitrary loss function independent of $\boldsymbol{\theta}_1$ such that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $L(\cdot; \boldsymbol{\theta}_2)$ is a convex function. Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a monotonic transformation, and $T' : \boldsymbol{\Theta}_1 \to \boldsymbol{\Theta}_1$ be a invertable transformation. Suppose that, for all $\boldsymbol{\theta}_2 \in \boldsymbol{\Theta}_2$ and all $\boldsymbol{\kappa}$ in the support of $f_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}$,*

$$|\det \boldsymbol{J}_T| f(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = f(\boldsymbol{\kappa}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2).$$

Let

$$R(D(\cdot); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \int_{\mathbb{R}^d} L(D(\cdot), \boldsymbol{\theta}_2) f(\boldsymbol{\kappa}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\kappa},$$

be the expected loss of estimator under $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Then at least one of the following holds:

1. $R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > R(D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$,
2. $R(D(\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > R(D(T^{-1}\boldsymbol{\kappa}; \boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$,
3. $R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \geq R(wD(\boldsymbol{\kappa}; \boldsymbol{\theta}_1) + (1-w)D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \quad \forall w \in [0, 1]$.

Proof. See appendix 2.B. □

Remark 2.3.2. Note that $D(T\boldsymbol{\kappa}; T'\theta_1)$ also defines an estimator for $\boldsymbol{\theta}_2$ under $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. If we wish to build a class of efficient estimators, then we expect that

$$R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \leq R(D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2). \tag{2.8}$$

Otherwise, $D(T\boldsymbol{\kappa}|T'\theta_1)$ is a more efficient estimator for $\boldsymbol{\theta}_2$ under $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Similarly, we also expect

$$\begin{aligned} R(D(\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &\leq R(D(T^{-1}\boldsymbol{\kappa}; \boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \\ R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &\leq R(wD(\boldsymbol{\kappa}; \boldsymbol{\theta}_1) + (1-w)D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \end{aligned}$$

for any $w \in [0, 1]$. Then by proposition 2.3.1, we can see that we can always achieve a more efficient estimator either given $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ or $(T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ unless we have

$$\begin{aligned} R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= R(D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \\ R(D(\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= R(D(T^{-1}\boldsymbol{\kappa}; \boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \\ R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= R(wD(\boldsymbol{\kappa}; \boldsymbol{\theta}_1) + (1-w)D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \forall w \in [0, 1]. \end{aligned} \tag{2.9}$$

By convexity of $L(\cdot|\boldsymbol{\theta}_1)$, equation (2.9) holds if and only if

$$D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1) = D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1), \tag{2.10}$$

almost surely.

Let $\boldsymbol{\kappa}_i = (\boldsymbol{h}_i, \boldsymbol{l}_i, \boldsymbol{c}_i),\ i = 1, 2$ and $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2)$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ where $\boldsymbol{\theta}_1 = (\sigma_1, \sigma_2), \boldsymbol{\theta}_2 = \rho$. Assume that the loss function $L(\cdot; \rho)$ is a convex function and is independent of standard deviations σ_1, σ_2 and $D(\cdot|\sigma_1, \sigma_2)$ be an estimator given σ_1, σ_2. Define $T(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2) = (\lambda_1\boldsymbol{\kappa}_1, \lambda_2\boldsymbol{\kappa}_2), T'(\sigma_1, \sigma_2) = (\lambda_1\sigma_1, \lambda_2\sigma_2)$, where $\lambda_1, \lambda_2 > 0$. We have the following corollary.

Corollary 2.3.3. *An efficient correlation estimator conditioning on σ_1 and σ_2 satisfy*

$$D(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2; \sigma_1, \sigma_2) = D(T(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2); T'(\sigma_1, \sigma_2)) \quad \forall \sigma_1, \sigma_2 \in \mathbb{R}^+ \quad a.s.. \tag{2.11}$$

Proof. Inmediately follows from Remark 2.3.2 □

If we wish to build an estimator which is feasible to apply to general cases, the estimator should be independent of σ_1, σ_2, i.e., we require

$$D_{\sigma_1,\sigma_2}(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2) = D_{\sigma_1',\sigma_2'}(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2) = D(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2) \quad \forall \sigma_1, \sigma_2, \sigma_1', \sigma_2' \in \mathbb{R}^+.$$

Combining with (2.11)

$$D(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2) = D \circ T(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2)$$

for any (λ_1, λ_2) almost surely.

Definition 2.3.4. *Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a fixed measure-preserving transformation. We call a transformation probability-invariant if for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{\kappa}$ in the support of $f_{\boldsymbol{\theta}}$,*

$$f(T(\boldsymbol{\kappa}); \boldsymbol{\theta}) = f(\boldsymbol{\kappa}; \boldsymbol{\theta}).$$

We call a transformation semi-probability-invariant if for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{\kappa}$ in the support of $f_{\boldsymbol{\theta}}$,

$$f(T(\boldsymbol{\kappa}); \boldsymbol{\theta}) = f(\boldsymbol{\kappa}; -\boldsymbol{\theta}).$$

Corollary 2.3.5. *T_1 and T_2 are both probability-invariant transformations. T_3 is a semi-probability-invariant transformation.*

We now define the following transformations:

$$\begin{aligned}
T_1(\boldsymbol{h}_1, \boldsymbol{l}_1, \boldsymbol{c}_1, \boldsymbol{h}_2, \boldsymbol{l}_2, \boldsymbol{c}_2) &= (\boldsymbol{h}_2, \boldsymbol{l}_2, \boldsymbol{c}_2, \boldsymbol{h}_1, \boldsymbol{l}_1, \boldsymbol{c}_1),\\
T_2(\boldsymbol{h}_1, \boldsymbol{l}_1, \boldsymbol{c}_1, \boldsymbol{h}_2, \boldsymbol{l}_2, \boldsymbol{c}_2) &= (\boldsymbol{h}_1 - \boldsymbol{c}_1, \boldsymbol{l}_1 - \boldsymbol{c}_1, -\boldsymbol{c}_1, \boldsymbol{h}_2 - \boldsymbol{c}_2, \boldsymbol{l}_2 - \boldsymbol{c}_2, -\boldsymbol{c}_2),\\
T_3(\boldsymbol{h}_1, \boldsymbol{l}_1, \boldsymbol{c}_1, \boldsymbol{h}_2, \boldsymbol{l}_2, \boldsymbol{c}_2) &= (\boldsymbol{h}_1, \boldsymbol{l}_1, \boldsymbol{c}_1, -\boldsymbol{l}_2, -\boldsymbol{h}_2, -\boldsymbol{c}_2).
\end{aligned}$$

T_1 defines a swap between asset 1 and asset 2. T_2 defines represents a time-reversed transform of asset 1 and asset 2. Consider the case $\sigma_1 = \sigma_2 = 1$. We have the following proposition.

Proposition 2.3.6. *Let $S_{1,t}, S_{2,t}$ be two geometric Brownian motions defined as*

$$d\log S_{1,t} = \sigma_1 dW_{1,t}, \quad d\log S_{2,t} = \sigma_2(\rho dW_t + \sqrt{1-\rho^2}dW_{2,t}),$$

where $W_{1,t}, W_{2,t}$ are independent Wiener processes. Let $r_{i,t+\tau^*} = \log(S_{i,t+\tau^*}/S_{i,t+0}),\ i = 1, 2,\ \tau^* \in (0, 1)$,

$$h_{i,t} = \max_{0\le\tau^*\le 1} r_{i,t+\tau^*}, \quad l_{i,t} = \min_{0\le\tau^*\le 1} r_{i,t+\tau^*}, \quad c_{i,t} = r_{i,t}, \quad i = 1, 2.$$

Define $\boldsymbol{\theta} = (\rho, \sigma_1, \sigma_2)$ and $\boldsymbol{\kappa}_t = (h_{1,t}, l_{1,t}, c_{1,t}, h_{2,t}, l_{2,t}, c_{2,t})$. Let $f(\cdot; \boldsymbol{\theta})$ be the density function of $\boldsymbol{\kappa}_t$. Then we have

$$f(T_2(\boldsymbol{\kappa}_t); \boldsymbol{\theta}) = f(\boldsymbol{\kappa}_t; \boldsymbol{\theta})$$

where $T : \boldsymbol{R}^6 \to \boldsymbol{R}^6$ is given by

$$T_2(\boldsymbol{\kappa}) = (h_{1,t} - c_{1,t}, l_{1,t} - c_{1,t}, -c_{1,t}, h_{2,t} - c_{2,t}, l_{2,t} - c_{2,t}, -c_{2,t})^\top.$$

Proof. See appendix 2.C. □

Using the following lemmas, we expect the following invariance property of an efficient estimator almost surely:

$$\begin{aligned} D(\boldsymbol{h}_1, \boldsymbol{l}_1, \boldsymbol{c}_1, \boldsymbol{h}_2, \boldsymbol{l}_2, \boldsymbol{c}_2) &= D(\boldsymbol{h}_2, \boldsymbol{l}_2, \boldsymbol{c}_2, \boldsymbol{h}_1, \boldsymbol{l}_1, \boldsymbol{c}_1), & (2.12)\\ D(\boldsymbol{h}_1, \boldsymbol{l}_1, \boldsymbol{c}_1, \boldsymbol{h}_2, \boldsymbol{l}_2, \boldsymbol{c}_2) &= D(\boldsymbol{h}_1 - \boldsymbol{c}_1, \boldsymbol{l}_1 - \boldsymbol{c}_1, -\boldsymbol{c}_1, \boldsymbol{h}_2 - \boldsymbol{c}_2, \boldsymbol{l}_2 - \boldsymbol{c}_2, -\boldsymbol{c}_2), & (2.13)\\ D(\boldsymbol{h}_1, \boldsymbol{l}_1, \boldsymbol{c}_1, \boldsymbol{h}_2, \boldsymbol{l}_2, \boldsymbol{c}_2) &= -D(\boldsymbol{h}_1, \boldsymbol{l}_1, \boldsymbol{c}_1, -\boldsymbol{l}_2, -\boldsymbol{h}_2, -\boldsymbol{c}_2). & (2.14) \end{aligned}$$

Now we present two important lemmas that link these invariance properties to the estimator efficiency.

Lemma 2.3.7. *Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a measure-preserving transformation such that, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{\kappa}$ in the support of $f_{\boldsymbol{\theta}}$,*

$$f(T(\boldsymbol{\kappa})) = f(\boldsymbol{\kappa}).$$

Let $D(\boldsymbol{\kappa})$ be any estimator and $L(\cdot, \boldsymbol{\theta})$ be an arbitrary loss function such that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $L(\cdot, \boldsymbol{\theta})$ is a convex function. Define $T^k = T \circ T^{k-1}$ and T^0 be the identity operator. Let A'_K be an average operator defined as

$$A_K(D(\boldsymbol{\kappa})) = \frac{1}{K} \sum_{k=1}^{K} D(T^{k-1}(\boldsymbol{\kappa})).$$

Then for any convex loss function $L(\cdot)$, we have

$$\mathbb{E}(L(A_K(D(\boldsymbol{\kappa})), \boldsymbol{\theta})) \leq \mathbb{E}(L(D(\boldsymbol{\kappa}), \boldsymbol{\theta})) \quad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Proof. See Garman and Klass (1980). □

Remark 2.3.8. Now let us supose that there exist M, such that

$$A_M(D(T(\boldsymbol{\kappa}))) = A_M(D(\boldsymbol{\kappa})). \quad (2.15)$$

Then, for any $K > 0$, we have

$$\begin{aligned} A_M(A_K(D(\boldsymbol{\kappa}))) &= \frac{1}{M} \sum_{m=1}^{M} \left[\frac{1}{K} \sum_{k=1}^{K} D\big(T^{k+m-2}(\boldsymbol{\kappa})\big) \right] \\ &= \frac{1}{K} \sum_{k=1}^{K} \left[\frac{1}{M} \sum_{m=1}^{M} D\big(T^{k-1}(T^{m-1}(\boldsymbol{\kappa}))\big) \right] \\ &= \frac{1}{K} \sum_{k=1}^{K} A_M(D(T^{k-1}(\boldsymbol{\kappa}))) \\ &= A_M(D(\boldsymbol{\kappa})). \quad (2.16) \end{aligned}$$

The result suggests the best estimator among the class of averaging estimators generated by $D(\boldsymbol{\kappa})$ can be characterized by (2.D.6). Note that if $T^M = I,\ M > 0$, then

$$\begin{aligned} A_M(D(T(\boldsymbol{\kappa}))) &= \frac{1}{M}\sum_{k=1}^{M} D(T^k(\boldsymbol{\kappa})) \\ &= \frac{1}{M}\left[\sum_{k=2}^{M} D(T^{k-1}(\boldsymbol{\kappa})) + D(T^M(\boldsymbol{\kappa}))\right] \\ &= \frac{1}{M}\left[\sum_{k=1}^{M} D(T^{k-1}(\boldsymbol{\kappa}))\right] \\ &= A_M(D(\boldsymbol{\kappa})), \end{aligned} \tag{2.17}$$

that is, if M is the period of transformation, then $A_M(D(\cdot))$ minimize the expected risk amount the class $\{A_K(D(\cdot))\}$.

Lemma 2.3.9. *Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a measure-preserving transformation such that, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{\kappa}$ in the support of $f(\cdot;\boldsymbol{\theta})$,*

$$f(T(\boldsymbol{\kappa});\boldsymbol{\theta}) = f(\boldsymbol{\kappa};-\boldsymbol{\theta}).$$

Let $D(\boldsymbol{\kappa})$ be any estimator and $L(\cdot,\boldsymbol{\theta})$ be an arbitrary loss function such that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $L(\cdot,\boldsymbol{\theta})$ is a convex function. Define $T^k = T \circ T^{k-1}$ and T^0 be the identity operator. Let A'_K be an alternating average operator defined as

$$A'_K(D(\boldsymbol{\kappa})) = \frac{1}{K}\sum_{k=1}^{K}(-1)^{k-1} D(T^{k-1}(\boldsymbol{\kappa})).$$

Let $\boldsymbol{\theta}$ be a random variable with a symetric densitity function $g_{\boldsymbol{\theta}}(\cdot)$. Then for any convex loss function, we have

$$B_{g_{\boldsymbol{\theta}}}(A'_k(D(\boldsymbol{\kappa}))) \leq B_{g_{\boldsymbol{\theta}}}(D(\boldsymbol{\kappa}))$$

where

$$B_{g_{\boldsymbol{\theta}}}(D(\boldsymbol{\kappa})) = \int_{\boldsymbol{\Theta}} \mathbb{E}(L(D(\boldsymbol{\kappa}),\boldsymbol{\theta}))\ g_{\boldsymbol{\theta}}(\boldsymbol{\theta})\ d\boldsymbol{\theta}$$

is the expected risk.

Proof. See appendix 2.D. □

Remark 2.3.10. Similar to the case of the invariant transform, if $T^M = I$ and M is an even number, we have

$$\begin{aligned}
A'_M(D(T(\boldsymbol{\kappa}))) &= \frac{1}{M}\sum_{m=1}^{M}(-1)^{m-1}D(T^m(\boldsymbol{\kappa})) \\
&= -\frac{1}{M}\Big[\sum_{m=2}^{M}(-1)^{m-1}D(T^{k-1}(\boldsymbol{\kappa})) + (-1)^M D(T^M(\boldsymbol{\kappa}))\Big] \\
&= -\frac{1}{M}\sum_{k=1}^{M}D(T^{k-1}(\boldsymbol{\kappa})) \\
&= -A'_M(D(\boldsymbol{\kappa})), \qquad (2.18) \\
A'_M(A_k(D(\boldsymbol{\kappa}))) &= \frac{1}{M}\sum_{m=1}^{M}(-1)^{m-1}\frac{1}{K}\Big[\sum_{k=1}^{K}(-1)^{k-1}D(T^{m+k-2}(\boldsymbol{\kappa}))\Big] \\
&= \frac{1}{K}\sum_{k=1}^{K}(-1)^{k-1}\frac{1}{M}\Big[\sum_{m=1}^{M}(-1)^{m-1}D(T^{m-1}(T^{k-1}(\boldsymbol{\kappa})))\Big] \\
&= \frac{1}{K}\sum_{k=1}^{K}(-1)^{k-1}A_M(D(T^{M-1}(\boldsymbol{\kappa}))) \\
&= \frac{1}{K}\sum_{k=1}^{K}(-1)^{2k-2}A_M(D(\boldsymbol{\kappa})) \qquad (2.19) \\
&= A'_M(D(\boldsymbol{\kappa})). \qquad (2.20)
\end{aligned}$$

That is the best estimator among the class of averaging estimators generated by $D(\boldsymbol{\kappa})$ will be $A_M(D(\boldsymbol{\kappa}))$. What would happen if M is an odd number? We have

$$\begin{aligned}
A'_{2M}(D(T(\boldsymbol{\kappa}))) &= \frac{1}{2M}\sum_{m=1}^{2M}(-1)^{m-1}D(T^{m-1}(\boldsymbol{\kappa})) \\
&= \frac{1}{2M}\sum_{m=1}^{M}(-1)^{m-1}D(T^{m-1}(\boldsymbol{\kappa})) + \frac{1}{2M}\sum_{m=1}^{M}(-1)^{m-1+M}D(T^{m-1}(\boldsymbol{\kappa})) \\
&= \frac{1}{2M}\sum_{m=1}^{M}(-1)^{m-1}D(T^{m-1}(\boldsymbol{\kappa})) - \frac{1}{2M}\sum_{m=1}^{M}(-1)^{m-1}D(T^{m-1}(\boldsymbol{\kappa})) \\
&= 0. \qquad (2.21)
\end{aligned}$$

That is all estimators will have higher Bayes risk compared to trivial decisions $D(\boldsymbol{\kappa}) = 0$. By (2.D.7), we have

$$f(\boldsymbol{\kappa};\boldsymbol{\theta}) = f(T^M(\boldsymbol{\kappa});\boldsymbol{\theta}) = f(\boldsymbol{\kappa};(-1)^M\boldsymbol{\theta}) = f(\boldsymbol{\kappa};-\boldsymbol{\theta}), \qquad (2.22)$$

that is, the posterior distribution will always be symmetric. We can hardly get an estimator for $\boldsymbol{\theta}$. From here we can also see why an estimator D consisting of $\boldsymbol{h}_1 - \boldsymbol{l}_1, \boldsymbol{h}_2 - \boldsymbol{l}_2$ only will fail, as $D(T_2(\boldsymbol{\kappa})) = D(\boldsymbol{\kappa})$.

2.4. Proposed correlation measures using (O, H, L, C) pairs

2.4.1. Eight correlations and their properties. Constructing an estimator that satisfies all the conditions outlined in (2.3), (2.12), (2.13), and (2.14) presents a significant challenge. Our strategy is to identify a simple estimator $D_1(\boldsymbol{\kappa})$ satisfying (2.3) that can effectively approximate ρ. Then we identify all unique $D_k(\boldsymbol{\kappa}) = D_1(T_k(\boldsymbol{\kappa}))$ where T_k is an probability invariant tranformation generated by T_1, T_2, T_3. Note that while T_3 is not a probability invariant transformation $T_3^2 = I$ and $T_3T_1T_3$ is a non-identity probability invariance transformation. Then we consider a simple average estimator $\tilde{D}(\boldsymbol{\kappa}) = \frac{1}{K}\sum_{k=1}^{K} D_k(\boldsymbol{\kappa})$ by setting the weight of all K unique estimator $D_k(\boldsymbol{\kappa})$ to be equal so that (2.D.4) in Appendix 2.D holds under all probability invariant transformations, which suggests that it is the best estimator generated by the class of probability invariant transformations. Finally, we set

$$D(\boldsymbol{\kappa}) = \frac{1}{2}\tilde{D}(\boldsymbol{\kappa}) - \frac{1}{2}\tilde{D}(T_2(\boldsymbol{\kappa})). \tag{2.23}$$

By lemma 2.D.3 in Appendix 2.D, $D(\boldsymbol{\kappa})$ is a more efficient estimator compared to $\tilde{D}$ in terms of Bayes risk under a symmetric prior. Note that only defines most efficient estimator in forms of linear combinations of $D_k(\boldsymbol{\kappa})$. We can still construct other forms of combinations which satisfy (2.12), (2.13) and (2.14), i.e. harmonic mean.

We first consider eight correlation measures $S_t = \{s_{k,t},\ k = 1, \ldots, 8\}$ on day t

Definition 2.4.1.

$$\begin{aligned}
s_{1,t} &= \operatorname{cor}(\boldsymbol{h}_1, \boldsymbol{h}_2), & s_{2,t} &= \operatorname{cor}(\boldsymbol{l}_1, \boldsymbol{l}_2),\\
s_{3,t} &= \operatorname{cor}(\boldsymbol{h}_1 - \boldsymbol{c}_1, \boldsymbol{h}_2 - \boldsymbol{c}_2), & s_{4,t} &= \operatorname{cor}(\boldsymbol{l}_1 - \boldsymbol{c}_1, \boldsymbol{l}_2 - \boldsymbol{c}_2),\\
s_{5,t} &= \operatorname{cor}(\boldsymbol{h}_1, \boldsymbol{l}_2), & s_{6,t} &= \operatorname{cor}(\boldsymbol{l}_1, \boldsymbol{h}_2),\\
s_{7,t} &= \operatorname{cor}(\boldsymbol{h}_1 - \boldsymbol{c}_1, \boldsymbol{l}_2 - \boldsymbol{c}_2), & s_{8,t} &= \operatorname{cor}(\boldsymbol{l}_1 - \boldsymbol{c}_1, \boldsymbol{h}_2 - \boldsymbol{c}_2)
\end{aligned} \tag{2.24}$$

where $\operatorname{cor}(\cdot,\cdot)$ *is the Peason correlation defined in* (1.24).

Remark 2.4.2. We rewrite (2.24) as

$$\begin{aligned}
s_{1,t} &= \operatorname{cor}(\log H_{1,t,1:n}/O_{1,t,1:n}, \log H_{2,t,1:n}/O_{2,t,1:n}),\ s_{2,t} = \operatorname{cor}(\log L_{1,t,1:n}/O_{1,t,1:n}, \log L_{2,t,1:n}/O_{2,t,1:n}),\\
s_{3,t} &= \operatorname{cor}(\log H_{1,t,1:n}/C_{1,t,1:n}, \log H_{2,t,1:n}/C_{2,t,1:n}),\ s_{4,t} = \operatorname{cor}(\log L_{1,t,1:n}/C_{1,t,1:n}, \log L_{2,t,1:n}/C_{2,t,1:n}),\\
s_{5,t} &= \operatorname{cor}(\log H_{1,t,1:n}/O_{1,t,1:n}, \log L_{2,t,1:n}/O_{2,t,1:n}),\ s_{6,t} = \operatorname{cor}(\log L_{1,t,1:n}/O_{1,t,1:n}, \log H_{2,t,1:n}/O_{2,t,1:n}),\\
s_{7,t} &= \operatorname{cor}(\log H_{1,t,1:n}/C_{1,t,1:n}, \log L_{2,t,1:n}/C_{2,t,1:n}),\ s_{8,t} = \operatorname{cor}(\log L_{1,t,1:n}/C_{1,t,1:n}, \log H_{2,t,1:n}/C_{2,t,1:n}).
\end{aligned} \tag{2.25}$$

The relation between $s_{i,t}$ is summarised in the following lemma.

Lemma 2.4.3. *The following equations hold:*

$$s_{2,t} = s_{1,t} \circ (T_3T_1T_3), \quad s_{3,t} = s_{1,t} \circ T_2, \quad s_{4,t} = s_{1,t} \circ (T_3T_1T_3T_2), \tag{2.26}$$

$$s_{k+4,t} = -s_{k,t} \circ T_3, \qquad k = 1, 2, 3, 4 \tag{2.27}$$

Remark 2.4.4. The distribution of $s_{1,t}$ is depending on the parameter ρ. T_2 changes the reference from $O_{i,t}$ to $C_{i,t}$ in the demoninator of $h_{i,t}, l_{i,t}$ and T_3 changes $h_{i,t}$ to $l_{i,t}$ and vice versa according (2.12) to (2.12).

Corollary 2.4.5. *Let $g_k(x,\rho)$ be the density function of $s_{k,t}$ when the true correlation is ρ, then by the invariance property we defined previously we have*

$$\begin{aligned} g_k(x;\rho) &= g_1(x;\rho) \qquad k=2,3,4 \\ g_k(x;\rho) &= g_1(-x;-\rho) \quad k=5,6,7,8. \end{aligned}$$

That is, given ρ, all $s_{1,t}, s_{2,t}, s_{3,t}, s_{4,t}$ have the same distribution, and also the same distribution as $-s_{1,t}, -s_{2,t}, -s_{3,t}, -s_{4,t}$ given $-\rho$.

Remark 2.4.6. Intuitively,

$$\begin{aligned} &\text{if} \quad \rho=1, \quad \boldsymbol{h}_1 = \mathfrak{c}\boldsymbol{h}_2 \Rightarrow \operatorname{cor}(\boldsymbol{h}_1,\boldsymbol{h}_2)=1; \\ &\text{if} \quad \rho=0, \quad \boldsymbol{h}_1,\boldsymbol{h}_2 \text{ are independent} \Rightarrow \mathbb{E}(\operatorname{cor}(\boldsymbol{h}_1,\boldsymbol{h}_2))=0; \\ &\text{if} \quad \rho=-1, \quad \boldsymbol{l}_1 = -\mathfrak{c}\boldsymbol{h}_2 \Rightarrow \operatorname{cor}(\boldsymbol{h}_1,\boldsymbol{h}_2) = -\operatorname{cor}(\boldsymbol{h}_1,\boldsymbol{l}_1) < 0 \end{aligned} \tag{2.28}$$

for some constant $\mathfrak{c}$.

2.4.2. Simulation experiments. To assess the properties of the eight proposed correlations in (2.24), simulation studies are conducted in the following steps. We assume that a price is observed whenever a trade occurs. We approximate the high and low of a Brownian motion over the interval $(\tau,\tau+1)$ using the maximum and minimum prices observed from a discretized price process within that interval. A common practical approach is to simulate prices over evenly spaced subintervals. However, such an approach may not accurately reflect real market conditions, as trading times at the microstructural level are inherently random rather than uniformly distributed. Moreover, in the multivariate setting, trades across different assets do not occur simultaneously, leading to temporal misalignment in observed prices. To more realistically capture trading dynamics, we model trading activity using independent Poisson processes for each asset. While this assumption of independence may not hold globally, since market-wide events, such as policy announcements by regulators, can cause a simultaneous increase in trading volume across all assets, we assume local independence. That is, within a short time window of high trading activity, each individual trade execution is assumed to originate from independent traders. Let $M_{t,\tau}$ denote the total number of trades. We further assume that there is always trading by the end of the day. In other words, no opening price jump is assumed to occur.

Simulation scheme

Step 1 The mean of daily return for both assets 1 and 2 are set to be $\mu_1 = 0$ and $\mu_2 = 0$. For the variance, they are fixed at

$$\text{Homoskedasticity:} \quad \sigma_1^2 = 0.00001 \ \text{ and } \ \sigma_2^2 = 0.00002. \tag{2.29}$$

The true correlation ρ ranges from -0.95 to 0.95 with steps of 0.025 in H levels. Write $\rho_h = -0.975 + 0.025h, h = 1,\ldots,76$.

Step 2 For each true ρ, $N = 10000$ days are simulated to get the distribution. In each day t, $n = 60$ intra-day observations are sampled. The total number of tradings at time τ in the day t, excluding the close trading, $M_{t,\tau} - 1$, are i.i.d sampled from a Poisson distribution with $\lambda = 10000$. This is used to simulate a price process at time τ of day t to obtain $\{O_{i,t,\tau}, H_{i,t,\tau}, L_{i,t,\tau}, C_{i,t,\tau}\}, i = 1,2$.

Step 3 Given $M_{t,\tau}$ tradings of the 2 assets in total at time τ of day t, the intermediate trading times $\mathfrak{t}_{t,\tau,1}, ..., \mathfrak{t}_{t,\tau,M_{t,\tau}-1}$ follow the same distribution as order statistics of $M_{t,\tau}-1$ uniform distributed random variables on the unit interval. Hence the time between tradings $\Delta\mathfrak{t}_{t,\tau,\iota} = \mathfrak{t}_{t,\tau,\iota} - \mathfrak{t}_{t,\tau,\iota-1}$ can be sampled in the following process. Let $E_{t,\tau,1}, ..., E_{t,\tau,M_{t,\tau}}$ be $M_{t,\tau}$

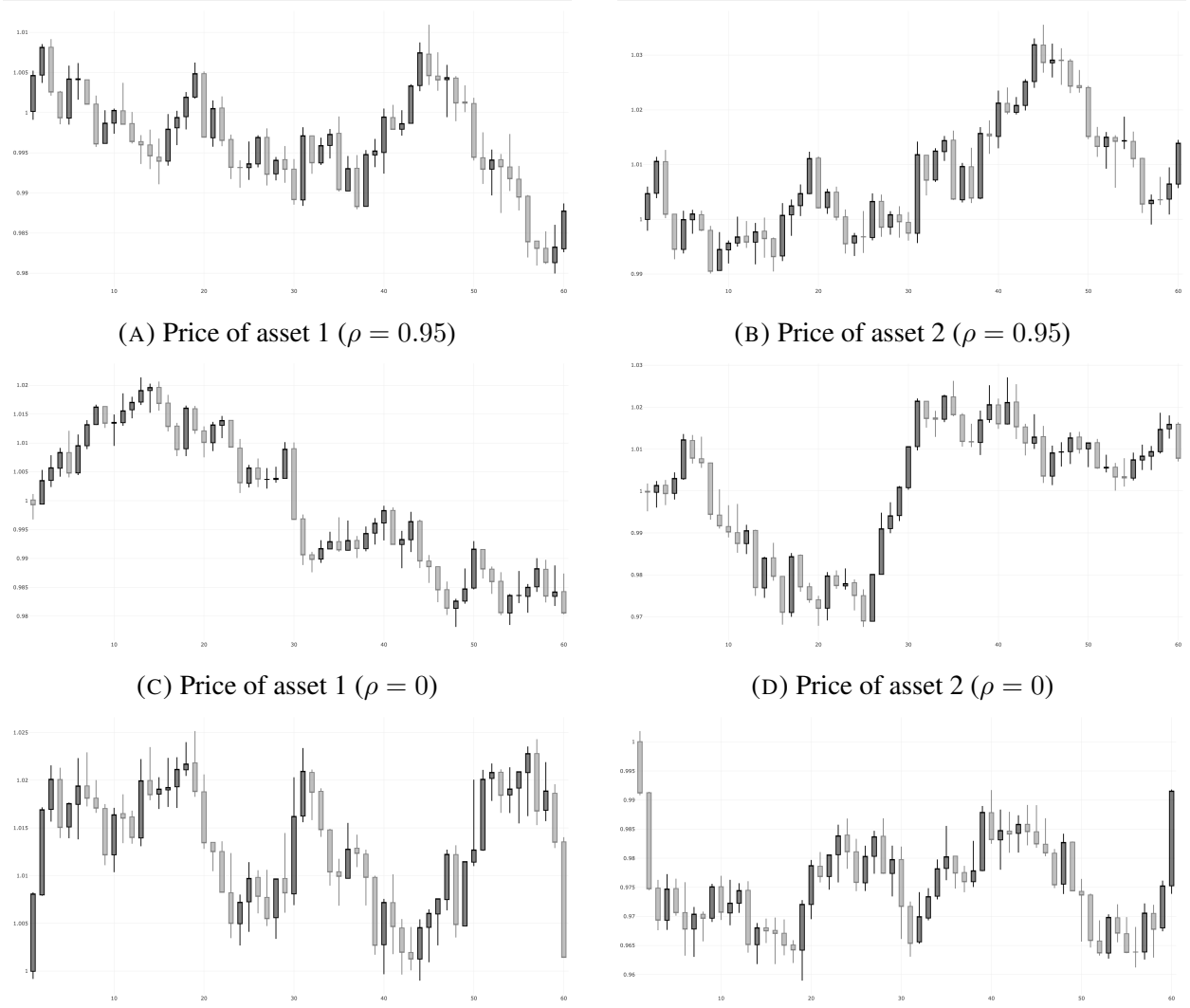
samples drawn independently from $\mathrm{Exp}(1)$. Then set $\Delta \mathrm{t}_{t,\tau,\iota} = E_{t,\tau,\iota}/(\sum_{\iota=1}^{M_{t,\tau}} E_{t,\tau,\iota})$. The rate of the exponential distribution can be chosen arbitrarily. These lengths of trading intervals $\Delta \mathrm{t}_{t,\tau,\iota}$ are used to calculate the price of assets 1 and 2 at time τ of day t.

Step 4 Returns $\boldsymbol{r}_{t,\tau,\iota} = (r_{1,t,\tau,\iota}, r_{2,t,\tau,\iota})$ are simulated from $\mathrm{N}_2(\boldsymbol{\mu}\Delta \mathrm{t}_{t,\tau,\iota}, \Delta \mathrm{t}^2_{t,\tau,\iota}\boldsymbol{\Sigma})$ where $\boldsymbol{\mu} = (\mu_1, \mu_2)$, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_h\sigma_1\sigma_2 \\ \rho_h\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

Step 5 The raw price series $P'_{i,t,\tau,\iota}, i = 1, 2$ are calculated from $r_{i,t,\tau,1:M_{t,\tau}}$ using $P'_{i,t,\tau,\iota} = P'_{i,t,\tau,\iota-1}e^{r_{i,t,\tau,\iota}}$. Then $M_{t,\tau} - 1$ samples $\mathcal{I}_{t,\tau,1:(M_{t,\tau}-1)}$ are simulated from Bernoulli distribution with $p = 0.6$, where $\mathcal{I}_{t,\tau,\iota} = 1$ and $\mathcal{I}_{t,\tau,\iota} = 0$ represent the trading belongs to asset 1 and 2 respectively. Then the actual price sequence is obtained by applying masks over raw price sequence: $\{P_{1,t,\tau,\iota}\} = \{P'_{1,t,\tau,\iota} | \mathcal{I}_{t,\tau,\iota} = 1 \text{ or } \iota = M_{t,\tau}\}, \{P_{2,t,\tau,\iota}\} = \{P'_{2,t,\tau,\iota} | \mathcal{I}_{t,\tau,\iota} = 0 \text{ or } \iota = M_{t,\tau}\}$.

Step 6 Lastly, summary statistics of asset i $\{O_{i,t,\tau}, H_{i,t,\tau}, L_{i,t,\tau}, C_{i,t,\tau}\}$ are calculated from the price set $\{P_{i,t,\tau,\iota}\}$.

The aim of the remaining section is to derive some correlation measures $D(\boldsymbol{\kappa})$ in (2.4.1), which will be used in Chapter 3 for model development and Chapter 4 for applications. The appendix 2.E provides supplementary material for the simulation scheme. Figure 1 visualises the simulated values of $\{O_{1,t,\tau}, H_{1,t,\tau}, L_{1,t,\tau}, C_{1,t,\tau}\}$ and illustrates how they are affected by different correlation assumptions $\rho = 0, 0.95, -0.95$ for $\tau = 1, \ldots, 60$ time periods using candle-stick plot. Each candlestick represents a specific time period and displays the opening, closing, highest, and lowest prices. The body of the candle shows the difference between the open and close, while the wicks (or shadows) show the highs and lows. A bullish dark grey candle indicates a price increase so that from the top, it reports the values $\{H_{i,t,\tau}, C_{i,t,\tau}, O_{i,t,\tau}, L_{i,t,\tau}\}$ whereas a bearish light grey candle indicates a price decrease showing instead, the values $\{H_{i,t,\tau}, O_{i,t,\tau}, C_{i,t,\tau}, L_{i,t,\tau}\}$.



(A) Price of asset 1 ($\rho = 0.95$)

(B) Price of asset 2 ($\rho = 0.95$)

(C) Price of asset 1 ($\rho = 0$)

(D) Price of asset 2 ($\rho = 0$)

(E) Price of asset 1 ($\rho = -0.95$)

(F) Price of asset 2 ($\rho = -0.95$)

FIGURE 1. Simulated $\{O_{i,t,\tau}, H_{i,t,\tau}, L_{i,t,\tau}, C_{i,t,\tau}\}$ from different correlation assumptions $\rho = 0.95, 0, -0.95$ at time $\tau = 1, \ldots, 60$ for a certain day t.

Figure 2 illustrates the distributions of $s_{1,t}|\rho$ to $s_{8,t}|\rho$ for day $t = 1, \ldots, N,\ (N = 10000)$. The y-axis is the observed correlation $s_{k,t}$, and the x-axis is the true correlation ρ. The black solid (grey dash) lines represent the mean $\bar{s}_{k,t}$ and percentiles of $s_{k,t}$ over N days. Results show that the distributions of $s_{k,t}, k = 1, \ldots, 4$ are the same and the distributions of $s_{k,t}, k = 5, \ldots, 8$ are also the same confirming the probability invariant property of the transformations T_2 and $T_3T_1T_3$ in (2.26) and the probability semi-invariant property of the transformation T_3 in (2.27).

We can further observe that sample means of $s_{k,t}$ over t at the true correlation ρ_h, $\bar{s}_{k,h},\ k = 1, \ldots, 4$ describe positive correlation and share a common relation to the true correlation ρ_h, while $\bar{s}_{k,h},\ k = 5, \ldots, 8$ describing negative correlation. The result is consistent with (2.28). One also observe that $s_{k,t}, k = 1, \ldots, 4\ (5, \ldots, 8)$ provide good correlation estimates when $\rho \to 1(-1)$. However, the bias tends to increase as $\rho \to -1(1)$.

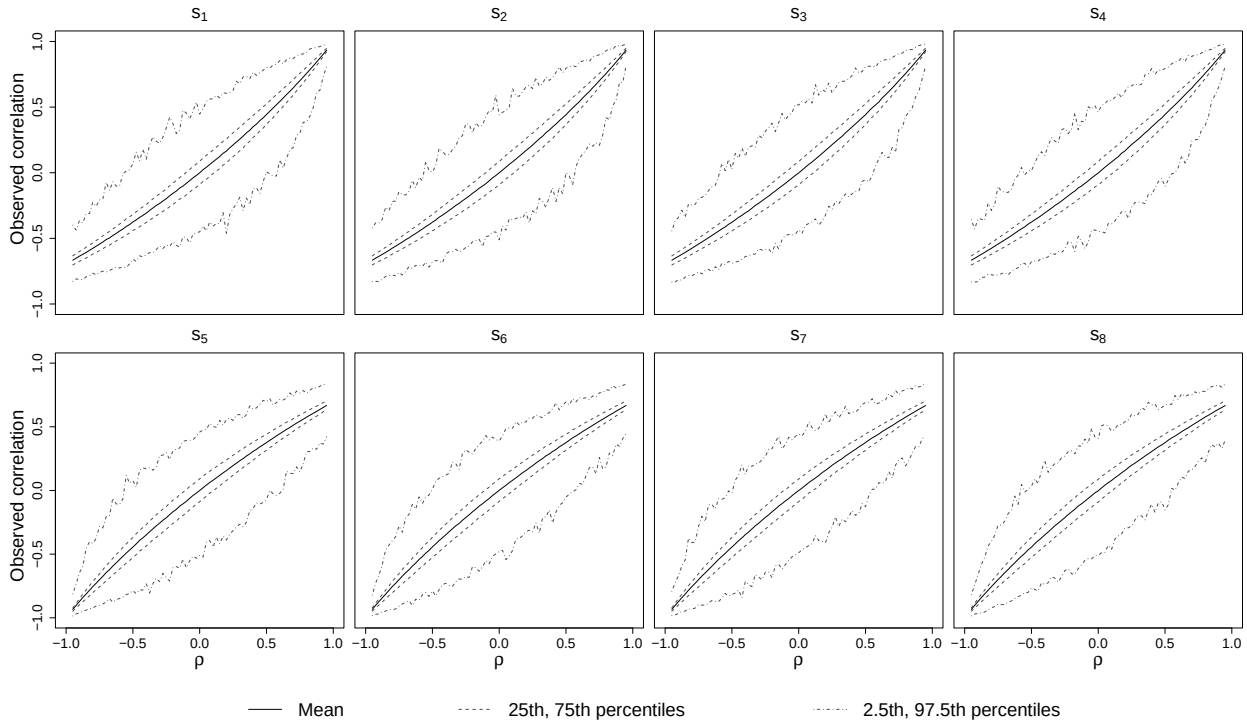


FIGURE 2. Distributions of $s_{1:8,t}$ across true values ρ from the simulation experiment.

Now we try to propose function f such that $f(s_{1,t})$ can be an adequate estimator of ρ. Let $o_1(\rho) = \mathrm{E}(s_{1,t})$, $o_2(\rho) = \underset{x}{\operatorname{argmax}}\, g_{\theta,1}(x)$. By moment matching, we propose our first mean estimator

$$f_\mu(x) = \begin{cases} o_1^{-1}(x) & x > \min o_1(\rho), \\ -1 & x \leq \min o_1(\rho). \end{cases} \tag{2.30}$$

The distributions of $s_{k,t}$ against ρ in Figure 2 show asymmetry when $\rho \to 1, -1$. Hence, we propose a mode estimator

$$f_m(x) = \begin{cases} o_2^{-1}(x) & x > \min o_2(\rho), \\ -1 & x \leq \min o_2(\rho), \end{cases} \tag{2.31}$$

which is equivalent to a pointwise ML estimator.

2.4.3. Estimation of mean, mode and variance functions.

2.4.3.1. *Fourth order polynomial fit of mean functions.* The theoretical derivation of o_1 is quite difficult. We aim to obtain an estimate of f_μ through $\bar{s}_{1,h}, h = 1, ..., 76$. Note that we can also get an estimate from $\bar{s}_{k,h}, h = 1, ..., 76$ for $k = 2, ..., 8$ as well. Note that $\bar{s}_{k_1,h}$ and $\bar{s}_{k_2,h}$ are not independent for $k_1 \neq k_2$. Thus, we only use one subset instead of getting an estimate by merging all 8 sets of data.

We first fit a polynomial regression model with order 10. The summary (see Appendix 2.F) of the fitting suggests that coefficients of order greater than 5 are not significant. We choose the order of polynomials to be 4. Furthermore, we have $f_\mu(1) = 1$ as suggested in (2.28). When assets 1 and 2 are independent, $\mathrm{E}(s_{1,t}) = 0$ by independence. Hence $f_\mu(0) = 0$.

We fitted a fourth-order polynomial

$$f_\mu(x) = a + bx + cx^2 + dx^3 + ex^4 \tag{2.32}$$

to the mean functions $\bar{s}_{1,h}$ by minimising the squared error, that is, $f_\mu(x)$ was obtained by

$$\begin{aligned} \min \quad & \sum_{h=1}^{76} \big(\rho_h - f_\mu(\bar{s}_{1,h})\big)^2, \\ \text{subject to} \quad & f_\mu(1) = 1,\ f_\mu(0) = 0. \end{aligned} \tag{2.33}$$

This is a typical OLS problem with linear constraints. Let

$$\boldsymbol{X} = \begin{pmatrix} 1 & \bar{s}_{1,1} & \bar{s}_{1,1}^2 & \bar{s}_{1,1}^3 & \bar{s}_{1,1}^4 \\ 1 & \bar{s}_{1,2} & \bar{s}_{1,2}^2 & \bar{s}_{1,2}^3 & \bar{s}_{1,2}^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix},\ \boldsymbol{Y} = \begin{pmatrix} \rho_1 & \rho_2 & \cdots \end{pmatrix}^\top,\ \boldsymbol{\beta} = \begin{pmatrix} a & b & c & d & e \end{pmatrix}^\top,$$

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix},\ \boldsymbol{C} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The minimisation problem in (2.42) is equivalent to

$$\begin{aligned} \min \quad & (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{C}. \end{aligned} \tag{2.34}$$

Let $\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$. Then the solution of (2.34) is given by

$$\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{A}^\top \big(\boldsymbol{A}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{A}^\top\big)^{-1} (\boldsymbol{C} - \boldsymbol{A}\tilde{\boldsymbol{\beta}}). \tag{2.35}$$

The estimated $f_\mu(x)$ is given by

$$f_\mu(x) = 1.2268x - 0.2410x^2 + 0.0543x^3 - 0.0401x^4. \tag{2.36}$$

The upper plot in Figure 4(a) displays the observed mean correlation $\bar{s}_{1,t}$ (black line) and the fitted correlation $\hat{\rho} = f_\mu(\bar{s}_{1,t})$ (grey line) across $s_{1,t}$ whereas the lower plot graphs the residuals $\big(\rho_h - f_\mu(\bar{s}_{1,h})\big)$. It can be observed that the fourth-order polynomial effectively captures the underlying trend in the data. The corresponding residual plot exhibits no discernible pattern, indicating that the fourth-order specification provides an adequate fit for the mean.

2.4.3.2. *Fourth order polynomial fit of the parametric mode* $\tilde{s}_k$. Previously proposed correlation measures are based on estimating the relationship between the mean of the distribution of $s_{k,t}|\rho$ and ρ. However, as illustrated in Figure 2, the distribution of $s_{1,t}|\rho$ becomes increasingly asymmetric as $\rho \to 1$. This asymmetry adversely affects the accuracy of $\hat{\rho}$ when using the estimation approach outlined in (2.36). Adopting Beta distribution to $s_{k,t}$ allows estimation of the mode $\tilde{s}_k$ of the distribution, which is less sensitive to skewed distribution. Hence, this mode estimator is preferred.

To adjust the range of $s_{1,t}$ to be within (0,1), Beta distribution is assigned to $(s_{1,t}+1)/2,\ t = 1, \ldots, N$. Using the moment matching method, estimates of Beta distribution parameters α and β can be obtained by solving the mean and variance equations

$$\frac{\mu_s + 1}{2} = \frac{\alpha}{\alpha + \beta} \quad \text{and} \quad \frac{\sigma_s^2}{4} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)},$$

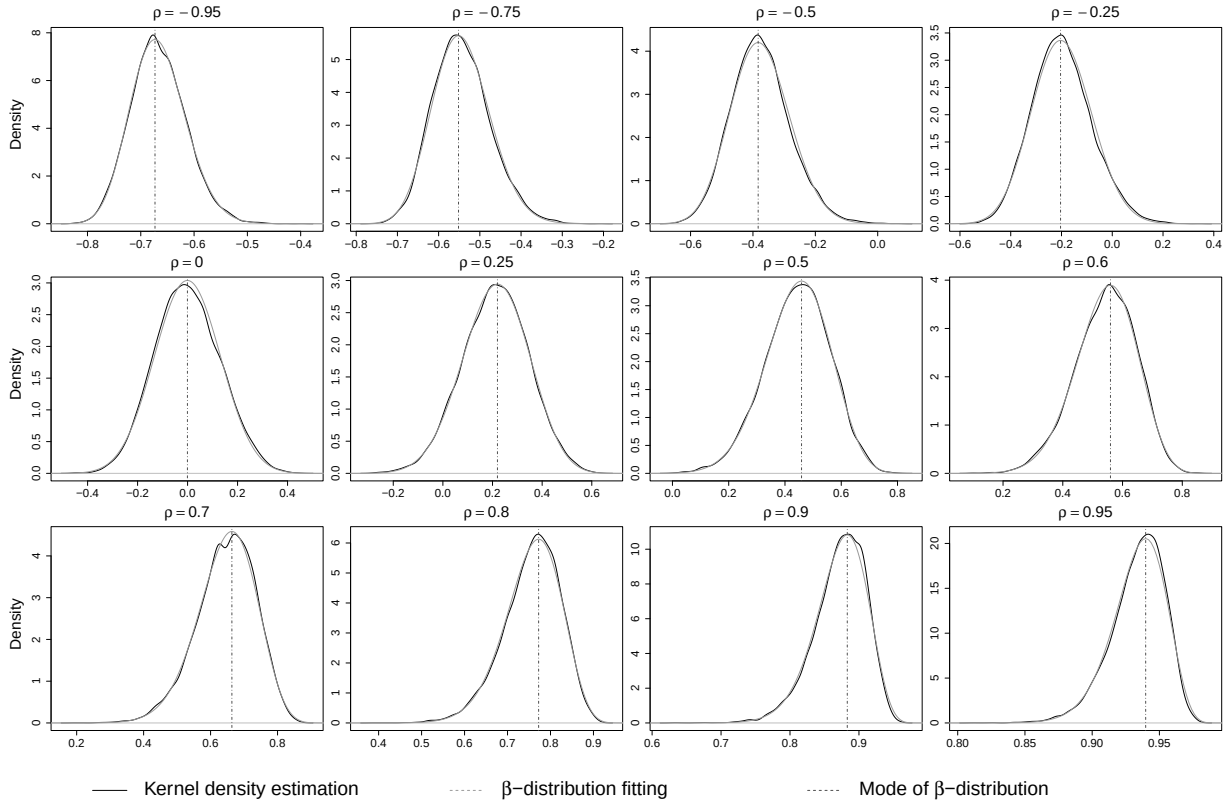


FIGURE 3. Distribution of s_1 across true ρ levels with kernel smoothed density and fitted density using Beta distribution. Dash lines indicate the mean of respective densities.

respectively to obtain

$$\alpha = \frac{\mu_s + 1}{2}\left[\frac{(\mu_s+1)(1-\mu_s)}{\sigma_s^2} - 1\right], \quad \beta = \frac{1-\mu_s}{2}\left[\frac{(\mu_s+1)(1-\mu_s)}{\sigma_s^2} - 1\right], \tag{2.37}$$

where the mean μ_s and variance σ_s^2 of $s_{1,t}$ can be estimated from the sample mean and sample variance respectively. Figure 3 graphs the kernel density estimation of $s_{1,t}$ and the fitted densities using Beta distribution, showing good approximation. The mode is given by

$$\tilde{s}_1 = \frac{2(\alpha-1)}{\alpha+\beta-2} - 1. \tag{2.38}$$

The trend of mode $\tilde{s}_1$ across ρ is graphed Figure 4(c). This curve is again approximated using a fourth-order polynomial given by

$$f_m(x) = 1.1859x - 0.2495x^2 + 0.1007x^3 - 0.0371x^4. \tag{2.39}$$

The coefficients in (2.39) are estimated in a similar way to (2.36). The upper plot of Figure 4(c) displays the estimated function using the mode estimator $\rho = f_m(\tilde{s}_{1,t})$ against mode $\tilde{s}_{1,t}$ showing good agreement with the observed $\tilde{s}_1$. The residuals $\big(\tilde{s}_1 - f_m(\tilde{s}_{1,t})\big)$ are plotted in the lower plot of Figure 4(c).

2.4.3.3. *Fourth order polynomial fit of RMSE function.* The $f_\mu(s_{1,t})$ and $f_m(s_{1,t})$ defines an point estimator of the correlation ρ. We are also interested in the RMSE of the estimator, which is given by

$$\mathrm{rmse}_\mu(\rho_h) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\big(f_\mu(s_{1,t}) - \rho_h\big)^2}, \;\; h = 1, \ldots, 76 \tag{2.40}$$

and

$$\mathrm{rmse}_m(\rho_h) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\big(f_m(s_{1,t}) - \rho_h\big)^2}, \;\; h = 1, \ldots, 76 \tag{2.41}$$

for different ρ values. We choose a fourth-order polynomial fit to the $\mathrm{rmse}(\cdot)$ function. Note that if $\rho = 1$, then $s_{1j} = 1$ almost surely. Hence, $\mathrm{rmse}_\mu(1) = \mathrm{rmse}_{m1}(1) = 0$. The fitted curve $g_{\mu_1}(\cdot)$ is obtained by

$$\begin{aligned} \min \quad & \sum_{h=1}^{76}\left[\sqrt{\frac{1}{N}\sum_{t=1}^{N}\big(f_\mu(s_{1,t}) - \rho_h\big)^2} - g_\mu(\rho_h)\right]^2, \\ \text{subject to} \quad & g_\mu(1) = 0. \end{aligned} \tag{2.42}$$

The fitted curve $g_\mu(\rho)$ mapping ρ to $\mathrm{rmse}_\mu(\rho)$ is given by

$$g_\mu(\rho) = 0.1596 - 0.0106\rho - 0.1222\rho^2 - 0.0221\rho^3 - 0.0047\rho^4. \tag{2.43}$$

Similarly, the fitted curve $g_m(\rho)$ mapping ρ to $\mathrm{rmse}_m(\rho)$ is given by

$$g_m(\rho) = 0.1542 - 0.0088\rho - 0.1002\rho^2 - 0.0252\rho^3 - 0.0200\rho^4. \tag{2.44}$$

When ρ is estimated by $f_\mu(\bar{s}_{1,t})$ in (2.36), the RMSE for $s_{1,t}$ is given by $g_\mu\big(f_\mu(\bar{s}_{1,t})\big)$. Similarly, when ρ is estimated by $f_m(\tilde{s}_{1,t})$ in (2.36), the RMSE for $s_{1,t}$ is given by $g_m\big(f_m(\tilde{s}_{1,t})\big)$. The fitted curve $g_\mu(\cdot)$ as graphed in the upper plot of Figure 4(b) and $g_m(\cdot)$ in the upper plot of Figure 4(d) show good estimation of the relationship of $\mathrm{rmse}_\mu(s_{1,t})$ and $\mathrm{rmse}_m(s_{1,t})$ against the true correlation ρ respectively. The residual plot for both fitting in Figure 4(b) and Figure 4(d) show no particular pattern.

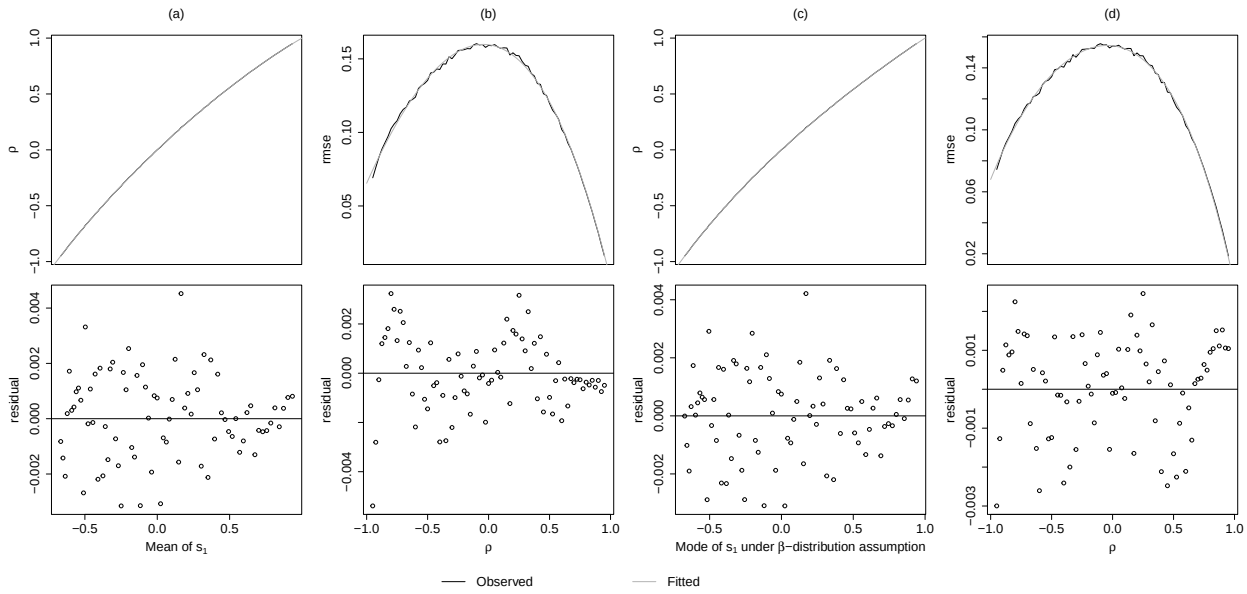


FIGURE 4. (a) Upper: Observed correlation $\bar{s}_{1,t}$ and fitted correlation $\hat{\rho} = f_\mu(\bar{s}_{1,t})$ against $s_{1,t}$; Lower: residuals $\big(\bar{s}_1 - f_\mu(\bar{s}_{1,t})\big)$ against $s_{1,t}$; (b) Upper: observed $\mathrm{RMSE}_\mu(s_{1,t})$ and fitted $g_\mu(\rho)$ against true correlation ρ; Lower: residuals $\big(\mathrm{RMSE}_\mu(s_{1,t}) - g_\mu(\rho)\big)$ against ρ; (c) Upper: Observed correlation $\tilde{s}_{1,t}$ and fitted correlation $\hat{\rho} = f_{m1}(\tilde{s}_{1,t})$ against $s_{1,t}$; Lower: residuals $\big(\tilde{s}_1 - f_{m1}(\tilde{s}_{1,t})\big)$ against $\tilde{s}_{1,t}$ which is the mode of the beta distribution fitted to the distribution of s_{1j}; (d) Upper: observed $\mathrm{RMSE}_{m1}(s_{1,t})$ and fitted $g_{m1}(\rho)$ against true correlation ρ; Lower: residuals $\big(\mathrm{RMSE}_{m1}(s_{1,t}) - g_{m1}(\rho)\big)$ against ρ.

2.4.4. Correlation measures based on the average of $\bar{s}_k$ or $\tilde{s}_k$. Under the assumption of symmetric priors, the simple average estimator is the linear combination of estimators s_k is given by

$$\mathrm{cor}(r_{1,t}, r_{2,t}) = \frac{1}{8}\left[\sum_{k=1}^{4} f(s_{k,t}) - \sum_{k=5}^{8} f(-s_{k,t})\right], \tag{2.45}$$

where $f(\cdot) = f_\mu(\cdot)$ in (2.36) for mean estimator, or $f(\cdot) = f_m(\cdot)$ in (2.39) for mode estimator. This corresponds to the average estimator $A'_K(D(\boldsymbol{\kappa}))$ in (2.D.8) with lower Bayes risk than $D(\boldsymbol{\kappa})$.

However, one issue with the simple average estimator is that the variance of estimators $\{s_{1,t}, s_{2,t}, s_{3,t}, s_{4,t}\}$ and $\{s_{5,t}, s_{6,t}, s_{7,t}, s_{8,t}\}$ differs a lot when there is a strong correlation. For instance, when the correlation is close to 1, we have more confidence in the estimation using estimators $\{s_{1,t}, s_{2,t}, s_{3,t}, s_{4,t}\}$ than using estimators $\{s_{5,t}, s_{6,t}, s_{7,t}, s_{8,t}\}$. For instance, BTC and ETH are strongly positively correlated so that the symmetric prior $g_\rho(\cdot)$ (see (2.D.9) in Lemma 2.D.3) assumption fails. While the simple average estimator will have a smaller Bayes risk when the prior is symmetric compared to

$$\mathrm{cor}_0(r_{1,t}, r_{2,t}) = \frac{1}{4}\sum_{k=1}^{4} f(s_{k,t}), \tag{2.46}$$

which will have a larger Bayes risk if the prior is strongly left-skewed. To overcome such an issue, instead of having an equal weight in (2.4.1), we set different weights on $\tilde{D}(\kappa)$ and $-\tilde{D}(T_2(\kappa))$

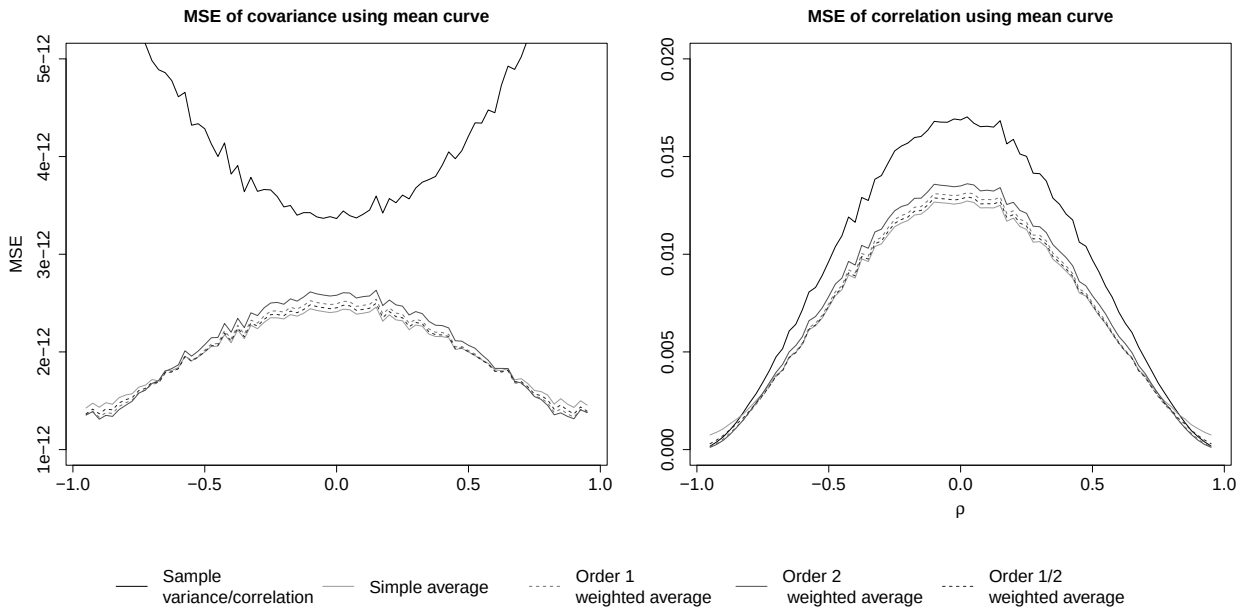


FIGURE 5. MSE of covariance (left) and correlation (right) measures for sample and range-based estimators using mean function and different weightings (method $\mathfrak{m} = 1, 2.2 - 2.5$ respectively) across true correlation ρ.

base on their efficiencies. Thus, we propose a weighted average estimator given by

$$\mathrm{cor}_{\mathfrak{m}}(r_{1,t}, r_{2,t}) = \frac{\sum_{k=1}^{4} f(s_{k,t}) \big/ f_r\big(f(s_{k,t})\big)^p - \sum_{k=5}^{8} f(-s_{k,t}) \big/ f_r\big(-f(s_{k,t})\big)^p}{\sum_{k=1}^{4} 1 \big/ f_r\big(f(s_{k,t})\big)^p + \sum_{k=5}^{8} 1 \big/ f_r\big(-f(-s_{k,t})\big)^p} \tag{2.47}$$

where $f_r(\rho) = g_m(\cdot)$ for mode estimator. Such a weighted average estimator still satisfies (2.12), (2.13) and (2.14), but more weight is distributed to $D_k(\boldsymbol{\tau})$ with lower error. For instance, when $\rho \to 1$, $D_k(\boldsymbol{\kappa}), k = 1, 2, 3, 4$ would be much more efficient comparing to $D_k(\boldsymbol{\kappa}), k = 5, 6, 7, 8$. It would be more appropriate to have a higher weight on $D_k(\boldsymbol{\kappa}), k = 1, 2, 3, 4$ compared to having an equal weight. Note that $f_r(f(s_{k,t})) \to 0$ for $k = 1, 2, 3, 4$, when $\rho \to 1$. Thus our proposed weighted average estimator would eventually degenerate to 1 for $p > 0$, while equally weighted average would still suffer from the error arising from $D_k(\boldsymbol{\kappa}), k = 5, 6, 7, 8$.

The simple average estimator with $p = 0$ and three weighted average estimators with $p = 1, 2, 1/2$ in (2.50) and $f(\cdot) = f_m(\cdot)\,(f_\mu(\cdot))$ are labelled as methods $\mathfrak{m} = 2.2, 2.3, 2.4, 2.5$ (1.2,1.3,1.4,1.5) respectively using the mode information $\tilde{s}_k$ (mean information $\bar{s}_k$).

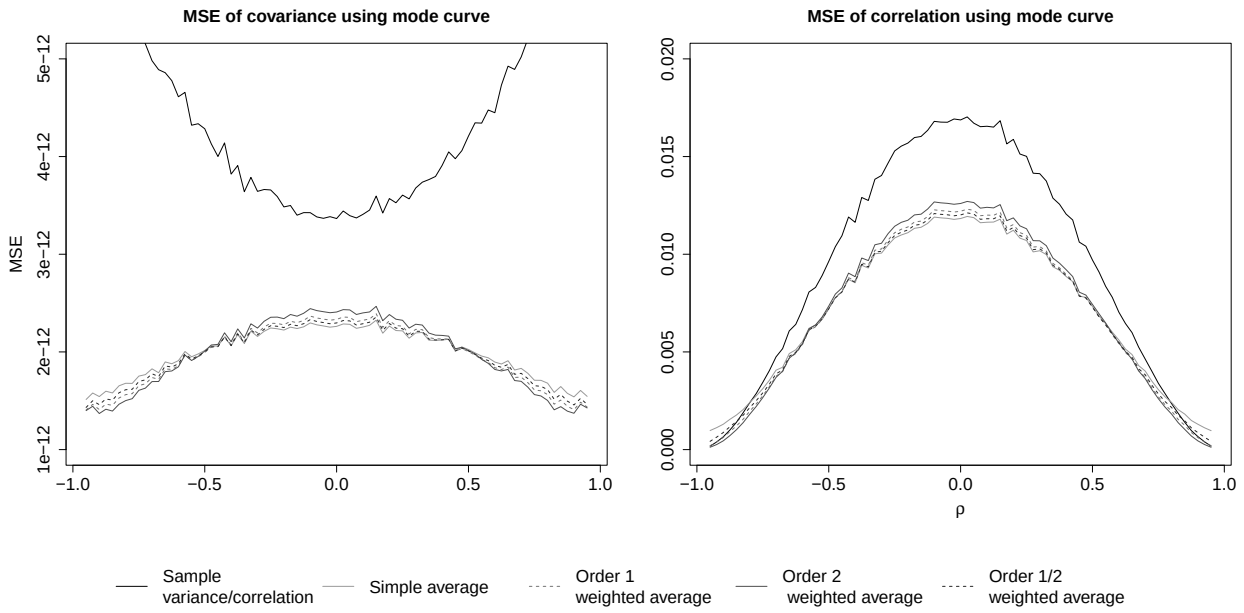


FIGURE 6. MSE of covariance (left) and correlation (right) measures for methods sample and range-based estimators using mode function and different weightings (method $\mathfrak{m} = 1, 2.2 - 2.5$ respectively) across true correlation ρ.

2.4.5. Comparison between mean and mode estimators. To assess the performance of these correlation estimates, we use the MSE for ρ as

$$\text{MSE}_{\mathfrak{m}} = \frac{1}{N}\sum_{t=1}^{N}(\rho - \hat{\rho}_{\mathfrak{m},t})^2 \tag{2.48}$$

where $\hat{\rho}_{\mathfrak{m},t} = \text{cor}_{\mathfrak{m}}(r_{1t}, r_{2t})$ and $\mathfrak{m}$ denotes methods 1.2-1.5, 2.2-2.5 and 1 (benchmark sample correlation estimator in (1.24)). These MSEs are graphed in Figure 5 (Figure 6) across true correlation ρ for methods 1.2-1.5 (2.2-2.5) using the mean (mode) function and method 1 using sample variance/correlation. For comparing covariances, the covariance estimation is calculated by

$$\text{cov}_{\mathfrak{m}}(r_{1,t}, r_{2,t}) = \text{cor}_{\mathfrak{m}}(r_{1,t}, r_{2,t})\hat{\sigma}_{\text{PK},1,t}\hat{\sigma}_{\text{PK},2,t} \tag{2.49}$$

where $\hat{\sigma}^2_{\text{PK},t}$ is given in (1.17). Figure 6 demonstrates that $\hat{\rho}$ achieves the best performance when estimated using Method 2.4, which corresponds to the variance-weighted mode estimator employing the function $f_{m1}(\cdot)$ as defined in (2.39). This method consistently performs well across all values of ρ, highlighting its robustness and reliability.

In summary, the proposed correlation estimates include the simple and weighted average estimates in (2.50) with $p = 0, 1, 2, 1/2$ respectively denoting the equal weight, rmse, mse and root rmse weights by matching the mean $\bar{s}_{k,t}$ of the distribution of $s_{k,t}$ (method 1.2 to 1.5) and matching the mode $\tilde{s}_{k,t}$ of distribution of $s_{k,t}$ (method 2.2 to 2.5). They are compared to the benchmark sample correlation estimate in (1.24).

Note that the variance of $s_{k,t}$ is a function of the number of observations n per day. We show that the MSE decays at a rate of $o(n^{-1})$ asymptotically in Appendix 2.G. Thus the weights in (2.47)

given by

$$\big[1/f_r(s_{k,t})\big]/\bigg[\sum_{k=1}^{8} 1/f_r(s_{k,t})\bigg]$$

do not depend on n asymptotically. Hence, we can also find that the weights are invariant to the number of intra-day observations n by rescaling (2.44) (similarly for (2.43)) to get a generic weighting function with unit intercept given by

$$f_{wr}(\rho) = 1 - 0.0571\rho - 0.6497\rho^2 - 0.1633\rho^3 - 0.1299\rho^4.$$

Then our proposed estimator with order p is given by

$$\mathrm{cor}_{\mathfrak{m}}(r_{1,t}, r_{2,t}) = \frac{\sum_{k=1}^{4} f_m(s_{k,t})/f_{wr}(f_m(s_{k,t}))^p - \sum_{k=5}^{8} f_m(-s_{k,t})/f_{wr}(-f_m(s_{k,t}))^p}{\sum_{k=1}^{4} 1/f_{wr}(f_m(s_{k,t}))^p + \sum_{k=5}^{8} 1/f_{wr}(-f_m(-s_{k,t}))^p}. \quad (2.50)$$

In general, our proposed correlation estimator for asset i, j is $\mathrm{cor}_{\mathfrak{m}}(r_{i,t}, r_{j,t})$.

2.5. Simulation studies on the properties of proposed estimators under more general conditions

Our proposed estimators are built under the bivariate normal assumption. In this section, we study the performance of our proposed estimators under a more general setting. We first test the accuracy of proposed estimators under the presence of heteroskedasticity. Then we address the issue of a possible non-positive definite correlation matrix in a high dimensional setting.

2.5.1. Heteroskedasticity. One can observe in Figure 2 that residual variance decreases as ρ tends to 1 or -1. Hence, heteroscedasticity may present since the variance of $s_{1:4}$ ($s_{5:8}$) are non-constant and converges to 0 as $\rho \to 1$ ($\rho \to -1$).

To assess the performance of the proposed correlation measures in Sections 2.4.3 and 2.4.4 the simulation study in Section 2.4.2 is repeated but the fixed $\sigma_i^2, i = 1, 2$ in (2.29) are replaced by $V_{i,t},\ t = 1, \ldots, N$ ($N = 10000$). The daily volatility models apply the basic CARR(1,1) structure as in (1.53) to (1.55) with $p = q = 1$, that is

$$\begin{aligned} V_{i,t} &= \lambda_{i,t}\epsilon_{i,t} \\ \lambda_{i,t} &= a_{0i} + a_i V_{i,t-1} + b_i \lambda_{i,t-1} \\ \epsilon_{i,t} &\sim f_{\mathrm{G}}(\cdot|\alpha, \beta) \end{aligned} \quad (2.51)$$

where $f_{\mathrm{G}}(\cdot|\alpha, \beta)$ denotes the gamma densities with shape parameter α and scale parameter β and

$$a_{01} = 0.00001, a_{02} = 0.00002, a_1 = 0.1, a_2 = 0.15, b_1 = 0.2, b_2 = 0.1.$$

The volatility is assumed to be constant within a day. The choices of these true parameter values are based on real applications. The gamma distribution is chosen for the errors $\epsilon_{i,t}$ due to its relationship with the Chi-square distribution. Let a random variable χ follow a chi-square distribution with degrees of freedom 200. Then Gamma(100,100) corresponds to the distribution of $\frac{\chi}{100}$ such that $\mathrm{E}(\frac{\chi}{100}) = \frac{100}{100} = 1$ and $\mathrm{Var}(\frac{\chi}{100}) = 2 \times 100/100^2 = 0.02$. Three Gamma distributions are considered with parameters $\alpha = \beta = 10, 20, 100$ so that the errors $e_{i,t,j}$ have the mean of 1 and decreasing variance of 0.2, 0.1 and 0.02 respectively. We note that the level of variance for Bitcoin returns is around 0.002, and the variances of equities in the stock market, which are known to be much less volatile than the cryptocurrency market, will be mostly lower than 0.002.

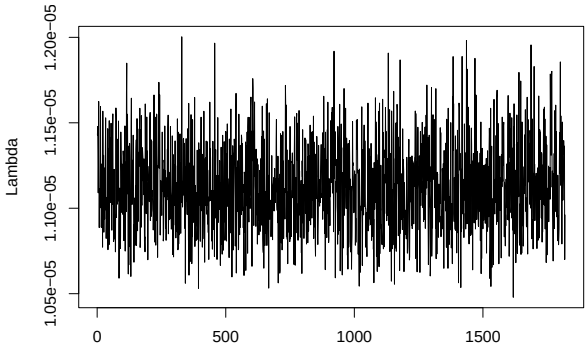


FIGURE 7. Simulated hourly volatility $V_{1,t}$ of asset 1 when the error distribution is Gamma $\Gamma(20, 20)$ in (2.51).

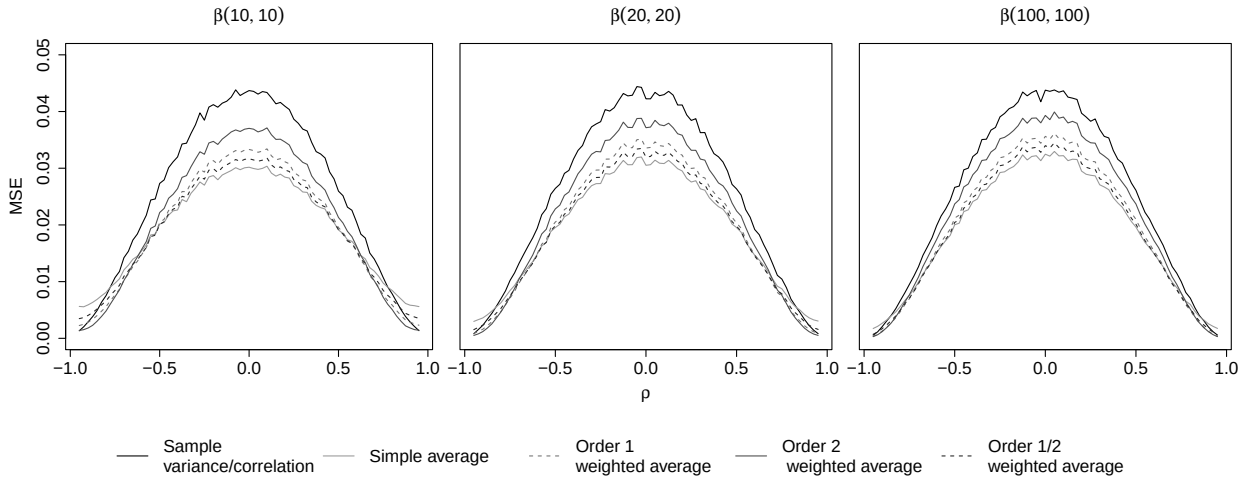


FIGURE 8. MSE of the five correlation estimators using different gamma error distributions for variances

Figure 7 plots the dynamics of $V_{1,t}$ in the long run for asset 1. Figure 8 plots the MSE in (2.48) for estimators $\mathfrak{m} = 1, 2.2 - 2.5$, and results show that the impact of heteroskedasticity on the proposed correlation measures will be minimal.

2.5.2. Adjustment to positive semi-definite correlation matrix. Our proposed methods estimate pairwise correlation $\varrho_{ij,t} = \mathrm{cor}_{\mathfrak{m}}(r_{i,t}, r_{j,t})$ in (2.50). Consequently, our proposed correlation matrix measures are given by

$$\boldsymbol{\mathcal{Q}}_{\mathfrak{m},t} = (\varrho_{ij,t}) \tag{2.52}$$

to estimate the true correlation matrix $\boldsymbol{Q}_t = (\rho_{ij,t})$. However, at any time τ of the day t when observations are made, data is rarely available for all stocks of interest. As the dimension d increases, this unavailability issue is more likely. One way to deal with this problem is to compute correlations of pairs of stocks using data drawn only from the time when data for both stocks are available. Then the resulting correlation matrix will be only an approximate correlation matrix because it was built from inconsistent data. Hence, it is possible that two correlation pairs ρ_{12}, ρ_{13} display high correlation while the third pair ρ_{23} displays relatively lower correlation (dropping day

subscript t). The constraint to ensure a positive definite correlation matrix for any 3 stocks is given by

$$\rho_{12}\rho_{13} - \sqrt{(1-\rho_{12}^2)(1-\rho_{13}^2)} \leq \rho_{23} \leq \rho_{12}\rho_{13} + \sqrt{(1-\rho_{12}^2)(1-\rho_{13}^2)}.$$

Hence, it is possible that the correlation matrix $\boldsymbol{Q}_{\mathfrak{m},t}$ in (2.52) is be non-positive definite when $d \geq 3$. We explain the non-positive definite issue by taking the simple average mode estimator ($\mathfrak{m} = 2.2$) as an example. One can extend the correlation pairs $s_{k,t}$ to correlation matrices $\boldsymbol{S}_{k,t}$ such that

$$\boldsymbol{\mathcal{Q}}_{\mathfrak{m},t} = \frac{1}{8}\sum_{k=1}^{8} \boldsymbol{\mathcal{Q}}_{kt}, \tag{2.53}$$

in (2.50) when the power of the weighting $p = 0$, and

$$\boldsymbol{\mathcal{Q}}_{kt} = f_m(\boldsymbol{S}_{k,t}) = 1.1859\,\boldsymbol{S}_{k,t} - 0.2495\,\boldsymbol{S}_{k,t}\circ\boldsymbol{S}_{k,t} + 0.1007\,\boldsymbol{S}_{k,t}\circ\boldsymbol{S}_{k,t}\circ\boldsymbol{S}_{k} - 0.0371\,\boldsymbol{S}_{k,t}\circ\boldsymbol{S}_{k,t}\circ\boldsymbol{S}_{k,t}\circ\boldsymbol{S}_{k,t}, \tag{2.54}$$

where $f_m(\cdot)$ is given in (2.39), and $\circ$ denotes the Hadamard product. We note that $\boldsymbol{Q}_{kt}$ is still not guaranteed to be positive definite as we cannot restrict the coefficients of the polynomial to be positive when $f_{m1}(\cdot)$ is a concave function, which means the second-order derivative evaluated for a positive input has to be negative.

For any symmetric matrix $\boldsymbol{A} \in \boldsymbol{\mathcal{S}}$ where $\boldsymbol{\mathcal{S}}$ is the set of symmetric matrices, Higham (2002) proposed to compute the nearest positive definite (NPD) correlation matrix $\boldsymbol{X} = \underset{\boldsymbol{X}^* \in \boldsymbol{\mathcal{C}}}{\operatorname{argmin}} \|\boldsymbol{A} - \boldsymbol{X}^*\|_{\mathrm{W}}$, $\boldsymbol{\mathcal{C}}$ is the set of correlation matrices, $\|\cdot\|_{\mathrm{w}}$ is the weighted Frobenius norm defined as $\|\boldsymbol{\Delta}\|_{\mathrm{W}} = \left\|\boldsymbol{W}^{1/2}\boldsymbol{\Delta}\boldsymbol{W}^{1/2}\right\|_{\mathrm{F}}$ for matrix $\boldsymbol{\Delta}$. The weight matrix $\boldsymbol{W} \in \boldsymbol{\mathcal{S}}$ is positive definite, and the Frobenius norm $\|\cdot\|_{\mathrm{F}}$, also called the Euclidean norm, is matrix norm defined as the square root of the sum of the absolute squares of its elements such that $\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2 = \sum_i \sum_j |\delta_{i,j}|^2$ and $\boldsymbol{\Delta} = \left(\delta_{i,j}\right)$.

The iterative procedures to compute $\boldsymbol{X}$ are given by

$$\text{Inital: } \boldsymbol{C}_0 = \boldsymbol{0},\ \boldsymbol{Y}_0 = \boldsymbol{A}, \tag{2.55}$$

$$\text{Update: } \boldsymbol{R}_q = \boldsymbol{Y}_{q-1} - \boldsymbol{C}_{q-1},\ \boldsymbol{X}_q = \mathrm{P}_S(\boldsymbol{R}_q),\ \boldsymbol{C}_q = \boldsymbol{X}_q - \boldsymbol{R}_q,\ \boldsymbol{Y}_q = \mathrm{P}_U(\boldsymbol{X}_q),\ q = 1, 2, .. \tag{2.56}$$

The approach involves projecting the symmetric matrix $\boldsymbol{A}$ onto the space of correlation matrices by identifying a correlation matrix $\boldsymbol{X}$ that minimises the weighted Frobenius norm between $\boldsymbol{A}$ and $\boldsymbol{X}$. The projections $\mathrm{P}_S(\cdot)$ into positive semidefinite matrices and $\mathrm{P}_U(\cdot)$ into correlation matrices are provided in (3.1) and (3.3) of Higham (2002) and $\boldsymbol{C}_q$ is the Dykstra's correction (Boyle and Dykstra, 1986).

In application, we compute the NPD matrix

$$\widehat{\boldsymbol{\mathcal{Q}}}_t = (\varrho_{ij,t}) \tag{2.57}$$

for the symmetric matrix $\boldsymbol{\mathcal{Q}}_t$ in (2.53) using (2.55) and (2.56) as our proposed correlation matrix for each estimator. The command
`nearPD(corr_matrix, keepDiag = TRUE)` in the R package `library(Matrix)` computes the NPD correlation matrix for $\boldsymbol{Q}$, input as `corr_matrix`.

2.5.3. Correlation in high-dimensional correlation matrix using one band structure. With the NPD transformed correlation matrix $\widehat{\boldsymbol{\mathcal{Q}}}_t$, we want to check if it outperforms $\boldsymbol{\mathcal{Q}}_t$ without NPD transformation through a simulation study conducted with the procedures following Section 2.4.2. As NPD transformation is more likely to apply when the dimension d is high, we set $d = 10$. To

reduce the number of ρ values, we assume a d-dimensional one-band correlation structure (ρ in band cells and 0 in non-band cells), that is, the true correlation matrix is given by

$$\boldsymbol{Q} = \begin{pmatrix} 1 & \rho & 0 & \cdots & 0 \\ \rho & 1 & \rho & \ldots & \vdots \\ 0 & \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \ldots & 0 & \rho & 1 \end{pmatrix}. \tag{2.58}$$

We show in Appendix 2.H that a necessary and sufficient condition for $\boldsymbol{Q}$ to be positively definite is $|\rho| < \frac{\sqrt{\tan^2(\pi/(d+1))+1}}{2}$, that is, $|\rho| < 0.5211$ when $d = 10$.

We study the impact of correlation level ρ on the accuracy and precision of the proposed correlation matrix estimators using NPD transformation and compare the impact to those without NPD transformation. To examine the precision of NPD estimators, we set the number of observations n to estimate the daily correlation matrix to be 24, which is much lower than $n = 60$ in the simulation experiments in Section 2.4.2 (Step 2). Based on $N = 10000$ days, Figure 9 plots the Frobenius norm of $\boldsymbol{Q} - \widehat{\boldsymbol{Q}}$, MSE of band cells and MSE of non-band cells using simple and weighted average mode estimators (methods 2.2 and 2.4) and compares them with the sample variance (method 1). Furthermore, the sample covariance matrix is known to exhibit poor performance when the dimensionality of the data is comparable to, or exceeds, the available sample size. Given the limitations of the sample covariance matrix in accurately estimating high-dimensional sparse covariance structures, Fan et al. (2013) proposed the Principal Orthogonal Complement Thresholding (POET) method. POET estimates the covariance matrix by decomposing it into a low-rank factor-driven part (via PCA) and a sparse residual component (via thresholding), and then combines them. It's designed for high-dimensional settings with latent common structures and weak residual correlations. We compare the performance of our proposed method against POET in Figure 9 as well.

The results from the left plot show that the simple average mode estimator (method 2.2) has the lowest difference after the NPD transformation. Moreover, the middle and right plots show that this NPD transformed estimator shows improvement after the NPD transformation. This improvement applies to both band and non-band cells, and the NPD transformed estimator consistently displays the lowest MSE across levels of ρ. Moreover, the middle plot also shows that MSEs of all estimators drop as $\rho \to 0.5$ in the band cells. These results agree with Figure 5 (right plot) that the simple average mode estimator is the best when ρ lies around -0.5 to 0.5 and MSE drops as $\rho \to 0.5$.

In summary, the NPD transformed simple average mode estimator ($\mathfrak{m} = 2.2$) provides the best performance with the lowest MSE and will be adopted in the applications because the level of correlation is expected to be moderate.

2.6. Relation to bivariate copula

Ultimately, a maximum likelihood estimator can be constructed based on the marginal distribution of $(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2)$. It is well known that the marginal distribution of $h_{i,t,\tau}$, the high price component of asset i at time τ on day t, follows a half-normal distribution, denoted $N_+(\sigma_{i,t})$, where $\sigma_{i,t}$ is the asset's daily volatility (Freedman, 2012). The corresponding low component, $l_{i,t,\tau}$, has the same distribution as the reflection of $h_{i,t,\tau}$ about zero. As a result, determining the joint marginal

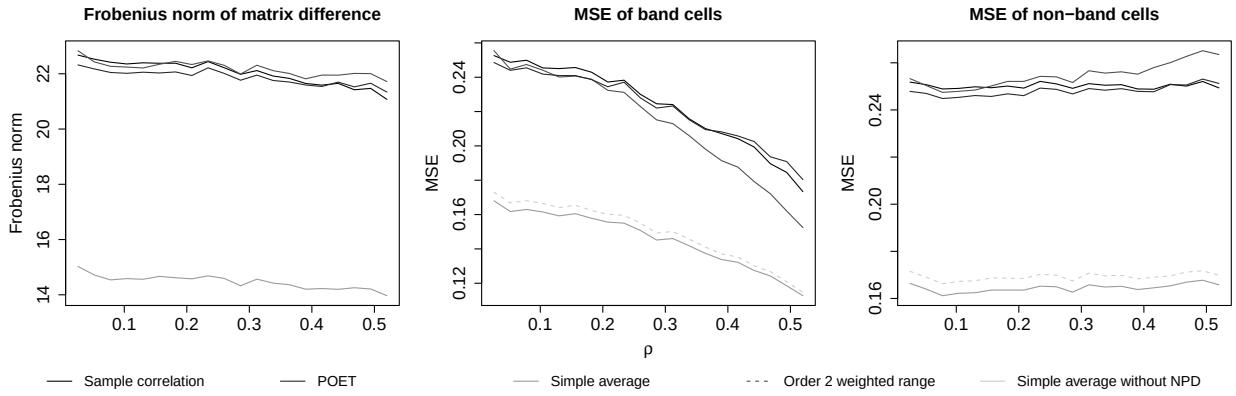


FIGURE 9. Frobenlus norm of $\boldsymbol{Q} - \widehat{\boldsymbol{Q}}$, MSE of correlation estimate $\hat{\rho}$ and MSE of non-band cells using estimators $\widehat{\boldsymbol{Q}}$ (sample $\mathfrak{m} = 1$, POET, simple average mode estimator $\mathfrak{m} = 2.2$ and MSE weighted average mode estimator $\mathfrak{m} = 2.4$) with NPD transformation and without NPD transformation against levels of $\rho \in (0, 0.5211)$ in (2.58).

distribution of $(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2)$ becomes equivalent to identifying the copula that links these marginal distributions. Similarly, if the bivariate copula capturing the dependence between $h_{1,t,\tau}$ and $h_{2,t,\tau}$ can be accurately specified, a more efficient estimator $D_1(\boldsymbol{\kappa})$ may be derived. Unfortunately, while the joint distributions of $(h_{1,t,\tau}, c_{1,t,\tau})$ and $(h_{1,t,\tau}, c_{2,t,\tau})$ can be explicitly characterized using the reflection principle, the joint distribution of $(h_{1,t,\tau}, h_{2,t,\tau})$ remains unknown for arbitrary correlation ρ. Therefore, in this section, we restrict our attention to the special case where $\rho = -1$ and illustrate how the choice of bivariate copula and the associated correlation measures influence the behaviour of our estimator. For the purpose of estimating $s_{1,t}$, we employ the sample correlation coefficient as our primary correlation estimator, as $h_{1,t,\tau}$ and $h_{2,t,\tau}$ exhibit not only a strict one-to-one correspondence but also a strictly linear relationship when $\rho = 1$. However, alternative dependence measures—such as Spearman's rank correlation or Kendall's τ—could also be considered, particularly under non-Gaussian settings. We further investigate the behaviour of these correlation measures in the case $\rho = -1$, to highlight how dependence structure influences the effectiveness of our proposed estimator.

As we discuss previously, when $\rho = -1$, we have $\operatorname{cor}(\boldsymbol{h}_{1,t}, \boldsymbol{h}_{2,t}) = -\operatorname{cor}(\boldsymbol{h}_{1,t}, \boldsymbol{l}_{1,t})$. Thus, the problem is reduced to the marginal distribution of the maximum and minimum of a Brownian motion. Now consider a standard Brownian W_t. Let $h = \max_{t\in(0,1)} W_t, l = \min_{t\in(0,1)} W_t$. Denote $\operatorname{Cor}(h, l)$ and $\operatorname{Cov}(h, l)$ as the population correlation and covariance respectively, then we have

$$\operatorname{Cor}(h, l) = \frac{\operatorname{Cov}(h, l)}{\sqrt{\operatorname{Var}(h)\operatorname{Var}(l)}} = \frac{\mathbb{E}(hl) - \mathbb{E}(h)\mathbb{E}(l)}{\sqrt{(\mathbb{E}(h^2) - \mathbb{E}(h)^2)(\mathbb{E}(l^2) - \mathbb{E}(l)^2)}} \tag{2.59}$$

We know that h and $-l$ follow the half-normal distributions.

Proposition 2.6.1. *The marginal cumulative distribution of l is given by*

$$F_1(x) = \begin{cases} 2\Phi(x) & x < 0 \\ 1 & x \leq 0 \end{cases}, \tag{2.60}$$

where Φ is the cumulative distribution function of the standard normal. The density function of l is given by

$$f_1(x) = 2\phi(x)\mathbb{1}_{x<0}, \tag{2.61}$$

where ϕ is the density function of the standard normal. The marginal cumulative distribution of h is given by

$$F_2(x) = \begin{cases} 0 & x \leq 0 \\ 2\Phi(x) - 1 & x > 0 \end{cases}, \tag{2.62}$$

The density function of h is given by

$$f_2(x) = 2\phi(x)\mathbb{1}_{x<0}. \tag{2.63}$$

Hence we have

$$\mathbb{E}(h^2) = \mathbb{E}(l^2) = 1, \quad \mathbb{E}(h) = -\mathbb{E}(l) = \sqrt{\frac{2}{\pi}} \tag{2.64}$$

Furthermore, we have $\mathbb{E}[(h-l)^2] = 4\log 2$, which is exactly the scaling coefficient in PK measure. See Feller (1951) for more details. Thus, we finally have

$$\begin{aligned} \mathbb{E}(hl) &= \frac{\mathbb{E}(h^2) + \mathbb{E}(l^2) - \mathbb{E}[(h-l)^2]}{2} \\ &= \frac{1 + 1 - 4\log 2}{2} \\ &= 1 - 2\log 2 \\ Cor(h,l) &= \frac{1 - 2\log 2 + \frac{2}{\pi}}{1 - \frac{2}{\pi}} \approx 0.6889. \end{aligned} \tag{2.65}$$

The empirical Pearson correlation from our simulations is 0.6887, close to the theoretical value. Furthermore, if we plug (2.65) into (2.36), we can find that $f_\mu(-0.6887) = -0.9327$, which is quite close to -1.

The precise joint distribution can be obtained through the reflection principle.

Theorem 2.6.2.

$$P(x < l, h \leq y) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \int_x^y (e^{-\frac{(z-2n(y-x))^2}{2}} - e^{-\frac{(z-2x-2n(y-x))^2}{2}}) dz \tag{2.66}$$

Proof. See Freedman (2012). □

While only a small number of n is required to get an adequate approximation of $P(x < l, h < y)$ when $|x|$ and y are large, the sequence experiences a low convergency rate when $|x|$ and y are small. Let

$$g(x,y) = \sum_{n=1}^{\infty} \frac{4}{(2n-1)\pi} e^{-\frac{(2n-1)^2\pi^2}{2(y+x)^2}} \sin\frac{(2n-1)\pi x}{y+x}. \tag{2.67}$$

Lowther (2023) transformed (2.6.2) so that expansion will also converge quickly for small values of $|x|$ and y.

Proposition 2.6.3.

$$P(x < l, h \leq y) = g(-x, y) \tag{2.68}$$

Proof. See Lowther (2023). □

Lemma 2.6.4. *The joint cumulative density of (h, l) is given by*

$$P(l \leq x, h \leq y) = P(h \leq y) - P(x < l, h \leq y) = F_2(y) - g(-x, y). \tag{2.69}$$

Proof. Follows immediately form Proposition 2.6.1 and 2.6.3. □

Let $C(u_1, u_2)$ denote the copula between h and l. By Sklar's theorem, we have

$$C(F_1(x), F_2(y) = F_2(y) - g(-x, y). \tag{2.70}$$

The Spearman correlation can be represented as a function of the underlying copula.

Theorem 2.6.5. *The population Spearman correlation is given by*

$$12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3. \tag{2.71}$$

Proof. See Nelsen (2006). □

Proposition 2.6.6. *The population Spearman correlation between h and l is given by*

$$3 - 48 \int_0^\infty \int_0^\infty g(x, y)\phi(x)\phi(y) dx dy, \tag{2.72}$$

where ϕ is the density of the standard normal distribution.

Proof. By taking transformation

$$u_1 = F_1(x), u_2 = F_2(y). \tag{2.73}$$

we have

$$\begin{aligned}
\int_0^1 \int_0^1 C(u_1, u_1) du_1 du_2 &= \int_0^\infty \int_{-\infty}^0 C(F_1(x), F_2(y)) f_1(x) f_2(y) dx dy \\
&= \int_0^\infty \int_\infty^0 P(h \leq x, l \leq y) f_1(x) f_2(y) dx dy \\
&= \int_0^\infty \int_\infty^0 (F_2(y) - g(-x, y)) f_1(x) f_2(y) dx dy \\
&= \int_0^\infty \int_\infty^0 F_2(y) f_1(x) f_2(y) dx dy - \int_0^\infty \int_\infty^0 g(-x, y) f_1(x) f_2(y) dx dy \\
&= \int_0^\infty F_2(y) f_2(y) \int_\infty^0 f_1(x) dx dy - 4 \int_0^\infty \int_0^\infty g(x, y)\phi(x)\phi(y) dx dy \\
&= \int_0^\infty F_2(y) f_2(y) dy - 4 \int_0^\infty \int_0^\infty g(x, y)\phi(x)\phi(y) dx dy \\
&= \frac{1}{2} - 4 \int_0^\infty \int_0^\infty g(x, y)\phi(x)\phi(y) dx dy
\end{aligned}$$

□

Let

$$g_k(x, y) = \frac{4}{(2n-1)\pi} e^{-\frac{(2n-1)^2\pi^2}{2(y+x)^2}} \sin \frac{(2n-1)\pi x}{y + x}. \tag{2.74}$$

TABLE 1. Summaries of integral of g_n for different n.

n	1	2	3	4	5
Integral	0.1830	-2.347028×10^4	6.7613×10^{-7}	-3.1914×10^{-9}	-2.5409×10^{-11}
Cumulative sum	0.1830	0.1828	0.1828	0.1828	0.1828
Correlation	0.8037	0.8065	0.8065	0.8065	0.8065

Assume that $\sum_{n=1}^{k} g_k(x,y)$ is uniformly bounded by a constant for all $x, y > 0$. Then we have

$$\int_0^\infty \int_0^\infty g(x,y)\phi(x)\phi(y)dxdy = \lim_{n\to\infty} \sum_{k=1}^{n} \int_0^\infty \int_0^\infty g_k(x,y)\phi(x)\phi(y)dxdy \tag{2.75}$$

by the dominated convergence theorem. We use `R` to evaluate $\int_0^\infty \int_0^\infty g_k(x,y)\phi(x)\phi(y)dxdy$ and the cumulative sum of $\int_0^\infty \int_0^\infty g_k(x,y)\phi(x)\phi(y)dxdy$. The result is summarised in Table 1. The third row is the approximation of Spearman correlation by substituting the approximation of the integral in 2.6.6.

We can observe that the summation converges quickly. The empirical Spearman correlation, which is 0.8063 from our simulation study, is consistent with the approximation.

Theorem 2.6.7. *The population Kendall's τ is given by*

$$4 \int_{[0,1]\times[0,1]} C(u_1,u_2)dC(u_1,u_2) - 1. \tag{2.76}$$

Proof. See Nelsen (2006). □

Proposition 2.6.8. *The population Spearman correlation between h and l is given by*

$$1 - 4 \int_{[0,\infty]\times[0,\infty]} g(x,y)dg(x,y). \tag{2.77}$$

Proof. By taking transformation

$$u_1 = F_1(x), u_2 = F_2(y). \tag{2.78}$$

we have

$$\begin{aligned}
\int_{[0,1]\times[0,1]} C(u_1,u_1)dC(u_1,u_2) &= \int_{[-\infty,0]\times[0,\infty]} C(F_1(x),F_2(y))dC(F_1(x),F_2(y)) \\
&= \int_{[-\infty,0]\times[0,\infty]} P(l \le x, h \le y)dP(l \le x, h \le y) \\
&= \int_{[-\infty,0]\times[0,\infty]} (F_2(y) - g(-x,y))dP(l \le x, h \le y) \\
&= \int_{[-\infty,0]\times[0,\infty]} F_2(y)dP(l \le x, h \le y) \\
&\quad - \int_{[-\infty,0]\times[0,\infty]} g(-x,y))dP(l \le x, h \le y).
\end{aligned}$$

Note that if we fix y and integrate over x, we shall get

$$\int_{[0,\infty]\times\{y\}} dP(l \le x, h \le y) = f_2(y)dy \tag{2.79}$$

Thus, the first term is given by

$$\int_{[0,\infty]\times[0,\infty]} F_2(y)dP(l \le x, h \le y) = \int_0^\infty F_2(y)f_2(y)dy = \frac{1}{2} \tag{2.80}$$

Furthermore, as (h, l) are continuous random variables, $P(h < x, l < y)$ must be differentiable. Hence we have

$$dP(h \le x, l \le y) = \frac{\partial^2}{\partial x \partial y}\big(F_2(y) - g(-x, y)\big) = -\frac{\partial^2}{\partial x \partial y} g(-x, y) = -dg(-x, y). \tag{2.81}$$

$$\begin{aligned}\int_{[0,1]\times[0,1]} C(u_1, u_1)dC(u_1, u_2) &= \frac{1}{2} - \int_{[-\infty,0]\times[0,\infty]} -g(-x, y)dg(-x, y) \\ &= \frac{1}{2} - \int_{[0,\infty]\times[0,\infty]} g(x, y)dg(x, y)\end{aligned}$$

□

To evaluate the integral, we first notice that

$$\int_{[-\infty,0]\times[a,\infty]} P(l \le x, h \le y)dP(l \le x, h \le y) < \int_{[-\infty,0]\times[a,\infty]} dP(l \le x, h \le y) \tag{2.82}$$

$$= P(h > a). \tag{2.83}$$

As $P(h > 5) \approx 6 \times 10^{-7}$, it is appropriate to approximate the integral by setting the range of y to be $[0, 5]$. A similar argument follows for setting the range of x to be $[0, 5]$ as well. Alomari (2019) proposed the numerical formula for the Riemann-Stieltjes integral with subdivision I_{m_1} : $a = x_0 < x_1 < \cdots < x_{m_1-1} < x_n = b$, and $J_{m_2} : c = y_0 < y_1 < \cdots < y_{m_2-1} < y_{m_2} = d$ is given by

$$\begin{aligned}&\int_a^b \int_c^d f(x, y)dh(x, y) \\ &\quad = \sum_{i=0}^{m_1-1} \sum_{j=0}^{m_2-1} \frac{h(x_{i+1}, y_{j+1}) - h(x_{i+1}, y_j) - h(x_i, y_{j+1}) + h(x_i, y_j)}{(x_{i+1} - x_i)(y_{j+1} - y_j)} \int_{x_i}^{x_{i+1}} \int_{y_j}^{y_{j+1}} f(t, s)dtds\end{aligned} \tag{2.84}$$

In our application, we set $m_1 = m_2 = m$ for simplicity. We evaluate the integral

$$\int_{[0,\infty]\times[0,\infty]} g(x, y)dg(x, y) \approx \int_{[0,5]\times[0,5]} g(x, y)dg(x, y) \tag{2.85}$$

using $f(x, y) = h(x, y) = \sum_{k=1}^n g_k(x, y)$ for different value of n and check the convergence. We set an evenly spaced subdivision. The summary of the Riemann-Stieltjes integral for m and n, which are different numbers of subintervals m and numbers of terms in $g(x, y)$ respectively, is provided in Table 2. We can find that for $n = 6, m = 200$, we can have an adequate approximation. The empirical Kendall's τ from our simulation is 0.5981, which is consistent with the theoretical value of population Kendall's τ, being $1 - 4 * 1.005 = 0.5980$.

TABLE 2. Values of Riemann-Stieljes interval of $\sum_{k=1}^{n} g_k$ with numbers of subintervals $m_1 = m_2 = m$.

m\n	4	6	8	10
50	0.09920	0.1015	0.1015	0.1015
100	0.0984	0.1007	0.1007	0.1007
200	0.0982	0.1005	0.1005	0.1005
300	0.0982	0.1005	0.1004	0.1005

2.7. Application of range-based correlation measures in statistical testing

In this section, we demonstrate applications of our proposed covariance matrix estimators in hypothesis testing. Firstly, we apply the estimators to estimate the covariance matrix of different market sectors of S&P 500, compare the estimated covariance matrices with those using exchange-traded funds (ETFs) and test if the two covariance matrices from the indices and ETFs are equal. Next, we also show that our proposed covariance matrix estimators have higher power than the sample covariance estimator to reject the null hypothesis of no difference when the covariance matrix does differ. An ETF is a type of investment fund that is traded on stock exchanges, similar to individual stocks. ETFs are designed to provide investors with exposure to a broad range of assets, including stocks, bonds, commodities, or other securities, which are either unavailable or impracticable to open a position. The performance of ETF is measured neither by the return nor the volatility. The ability to replicate the performance of an underlying index, asset class, or benchmark is the key performance indicator. For widely traded ETFs issued by well-known brokers, we would expect exceptional trackability.

2.7.1. Data. We collect 15-minute OHLC data from TradingView for ten S&P500 sector indices, namely information technology, health care, financials, real estate, energy, materials, consumer discretionary, industrials, utilities and consumer staples. We choose ETFs issued by State Street to track S&P500 sector indices. We do not include the communications services sector in our study as the trading volume of the corresponding ETF is much smaller compared to other sectors. We use intraday data to estimate the covariance matrix of a particular day. The period of interest is from January 1, 2022, to March 1, 2024, with $N = 543$ trading days. The abbreviations for the sectors of S&P 500 indices and ETFs are provided in Table 3 under the columns labelled $\boldsymbol{x}$ and $\boldsymbol{y}_1$ respectively. We also consider three alternative constructions for $\boldsymbol{y}_1$. For $\boldsymbol{y}_2$, we replace XLK with Apple (AAPL), a leading company in the information technology sector, to represent a more specific asset within the sector. In $\boldsymbol{y}_3$, we swap the roles of XLV and XLK, altering their relative positions in the analysis. Finally, in $\boldsymbol{y}_4$, we replace XLK with a white noise process, introducing a random, non-sectoral component into the framework. Under this construction, $\boldsymbol{y}_2$ closely aligns with the null hypothesis, as it retains a similar underlying structure with a meaningful representation of the information technology sector. Conversely, $\boldsymbol{y}_4$ deviates significantly from the null hypothesis due to the inclusion of a white noise process, which lacks any economic or sectoral significance. This variation allows us to systematically evaluate the sensitivity of the test to different configurations and assess its robustness across varying assumptions. Note that trading volumes in 15 minutes are larger than 1000 transactions in general, with only one exception for the communication sector.

TABLE 3. Abbreviation for the sectors of S&P500 ($\boldsymbol{x}$) and ETFs ($\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \boldsymbol{y}_4$)

Sector	$\boldsymbol{x}$	$\boldsymbol{y}_1$	$\boldsymbol{y}_2$	$\boldsymbol{y}_3$	$\boldsymbol{y}_4$
Information Technology	S5INFT	XLK	**APPL**	**XLV**	**WN**†
Health care	S5HLTH	XLV	XLV	**XLK**	XLV
Financials	SPF	XLF	XLF	XLF	XLF
Real Estate	S5REAS	XLRE	XLRE	XLRE	XLRE
Energy	SPN	XLE	XLE	XLE	XLE
Materials	S5MATR	XLB	XLB	XLB	XLB
Consumer Discretionary	S5COND	XLY	XLY	XLY	XLY
Industrials	S5INDU	XLI	XLI	XLI	XLI
Utilities	S5UTIL	XLU	XLU	XLU	XLU
Consumer Staples	S5CONS	XLP	XLP	XLP	XLP

† white noise

2.7.2. Two sample test. We adopt the test procedure proposed in Cai et al. (2013). The test is repeated for each day t, and we suppress this subscript for simplicity. Let $\{\boldsymbol{X}_{1:n_1}\}$ be the d-dimensional i.i.d samples from $\mathrm{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $\{\boldsymbol{Y}_{1:n_2}\}$ be the d-dimensional i.i.d samples from $\mathrm{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ and $\{\boldsymbol{X}_\tau\}$ are independent from $\{\boldsymbol{Y}_\tau\}$. This test is designed in the high-dimensional setting where d can be much larger than $\max(n_1, n_2)$. We note that the sample size n corresponds to the number of intra-day observations and sub-sample sizes are n_1, n_2 such that $\max(n_1, n_2) < n$. The null and alternative hypotheses of the test are given by:

$$H_0: \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \quad \text{vs} \quad H_1: \quad \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2. \tag{2.86}$$

Given the d-dimensional realised vectors $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_{n_1}\}$ and $\{\boldsymbol{y}_1, ..., \boldsymbol{y}_{n_2}\}$, the sample covariance estimates of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are given by

$$\widehat{\boldsymbol{\Sigma}}_1 = \frac{1}{n_1 - 1} \sum_{\tau=1}^{n_1} (\boldsymbol{x}_\tau - \bar{\boldsymbol{x}})(\boldsymbol{x}_\tau - \bar{\boldsymbol{x}})^\top, \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_2 = \frac{1}{n_2 - 1} \sum_{\tau=1}^{n_2} (\boldsymbol{y}_\tau - \bar{\boldsymbol{y}})(\boldsymbol{y}_\tau - \bar{\boldsymbol{y}})^\top, \tag{2.87}$$

where $\bar{\boldsymbol{x}} = \frac{1}{n_1} \sum_{\tau=1}^{n_1} \boldsymbol{x}_\tau$ and $\bar{\boldsymbol{y}} = \frac{1}{n_2} \sum_{\tau=1}^{n_2} \boldsymbol{y}_\tau$. Then, we define the variance of covariance estimators by

$$\begin{aligned} \hat{\theta}_{i,j,1} &= \frac{1}{n_1 - 1} \sum_{\tau=1}^{n_1} \left[(x_{i,\tau} - \bar{x}_i)(x_{j,\tau} - \bar{x}_j) - \hat{\sigma}_{i,j,1} \right]^2, \\ \hat{\theta}_{i,j,2} &= \frac{1}{n_2 - 1} \sum_{\tau=1}^{n_2} \left[(y_{i,\tau} - \bar{y}_i)(y_{j,\tau} - \bar{y}_j) - \hat{\sigma}_{i,j,2} \right]^2, \end{aligned} \tag{2.88}$$

where $x_{i,\tau}, y_{i,\tau}$ are the i-th entry of $\boldsymbol{x}_\tau$ and $\boldsymbol{y}_\tau$ respectively, $\bar{x}_i = \frac{1}{n_1} \sum_{\tau=1}^{n_1} x_{i,\tau}$ and $\bar{y}_i = \frac{1}{n_2} \sum_{\tau=1}^{n_2} y_{i,\tau}$, and $\hat{\sigma}_{i,j,1}$ and $\hat{\sigma}_{i,j,2}$ are the (i, j) entry of $\widehat{\boldsymbol{\Sigma}}_1$ and $\widehat{\boldsymbol{\Sigma}}_2$ respectively. Then the test statistic based on the maximum of the standardised differences between the entries of the two-sample covariance matrices is given by

$$\mathcal{M} = \max_{i,j} \frac{(\hat{\sigma}_{i,j,1} - \hat{\sigma}_{i,j,2})^2}{\hat{\theta}_{i,j,1}/n_1 + \hat{\theta}_{i,j,2}/n_2}. \tag{2.89}$$

Under H_0, $\mathcal{M}$ follows a Gumbel distribution, $\text{Gumbel}(\mu, \beta)$, where $\mu = 4\log d - \log 8\pi - \log\log d$ and $\beta = 2$ asymptotically. The cumulative density of $\text{Gumbel}(\mu, \beta)$ is given by $F(x) = e^{-e^{(x-\mu)/\beta}}$. Thus, the p-value of the test is given by

$$p\text{-value} = 1 - F(\mathcal{M}) = 1 - e^{-e^{-(\mathcal{M} - 4\log d + \log 8\pi + \log\log d)/2}}. \tag{2.90}$$

This test is developed for testing the sample covariance matrix with independent samples. When using the proposed correlation matrix estimators $\widehat{\boldsymbol{\mathcal{Q}}}_{\mathfrak{m},\varsigma}$ which is the NPD transformed of $\boldsymbol{\mathcal{Q}}_{\mathfrak{m},\varsigma}$ in (2.53) and (2.54), the covariance matrix estimates are given by

$$\widehat{\boldsymbol{\Sigma}}_{\mathfrak{m},\varsigma} = \big(\hat{\sigma}'_{i,j,\varsigma}\big) = \text{Diag}\big([v_{i,\varsigma}^{\text{PK}}]^{1/2}\big)\ \widehat{\boldsymbol{\mathcal{Q}}}_{\mathfrak{m},\varsigma}\ \text{Diag}\big([v_{i,\varsigma}^{\text{PK}}]^{1/2}\big) \tag{2.91}$$

where PK volatility measures v_i^{PK} are given in (1.17) for component i in samples $\varsigma = 1, 2$ using $(\boldsymbol{x}, \boldsymbol{y})$. In our test settings, the two data sets are matched pairs. So they are not independent samples. To adjust for the dependency, we perform Bootstrap sampling with replacement of sizes (n_1, n_2) independently to $(\boldsymbol{x}, \boldsymbol{y})$ containing $n = 60$ 5-minute samples in each day t.

Moreover, the test using (2.89) favours the sample covariance matrix estimator because the variances $\hat{\theta}_{i,j,\varsigma}, \varsigma = 1, 2$ in the denominator are derived using sample variances. So it should be adjusted when applying to our proposed range-based mode estimator. In this case, estimates of variance $\hat{\theta}_{i,j,\varsigma}$ are proposed to be replaced by $\widehat{V}_{i,j,\varsigma} = \text{Var}(\boldsymbol{x}_{i,j,\varsigma,1:n_1}, \boldsymbol{y}_{i,j,\varsigma,1:n_2})$ using Bootstrap samples $\boldsymbol{x}_{i,j,\varsigma,1:n_1}, \boldsymbol{y}_{i,j,\varsigma,1:n_2}$, simulated at $\rho_{i,j,\varsigma,1:n_\varsigma}$, $\varsigma = 1, 2$. Hence, the revised test statistics for range-based covariance matrices are given by (dropping the day t subscript)

$$\mathcal{M}'_{i,j} = \max_{i,j} \frac{(\hat{\sigma}'_{i,j,1} - \hat{\sigma}'_{i,j,2})^2}{\widehat{V}_{i,j,1} + \widehat{V}_{i,j,2}}. \tag{2.92}$$

Appendix 2.I shows the convergence of $\mathcal{M}'_{i,j}$ in distribution using $\widehat{V}_{i,j}$.

We apply the two-sample test to test the hypotheses in (2.86), evaluate the p-value using (2.90), repeat it over $N = 543$ trading days and count the number of rejections of H_0 at 5% significance level. In conducting this test, we set $\boldsymbol{x}$ to be the vector of S&P index returns with $d = 10$ sectors, and $\boldsymbol{y}_1$ to be the vector of ETF returns of the 10 sectors.

To compare the performances of our proposed range-based covariance estimators using simple average mode estimates with the sample covariance estimator under different scenarios, we also modify $\boldsymbol{y}_1$ to obtain $\boldsymbol{y}_2, \boldsymbol{y}_3$ and $\boldsymbol{y}_4$. Details of the data are provided in Table 3. The various tests applying different data pairs $(\boldsymbol{x}, \boldsymbol{y})$ and estimators are defined below.

Standard test: We are interested in comparing the power of the test, that is, the proportion of time rejecting H_0, with different covariance matrix estimators. The test is repeated with $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ estimated by sample covariance matrix in (2.87) as well as the range-based covariance matrix in (2.91) using simple average mode estimator ($\mathfrak{m} = 2.2$). Results are displayed in Table 4 under $\boldsymbol{y}_1$. As the two tests give the same proportion, which are both close to 0.05, we conclude that ETFs replicate the S&P 500 indices successfully.

Replacement test: Now $\boldsymbol{\Sigma}_2$ is constructed replacing XLK (Information technology) with APPL, Apple Inc. In this case, we use $\boldsymbol{x}$ and $\boldsymbol{y}_2$ in Table 3. From Table 4, the proportion of rejecting H_0 using the range-based estimator is 0.127, which is larger than 0.105 using the sample covariance estimator. This shows that the range-based estimator has a higher power than the sample covariance matrix estimator in detecting inconsistency.

Swap test: We swap XLK (Information technology) and XLV (Health care) for $\boldsymbol{\Sigma}_2$ using $\boldsymbol{x}$ and $\boldsymbol{y}_3$ in Table 3. Again, Table 4 under $\boldsymbol{y}_3$ shows that the proportion of rejecting H_0 using the range-based estimator is 0.269, which is larger than 0.1995 using the sample covariance estimator, confirming the higher power of the range-based estimator.

White noise test: Lastly, we replace XLK of ETFs with a white noise (WN) process and apply data $\boldsymbol{x}$ and $\boldsymbol{y}_4$ to the test. The results in Table 4 under $\boldsymbol{y}_4$ show that the power using the proposed range-based estimator is again higher.

Applying the WN to different components, results consistently confirm the higher power of our proposed range-based estimator. Consistent results are also obtained when the components to be replaced by a WH process are increased to 2.

TABLE 4. Proportion of rejection of H_0 out of 543 trading days for the two-sample tests using $\boldsymbol{x}$ and various $\boldsymbol{y}$

Estimator	$\boldsymbol{y}_1$	$\boldsymbol{y}_2$	$\boldsymbol{y}_3$	$\boldsymbol{y}_4$
Range-based simple average mode estimator ($\mathfrak{m} = 2.2$)	0.070	0.127	0.269	0.353
Sample correlation estimator ($\mathfrak{m} = 1$)	0.070	0.105	0.199	0.273

2.7.3. One sample test. The `covTestR` package contains seven one-sample tests of covariance matrices under different conditions for d and n and different distribution assumptions (Ahmad and von Rosen, 2015; Chen et al., 2010a; Fisher, 2012; Ledoit and Wolf, 2002; Nagao, 1973; Srivastava, 2005; Srivastava et al., 2011). The commands of these tests in `R` are `Ahmad2015(x,Sigma= "identity",...)`, `Chen2020(...)`, `Fisher2012(...)`, `LedoitWolf2` `Nagao1973(...)`, `Srivastava2005(...)`, `Srivastava2011(...)`, respectively. We first check the performance of these tests and then apply the tests to assess the performance of our proposed range-based estimator focusing again on the simple average mode estimator.

Simulation study: Firstly, a simulation study is conducted to check the performance of various proposed tests using the one-band matrix structure in (2.58) with increasing ρ, dimension $d = 10$, sample size $n = 26$, and repeats for $N = 10000$ days. For each day t, assume $\boldsymbol{X}_\tau \sim \mathrm{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ dropping the subscript t. The hypotheses are

$$H_0 : \boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_0 \quad \text{vs} \quad H_1 : \boldsymbol{\Sigma}_x \neq \boldsymbol{\Sigma}_0 \tag{2.93}$$

where $\boldsymbol{\Sigma}_0$ is the true covariance matrix. Writing $\boldsymbol{\Sigma}_0 = \boldsymbol{L}\boldsymbol{L}^\top$ in Cholesskey decomposition,

$$\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_0 \Leftrightarrow \boldsymbol{L}^{-1}\boldsymbol{\Sigma}_x\boldsymbol{L}^{-\top} = \boldsymbol{L}^{-1}\boldsymbol{\Sigma}_0\boldsymbol{L}^{-\top} \Leftrightarrow \boldsymbol{\Sigma}_x^* = \boldsymbol{I}. \tag{2.94}$$

Hence, the hypotheses in (2.93) can be written as

$$H_0 : \boldsymbol{\Sigma}_z = \boldsymbol{I} \quad \text{vs} \quad H_1 : \boldsymbol{\Sigma}_z \neq \boldsymbol{I} \tag{2.95}$$

using $\boldsymbol{Z} = \boldsymbol{L}^{-1}\boldsymbol{X}$. When $n \to \infty$, d is fixed and the random vector $\boldsymbol{X}$ follows multivariate normal, Nagao (1973) proposed the test statistics

$$\mathcal{T}_1 = \frac{1}{d}\mathrm{tr}(\boldsymbol{S} - \boldsymbol{I})^2$$

where $\boldsymbol{S}$ is the sample estimate of $\boldsymbol{\Sigma}_x^*$ and $\mathrm{tr}(\cdot)$ denotes the trace operator. The asymptotic distribution of $\mathcal{T}_1$ is given in (4.13) of Nagao (1973) and is a function of d and a few chi-square distributions with different degrees of freedom.

Ledoit and Wolf (2002) examined the behaviour of $\mathcal{T}_1$ when $d > n$ and $d/n \to c \in (0, \infty)$, proposed the modified statistic

$$\mathcal{T}_2 = \frac{1}{p}\mathrm{tr}[(\boldsymbol{S}-\boldsymbol{I})^2] - \frac{d}{n}\Big[\frac{1}{d}\mathrm{tr}(\boldsymbol{S})\Big]^2 + \frac{d}{n} \text{ with } \frac{nT_2 - d - 1}{2} \xrightarrow{D} \mathrm{N}(0,1)$$

and showed the test is robust (no normality assumption) against the condition $d/n \to c \in (0, \infty)$.

Later on, Srivastava (2005) proposed the test with test statistic

$$\mathcal{T}_3 = \frac{n}{2}(\hat{a}_2 - 2\hat{a}_1 + 1) \xrightarrow{D} N(0,1), \text{ where } \hat{a}_1 = \frac{1}{d}\mathrm{tr}\boldsymbol{S} \text{ and } \hat{a}_2 = \frac{n^2}{(n-1)(n+2)d}\Big[\mathrm{tr}\boldsymbol{S}^2 - \frac{1}{n}(\mathrm{tr}\boldsymbol{S})^2\Big].$$

Following the idea of Srivastava (2005), Fisher (2012) proposed another test for $d > n$

$$\mathcal{T}_4 = \frac{n}{\sqrt{8(c^2+12c+8)}}(\hat{a}_4 - 2\hat{a}_2 + 1) \xrightarrow{D} \mathrm{N}(0,1)$$

where

$$\hat{a}_4 = \frac{\gamma}{d}\Big[\mathrm{tr}\boldsymbol{S}^4 - \frac{4}{n}\mathrm{tr}\boldsymbol{S}^3\mathrm{tr}\boldsymbol{S} - \frac{2n^2+3n-6}{n(n^2+n+2)}(\mathrm{tr}\boldsymbol{S}^2)^2 + \frac{2(5n+6)}{n(n^2+n+2)}\mathrm{tr}\boldsymbol{S}^2(\mathrm{tr}\boldsymbol{S})^2 - \frac{5n+6}{n^2(n^2+n+2)}(\mathrm{tr}\boldsymbol{S})^4\Big]$$

and

$$\gamma = \frac{n^5(n^2+n+2)}{(n+1)(n+2)(n+4)(n+6)(n-1)(n-2)(n-3)}.$$

Dropping normality assumption and assuming only $d \to \infty$ as $n \to \infty$ and $\mathrm{tr}(\boldsymbol{\Sigma}^4) = o\big(\mathrm{tr}^2(\boldsymbol{\Sigma}^2)\big)$, Chen et al. (2010a) proposed

$$\mathcal{T}_5 = \frac{1}{d}C_{2,n} - \frac{2}{d}C_{1,n} + 1, \ \ \sigma_{2,n}^{-1}[d\mathcal{T}_5 - \mathrm{tr}(\boldsymbol{\Sigma}-\boldsymbol{I})^2] \xrightarrow{D} \mathrm{N}(0,1)$$

where

$$\begin{aligned} C_{2,n} &= Y_{2,n} - 2Y_{4,n} + Y_{5,n}, \\ Y_{2,n} &= \frac{1}{P_n^2}\sum_{i\neq j}(\boldsymbol{X}_i^\top \boldsymbol{X}_j)^2, \ \ Y_{4,n} = \frac{1}{P_n^3}\sum_{i,j,k}\boldsymbol{X}_i^\top \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{X}_k, \ \ Y_{5,n} = \frac{1}{P_n^4}\sum_{i,j,k,l}\boldsymbol{X}_i^\top \boldsymbol{X}_j \boldsymbol{X}_k^\top \boldsymbol{X}_l \\ P_n^r &= n!/(n-r)!. \end{aligned}$$

Under the condition of $d > n$ and non-normality, Srivastava et al. (2011) further proposed

$$\mathcal{T}_6 = \frac{n}{2}(\hat{a}_2 - 2\hat{a}_1 + 1), \ \ n^{-1}\mathcal{T}_6 \xrightarrow{D} \mathrm{N}(0,4)$$

which is the same as $\mathcal{T}_3$ under the normality condition.

Again under the non-normality condition, Ahmad and von Rosen (2015) proposed

$$\mathcal{T}_7 = \frac{1}{d}C_3 - \frac{2}{d}C_1 + 1, \ \ n\mathcal{T}_7 \xrightarrow{D} \mathrm{N}\Big(0, 4\frac{2}{c+1}\Big)$$

when $d/n \to c$. Table 5(a) reports the type I error rate (proportion of times rejecting H_0) of the test at a 5% significance level. Results show that the test proposed by Fisher (2012) is most robust (closest to 5%) but the test proposed by Srivastava (2005) is furthest away from 5%.

Compound test: we assume $\boldsymbol{X}_\tau, \boldsymbol{Y}_\tau \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_0)$ and apply the seven one sample tests to test if the true covariance matrix $\boldsymbol{\Sigma}_x$ for S & P 500 indices equals to $\boldsymbol{\Sigma}_0$ which is estimated by $\boldsymbol{\Sigma}_0 = \widehat{\boldsymbol{\Sigma}}_y$ using $\boldsymbol{y}_1$ in Table 3 and both sample covariance matrix and our proposed range covariance matrix estimators. Let $\widehat{\boldsymbol{\Sigma}}_y = \widehat{\boldsymbol{L}}_y \widehat{\boldsymbol{L}}_y^\top$, where $\widehat{\boldsymbol{L}}_y$ is the Cholesky decomposition of $\widehat{\boldsymbol{\Sigma}}_y$. Asymptotically, we have $\widehat{\boldsymbol{\Sigma}}_y \xrightarrow{D} \boldsymbol{\Sigma}_0$ and then $\boldsymbol{Z}_\tau = \widehat{\boldsymbol{L}}_y^{-1} \boldsymbol{X}_\tau \xrightarrow{D} \mathrm{N}(\boldsymbol{0}, \boldsymbol{I})$. Assuming $\boldsymbol{Z}_\tau \to \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_z)$, we test the hypotheses in (2.95).

Using $\boldsymbol{x}$ and $\boldsymbol{y}_1$, the two-sample test in Table 4 shows the lowest rejection or type I error rate. Results in Table 5(b) show that the proposed range-based estimator is better because the type I error rate is smaller than the sample covariance estimator among all except the test proposed by Ledoit and Wolf (2002).

TABLE 5. One-sample tests for covariance matrices

ρ	Ahmad (2015)	Chen (2010)	Fisher (2012)	LedoitWolf (2002)	Nagao (1973)	Srivastava (2005)	Srivastava (2011)
(a) Simulation study							
0.026	0.0009	0.1565	0.0624	0.0754	0.1319	0.9847	0.0998
0.052	0.0004	0.1472	0.0617	0.0722	0.1381	0.9836	0.0943
0.078	0.0005	0.1487	0.0641	0.0752	0.1414	0.9831	0.0956
0.104	0.0008	0.1513	0.0669	0.0739	0.1406	0.9832	0.098
0.130	0.0007	0.1535	0.0686	0.0756	0.1392	0.9844	0.0989
0.156	0.0007	0.1600	0.0709	0.0816	0.1440	0.9832	0.1082
0.182	0.0005	0.1517	0.0643	0.0722	0.1305	0.9854	0.0954
0.208	0.0003	0.1552	0.0674	0.0743	0.1405	0.9853	0.0986
0.234	0.0009	0.1565	0.0656	0.0765	0.1357	0.9834	0.1073
0.260	0.0005	0.1556	0.0655	0.0792	0.1443	0.986	0.1002
0.286	0.001	0.1579	0.0653	0.0745	0.1354	0.9845	0.0962
0.312	0.0004	0.1577	0.0672	0.0752	0.1380	0.9838	0.1011
0.338	0.0008	0.1524	0.0663	0.0766	0.1334	0.9855	0.0989
0.364	0.0009	0.1584	0.0662	0.0769	0.1368	0.9840	0.0997
0.390	0.0007	0.1506	0.0704	0.0761	0.139	0.9843	0.0959
0.416	0.0007	0.1519	0.0625	0.0709	0.1325	0.9856	0.0954
0.442	0.0014	0.1505	0.0667	0.0733	0.1386	0.9831	0.0956
0.468	0.0002	0.1528	0.0618	0.0729	0.1357	0.9851	0.0993
0.494	0.0011	0.1559	0.0691	0.0752	0.1353	0.9834	0.0971
0.52	0.0003	0.1565	0.0709	0.0804	0.1433	0.9850	0.1027
(b) EFT data							
Sample	0.2523	0.7901	0.7440	0.6059	0.7053	0.8785	0.8250
Proposed	0.0184	0.4217	0.3831	0.6703	0.3278	0.2891	0.6998

2.8. Summary

This chapter extends the PK measure in (1.17) for variances to covariance matrices in a multi-dimensional setting. Considering dimension $d = 2$, correlation estimators are considered because they can be easily converted to covariance estimators in (2.49) using the efficient PK measures.

We propose different estimators using a simple average and weighted averages of $s_k, k = 1, \ldots, 8$ in (2.50) with different power p to estimate ρ. The RMSE function $f_r(\rho)$ in (2.44) describing the relationship with ρ is again estimated using a quartic function. Using $\hat{\rho} = f_{m1}(s_k)$ in (2.39), estimators $\mathfrak{m} = 1, 2.2 - 2.5$, denote the sample correlation matrix in (1.24) and RMSE weighted mode estimator in (2.50) with $p = 0, 1, 2, 1/2$ respectively. The MSE are displayed in Figure 5 (right plot) across ρ for estimators $\mathfrak{m} = 1, 2.2 - 2.5$. Results show that the range-based estimator using simple average mode estimate ($\mathfrak{m} = 2.2$) is the best when ρ is moderate but when ρ is high, the range-based estimator using MSE weighted mode estimate ($\mathfrak{m} = 2.4$) is slightly better.

Next, we assess the impact of heteroskedastic variance on the performance of various estimators using a simulation study with CARR(1,1) model and Gamma error in (2.51). Results that the MSE is not inflated by heteroskedasticity. Moreover, we also assess by simulation study the impact of NPD transformation to tackle the non-positive definite issue when $d \leq 3$ on the performance of various estimators. With $d = 10$ and the one-band structure in (2.58), results displayed in Figure 9 show that the NPD transformed simple average mode estimates is the best for moderate correlation and NPD transformation improves the performance of the estimator.

Lastly, we apply two and one sample tests to compare the performance of the sample and range-based estimators under different scenarios. We modify the two-sample test originally proposed for two-sample covariance matrices from independent samples to adopt two range-based covariance matrices from match pairs. Results in Table 4 show that the range-based covariance matrices display higher power in terms of higher rejection rate to detect inconsistency than the sample covariance matrices under different scenarios. Moreover, the seven one-sample tests in Table 5 further confirm the lower type I error for the range-based covariance matrices for all except one test.

In summary, this chapter proposes efficient range-based correlation and covariance matrix estimators with higher power and lower type I errors. The choice is the NPD transformed simple average mode estimator when the general correlation level is moderate and the MSE weighted mode estimator with NPD transformation when the correlation is higher. In case of uncertainty, both estimators can be applied and the one providing the best results in terms of forecasting, for example, can be chosen.

Using the proposed estimators, a time series of covariance matrices can be evaluated to reveal the pattern of correlation dynamics across time and facilitate portfolio optimisation to hedge risk. However, such a time series is subject to random noise from the data. Hence, a covariance matrix model should be developed to eliminate the random noise and provide better modelling of the covariance/correlation dynamics. Our proposed range-based covariance and correlation estimators depend strongly on the data generation process in Section 2.4.2. We have explored the impact of heteroskedasticity in Section 2.5.1 but other characteristics such as drift and non-normality should also be explored in future research.

Appendices

2.A. Proof of Proposition 2.2.2

Proposition 2.A.1. *Assume that $P_{3,t,\tau} = P_{1,t,\tau}/P_{2,t,\tau}$ for all t, τ. Then we have*

$$-1 \leq \mathrm{cor_a}(r_{1,t}, r_{2,t}) \leq 1,$$

where $\mathrm{cor_a}(r_{1,t}, r_{2,t})$ is defined in (2.4)*.*

Proof. For day t we have

$$\frac{L_{1,t}}{H_{2,t}} \leq \frac{P_{1,t+\tau^*}}{P_{2,t+\tau^*}} \leq \frac{H_{1,t}}{L_{2,t}} \tag{2.A.1}$$

where $P_{i,t+\tau^*}$ is the price of asset i at time τ^* within day t. Thus

$$H_{3,t} \leq \frac{H_{1,t}}{L_{2,t}}, L_{3,t} \geq \frac{L_{1,t}}{H_{2,t}} \Rightarrow \frac{H_{3,t}}{L_{3,t}} \leq \frac{H_{1,t}H_{2,t}}{L_{1,t}L_{2,t}}, \tag{2.A.2}$$

which yields

$$h_{1,t} - l_{1,t} + (h_{2,t} - l_{2,t}) \geq h_{3,t} - l_{3,t} \geq 0 \tag{2.A.3}$$

Substitude 2.A.3 into (2.4), we have

$$\begin{aligned} \mathrm{cor_a}(r_{1,t}, r_{2,t}) &\geq \frac{\sum_{\tau=1}^{n}(h_{1,t,\tau} - l_{1,t,\tau})^2 + (h_{2,t,\tau} - l_{2,t,\tau})^2 - (h_{1,t} - l_{1,t} + (h_{2,t} - l_{2,t}))^2}{2\sqrt{\sum_{\tau=1}^{n}(h_{1,t,\tau} - l_{1,t,\tau})^2 \times \sum_{\tau=1}^{n}(h_{2,t,\tau} - l_{2,t,\tau})^2}} \\ &= \frac{-2\sum_{\tau=1}^{n}(h_{1,t,\tau} - l_{1,t,\tau})(h_{1,t,\tau} - l_{1,t,\tau})}{2\sqrt{\sum_{\tau=1}^{n}(h_{1,t,\tau} - l_{1,t,\tau})^2 \times \sum_{\tau=1}^{n}(h_{2,t,\tau} - l_{2,t,\tau})^2}} \geq -1, \end{aligned} \tag{2.A.4}$$

by Cauchy–Schwarz inequality. Furthermore, for any $0 < \tau^*, \tau^{*\prime} < 1$ we have

$$\frac{H_{3,t}}{L_{3,t}} \geq \frac{P_{3,t+\tau^*}}{P_{3,t+\tau^{*\prime}}} = \frac{P_{1,t+\tau^*}P_{2,t+\tau^{*\prime}}}{P_{2,t+\tau}P_{1,t+\tau^{*\prime}}} \geq \frac{P_{1,t+\tau^*}L_{2,t}}{H_{2,t}P_{1,t+\tau^{*\prime}}} \geq \frac{L_{2,t}}{H_{2,t}} \max\left(\frac{P_{1,t+\tau^*}}{P_{1,\tau^{*\prime}}}\right) = \frac{H_1L_2}{L_1H_2} \tag{2.A.5}$$

$$\frac{H_{3,t}}{L_{3,t}} \geq \frac{P_{1,t+\tau^*}P_{2,t+\tau^{*\prime}}}{P_{2,t+\tau^*}P_{1,t+\tau^{*\prime}}} \geq \frac{L_{1,t}P_{2,t+\tau^{*\prime}}}{P_{2,t+\tau^*}H_{1,t}} \geq \max\frac{L_{1,t}}{H_{1,t}}\left(\frac{P_{2,t+\tau^{*\prime}}}{P_{2,t+\tau^*}}\right) = \frac{L_1H_2}{L_2H_1}. \tag{2.A.6}$$

Hence again by Cauchy–Schwarz inequality

$$(h_{3,t} - l_{3,t})^2 \geq (h_{1,t} - l_{1,t} - (h_{2,t} - l_{2,t}))^2 \tag{2.A.7}$$

$$\mathrm{cor_a}(r_{1,t}, r_{2,t}) \leq \frac{2\sum_{\tau=1}^{n}(h_{1,t,\tau} - l_{1,t})(h_{1,t,\tau} - l_{1,t,\tau})}{2\sqrt{\sum_{\tau=1}^{n}(h_{1,t,\tau} - l_{1,t,\tau})^2 \times \sum_{\tau=1}^{n}(h_{2,t,\tau} - l_{2,t,\tau})^2}} \leq 1. \tag{2.A.8}$$

□

2.B. Proof of Proposition 2.3.1

Proposition 2.B.1. *Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \boldsymbol{\Theta} = \boldsymbol{\Theta}_1 \times \boldsymbol{\Theta}_2$. Let $D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1)$ be any estimator for $\boldsymbol{\theta}_2$ given $\boldsymbol{\theta}_1$ and $L(\boldsymbol{D(\kappa; \theta_1), \theta_2})$ be a loss function independent of $\boldsymbol{\theta}_1$ such that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $L(\cdot; \boldsymbol{\theta}_2)$ is a convex function. Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a monotonic transformation, and $T' : \boldsymbol{\Theta}_1 \to \boldsymbol{\Theta}_1$ be a invertable transformation. Suppose that, for all $\boldsymbol{\theta}_2 \in \boldsymbol{\Theta}_2$ and all $\boldsymbol{\kappa}$ in the support of $f(\cdot; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$,*

$$|\det \boldsymbol{J}_T| f(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = f(\boldsymbol{\kappa}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2).$$

Consider the risk given by:

$$R(D(\cdot); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \int_{\mathbb{R}^d} L(D(\cdot), \boldsymbol{\theta}_2) f(\boldsymbol{\kappa}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\kappa},$$

as the expected loss of estimator D under the distribution $f(\cdot; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Then at least one of the following holds:

1. $R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > R(D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$,
2. $R(D(\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > R(D(T^{-1}\boldsymbol{\kappa}; \boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$,
3. $R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \geq R(wD(\boldsymbol{\kappa}; \boldsymbol{\theta}_1) + (1-w)D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \quad \forall w \in [0, 1]$.

Proof. Suppose that neither statement 1 and 2 holds. Then one has

$$\begin{aligned}
R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &\leq R(D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&= \int_{\mathbb{R}^d} L(D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1), \boldsymbol{\theta}_2) f(\boldsymbol{\kappa}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\kappa} \\
&= \int_{\mathbb{R}^d} L(D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1), \boldsymbol{\theta}_2) |\det \boldsymbol{J}_T| f(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\kappa} \\
&= \int_{\mathbb{R}^d} L(D(\boldsymbol{\kappa}'; T'\boldsymbol{\theta}_1), \boldsymbol{\theta}_2) f(\boldsymbol{\kappa}'; T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\kappa}' \\
&= R(D(\boldsymbol{\kappa} | T'\boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2). \qquad \text{(2.B.1)}
\end{aligned}$$

by taking transform of variable $\boldsymbol{\kappa}' = T\boldsymbol{\kappa}$. Furthermore, we also have

$$\begin{aligned}
R(D(\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &\leq R(D(T^{-1}\boldsymbol{\kappa}; \boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&= \int_{\mathbb{R}^d} L(D(T^{-1}\boldsymbol{\kappa}; \boldsymbol{\theta}_1), \boldsymbol{\theta}_2)) f(\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\kappa} \\
&= \int_{\mathbb{R}^d} L(D(\boldsymbol{\kappa}'; \boldsymbol{\theta}_1), \boldsymbol{\theta}_2) |\det \boldsymbol{J}_T| f(T\boldsymbol{\kappa}'; T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\kappa}' \\
&= \int_{\mathbb{R}^d} L(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1), \boldsymbol{\theta}_2) f(\boldsymbol{\kappa}'; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\kappa}' \\
&= R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \qquad \text{(2.B.2)}
\end{aligned}$$

by taking transform of variable $\boldsymbol{\kappa} = T\boldsymbol{\kappa}'$. Hence by 2.B.1 and 2.B.2 we have

$$R(D(\boldsymbol{\kappa}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = R(D(\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); T'\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = R(D(T\boldsymbol{\kappa}; T'\boldsymbol{\theta}_1); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2).$$

Then by convexity of $L(\cdot, \boldsymbol{\theta}_2)$ we have

$$\begin{aligned} R(wD(\boldsymbol{\kappa};\boldsymbol{\theta}_1) &+ (1-w)D(T\boldsymbol{\kappa};T'\boldsymbol{\theta}_1);\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \\ &= \int_{\mathbb{R}^d} L(wD(\boldsymbol{\kappa};\boldsymbol{\theta}_1)+(1-w)D(T\boldsymbol{\kappa};T'\boldsymbol{\theta}_1),\boldsymbol{\theta}_2)f(\boldsymbol{\kappa};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)d\boldsymbol{\kappa} \\ &\leq \int_{\mathbb{R}^d} (wL(D(\boldsymbol{\kappa};\boldsymbol{\theta}_1),\boldsymbol{\theta}_2)+(1-w)L(D(T\boldsymbol{\kappa};T'\boldsymbol{\theta}_1),\boldsymbol{\theta}_2))f(\boldsymbol{\kappa};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)d\boldsymbol{\kappa} \\ &= wR(D(\boldsymbol{\kappa}|\boldsymbol{\theta}_1);\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)+(1-w)R(D(T\boldsymbol{\kappa}|T'\boldsymbol{\theta}_1);\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \\ &= R(D(\boldsymbol{\kappa};\boldsymbol{\theta}_1);\boldsymbol{\theta}_1,\boldsymbol{\theta}_2), \end{aligned}$$

which contradicts our assumption. □

2.C. Proof of Proposition 2.3.6

Proposition 2.C.1. *Let $S_{1,t}, S_{2,t}$ be two geometric Brownian motions defined as*

$$d\log S_{1,t} = \sigma_1 dW_{1,t}, \quad d\log S_{2,t} = \sigma_2(\rho dW_{1,t} + \sqrt{1-\rho^2}dW_{2,t}),$$

where $W_{1,t}, W_{2,t}$ are independent Wiener processes. Let

$$h_{i,t} = \max_{0\leq\tau^*\leq 1} r_{i,t+\tau^*}, \quad l_{i,t} = \min_{0\leq\tau^*\leq 1} r_{i,t+\tau^*}, \quad c_{i,t} = r_{i,t}, \quad i=1,2, \tag{2.C.1}$$

where

$$r_{i,t+\tau^*} = \log(S_{i,t+\tau^*}/S_{i,t+0}),\ i=1,2,\ \tau^*\in(0,1). \tag{2.C.2}$$

Define $\boldsymbol{\theta} = (\rho, \sigma_1, \sigma_2)$ and $\boldsymbol{\kappa}_t = (h_{1,t}, l_{1,t}, c_{1,t}, h_{2,t}, l_{2,t}, c_{2,t})$. Let $f(\cdot;\boldsymbol{\theta})$ be the density function of $\boldsymbol{\kappa}$. Then we have

$$f(T_2(\boldsymbol{\kappa}_t);\boldsymbol{\theta}) = f(\boldsymbol{\kappa}_t;\boldsymbol{\theta}) \tag{2.C.3}$$

where $T : \boldsymbol{R}^6 \to \boldsymbol{R}^6$ is given by

$$T_2(\boldsymbol{\kappa}) = (h_{1,t}-c_{1,t}, l_{1,t}-c_{1,t}, -c_{1,t}, h_{2,t}-c_{2,t}, l_{2,t}-c_{2,t}, -c_{2,t})^\top. \tag{2.C.4}$$

Proof. Let $r_{i,t+\tau^*} = \log(S_{i,t+\tau^*}/S_{i,t+0}),\ i=1,2,\ \tau^*\in(0,1)$. Then we have

$$h_{i,t} = \max_{0\leq\tau^*\leq 1} r_{i,t+\tau^*}, \quad l_{i,t} = \min_{0\leq\tau^*\leq 1} r_{i,t+\tau^*}, \quad c_{i,t} = r_{i,t}, \quad i=1,2 \tag{2.C.5}$$

by definition. Now we define $\tilde{r}_{i,t+\tau^*} = r_{i,t+1-\tau^*} - r_{i,t+1}$. Then we have

$$\begin{aligned} h_{i,t} - c_{i,t} &= \max_{0\leq\tau^*\leq 1} r_{i,t+\tau^*} - r_{i,t+1} = \max_{0\leq\tau^*\leq 1} \tilde{r}_{i,t+\tau^*}, \\ l_{i,t} - c_{i,t} &= \min_{0\leq\tau^*\leq 1} r_{i,t+\tau^*} - r_{i,t+1} = \min_{0\leq t^*\leq 1} \tilde{r}_{i,t+\tau^*}, \\ -c_{i,t} &= -r_{i,t+1} = \tilde{r}_{i,t+\tau^*}, \quad i=1,2. \end{aligned}$$

Hence,

$$T_2(\boldsymbol{\kappa}_t) = (\tilde{h}_{1,t}, \tilde{l}_{1,t}, \tilde{c}_{1,t}, \tilde{h}_{2,t}, \tilde{l}_{2,t}, \tilde{c}_{2,t}) = \tilde{\boldsymbol{\kappa}}_t \tag{2.C.6}$$

where $\tilde{h}_{i,t}, \tilde{l}_{i,t}, \tilde{c}_{i,t}$ are defined in similar fashion as (2.C.5) replacing $r_{i,t+\tau^*}$ with $\tilde{r}_{i,t+\tau^*}$.
Let $\tilde{W}_{i,t+\tau^*} = W_{i,t+1-\tau^*} - W_{i,t+1}, i=1,2$. Then we have

$$\tilde{r}_{1,t+\tau^*} = \sigma_1(W_{1,t+1-\tau^*} - W_{1,t+1}) = \sigma_1\tilde{W}_{1,t+\tau^*},\ \ \tilde{r}_{2,t+\tau^*} = \sigma_2(\rho\tilde{W}_{1,t+\tau^*} + \sqrt{1-\rho^2}\tilde{W}_{2,t+\tau^*}). \tag{2.C.7}$$

Clearly $\tilde{W}_{1,t+\tau^*}$ and $\tilde{W}_{2,t+\tau^*}$ are independent continuous Gaussian process. Furthermore for any $s^* < \tau^*$, we have

$$\begin{aligned}
\mathrm{E}(\tilde{W}_{i,t+\tau^*}) &= \mathrm{E}(W_{i,t+1-\tau^*}) - \mathrm{E}(W_{i,t+1}) = 0, && \text{(2.C.8)}\\
\mathrm{E}(\tilde{W}_{i,t+s^*}\tilde{W}_{1,t+\tau^*}) &= \mathrm{E}(W_{i,t+1-s^*} - W_{i,t+1})(W_{i,t+1-t^*} - W_{i,t+1})\\
&= \mathrm{E}(W_{i,t+1-s^*}W_{i,t+1-\tau^*}) - \mathrm{E}(W_{i,t+1-s^*}W_{i,t+1}) - \mathrm{E}(W_{i,t+1}W_{i,t+1-\tau^*}) + \mathrm{E}(W_{i,t+1}W_{i,t+1})\\
&= 1 - \tau^* - (1 - s^*) - (1 - \tau^*) + 1\\
&= s^*. && \text{(2.C.9)}
\end{aligned}$$

Hence $\tilde{W}_{1,t+\tau^*}$ and $\tilde{W}_{2,t+\tau^*}$ are Weiner processes. Hence $\tilde{\boldsymbol{\kappa}}_t$ should have the same distribution as $\boldsymbol{\kappa}_t$, that is

$$\int_A f'_{\boldsymbol{\theta}}(\boldsymbol{\kappa}_t;\boldsymbol{\theta})d\boldsymbol{\kappa}_t = \int_A f'_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\kappa}}_t;\boldsymbol{\theta})d\tilde{\boldsymbol{\kappa}}_t = \int_A f'_{\boldsymbol{\theta}}(T_2(\boldsymbol{\kappa}_t);\boldsymbol{\theta})|\det \boldsymbol{J}_{T_2}|d\boldsymbol{\kappa}_t \tag{2.C.10}$$

for any set A, which yields

$$f'_{\boldsymbol{\theta}}(\boldsymbol{\kappa}_t|\boldsymbol{\theta}) = |\det \boldsymbol{J}_{T_2}| f'_{\boldsymbol{\theta}}\big(T_2(\boldsymbol{\kappa}_t)|\boldsymbol{\theta}\big). \tag{2.C.11}$$

Furthermore, we have

$$\boldsymbol{J}_{T_2} = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0\\ 0 & 1 & -1 & 0 & 0 & 0\\ 0 & 0 & -1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & -1\\ 0 & 0 & 0 & 0 & 1 & -1\\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}, \quad \det \boldsymbol{J}_{T_2} = 1. \tag{2.C.12}$$

Hence (2.C.3) follows.

The invariance of T_2 follows immediately as

$$\begin{aligned}
f_{\boldsymbol{\theta}}\big(T_2(\boldsymbol{h}_1,\boldsymbol{l}_1,\boldsymbol{c}_1,\boldsymbol{h}_2,\boldsymbol{l}_2,\boldsymbol{c}_2);\boldsymbol{\theta}\big) &= \prod_{t=1}^{N} f'_{\boldsymbol{\theta}}\big(T_2(h_{1,t},l_{1,t},c_{1,t},h_{2,t},l_{2,t},c_{2,t});\boldsymbol{\theta}\big)\\
&= \prod_{t=1}^{N} f'_{\boldsymbol{\theta}}\big((h_{1,t},l_{1,t},c_{1,t},h_{2,t},l_{2,t},c_{2,t});\boldsymbol{\theta}\big)\\
&= f_{\boldsymbol{\theta}}\big((\boldsymbol{h}_1,\boldsymbol{l}_1,\boldsymbol{c}_1,\boldsymbol{h}_2,\boldsymbol{l}_2,\boldsymbol{c}_2);\boldsymbol{\theta}\big). && \text{(2.C.13)}
\end{aligned}$$

□

2.D. Proof of Lemma 2.D.3

In this appendix, we first present Lemma 2.D.1 from Garman and Klass (1980), which demonstrates that for any probability invariance transformation T defined in (2.B.1) and any estimator $D(\cdot)$, the expected risk of the average operator $A_K(D(\cdot))$, as defined in (2.D.2), is no greater than the expected risk of $D(\cdot)$. We then extend the results in Lemma 2.D.1 to any measure-preserving transformation even without probability invariance such as transformation T_3 in (2.12). However, the proof from Garman and Klass (1980) and the Lemma 2.D.3 is meant to be general which can be applied to any matrix $\boldsymbol{X}$ (Garman and Klass, 1980) but we adopt the notation $\boldsymbol{\kappa}$ to be consistent with the application in Section 2.3 and Appendix 2.C.

Now we would like to extend the result to a case in which invariance condition (2.D.7) is slightly modified.

Lemma 2.D.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\boldsymbol{\kappa} : \Omega \to \mathbb{R}^d$ be a d-dimension random vector, whose joint density $f_{\boldsymbol{\kappa}}(\cdot; \boldsymbol{\theta})$ depends on an parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a measure-preserving transformation such that, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{\kappa}$ in the support of $f_{\boldsymbol{\theta}}$,*

$$f_{\boldsymbol{\theta}}(T(\boldsymbol{\kappa})) = f_{\boldsymbol{\theta}}(\boldsymbol{\kappa}). \tag{2.D.1}$$

Let $D(\boldsymbol{\kappa})$ be any estimator and $L(\cdot|\boldsymbol{\theta})$ be an arbitrary loss function such that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $L(\cdot|\boldsymbol{\theta})$ is a convex function. Define $T^k = T(T^{k-1})$ and T^0 be the identity operator. Let A'_K be an average operator defined as

$$A_K(D(\boldsymbol{\kappa})) = \frac{1}{K}\sum_{k=1}^{K} D(T^{k-1}(\boldsymbol{\kappa})). \tag{2.D.2}$$

Then for any convex loss function $L(\cdot)$, we have

$$\mathrm{E}_{\boldsymbol{\theta}}(L(A_K(D(\boldsymbol{\kappa}))|\boldsymbol{\theta})) \leq \mathrm{E}_{\boldsymbol{\theta}}(L(D(\boldsymbol{\kappa})|\boldsymbol{\theta})) \quad \forall \boldsymbol{\theta}. \tag{2.D.3}$$

Proof. See Garman and Klass (1980). □

Remark 2.D.2. Now let us supose that there exist M, such that

$$A_M(D(T(\boldsymbol{\kappa}))) = A_M(D(\boldsymbol{\kappa})). \tag{2.D.4}$$

Then, for any $K > 0$, we have

$$\begin{aligned} A_M(A_K(D(\boldsymbol{\kappa}))) &= \frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{K}\sum_{k=1}^{K} D\big(T^{k+m-2}(\boldsymbol{\kappa})\big)\right] \\ &= \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{M}\sum_{m=1}^{M} D\big(T^{k-1}(T^{m-1}(\boldsymbol{\kappa}))\big)\right] \\ &= \frac{1}{K}\sum_{k=1}^{K} A_M(D(T^{k-1}(\boldsymbol{\kappa}))) \\ &= A_M(D(\boldsymbol{\kappa})). \end{aligned} \tag{2.D.5}$$

The result suggests the best estimator among the class of averaging estimators generated by $D(\boldsymbol{\kappa})$ can be characterised by (2.D.6). Note that if $T^M = I,\ M > 0$, then

$$\begin{aligned} A_M(D(T(\boldsymbol{\kappa}))) &= \frac{1}{M}\sum_{k=1}^{M} D(T^k(\boldsymbol{\kappa})) \\ &= \frac{1}{M}\left[\sum_{k=2}^{M} D(T^{k-1}(\boldsymbol{\kappa})) + D(T^M(\boldsymbol{\kappa}))\right] \\ &= \frac{1}{M}\left[\sum_{k=1}^{M} D(T^{k-1}(\boldsymbol{\kappa}))\right] \\ &= A_M(D(\boldsymbol{\kappa})), \end{aligned} \tag{2.D.6}$$

that is, if M is the period of transformation, then $A_M(D(\cdot))$ minimize the expected risk amount the class $\{A_K(D(\cdot))\}$.

Lemma 2.D.3. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\boldsymbol{\kappa} : \Omega \to \mathbb{R}^d$ be a d-dimension random vector, whose joint density $f_{\boldsymbol{\kappa}}(\cdot; \boldsymbol{\theta})$ depends on an parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a measure-preserving transformation such that, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{\kappa}$ in the support of $f_{\boldsymbol{\theta}}$,*

$$f_{\boldsymbol{\theta}}(T(\boldsymbol{\kappa})) = f_{-\boldsymbol{\theta}}(\boldsymbol{\kappa}). \tag{2.D.7}$$

Let $D(\boldsymbol{\kappa})$ be any estimator and $L(\cdot|\boldsymbol{\theta})$ be an arbitrary loss function such that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $L(\cdot|\boldsymbol{\theta})$ is a convex function. Define $T^k = T(T^{k-1})$ and T^0 be the identity operator. Let A'_K be an alternating average operator defined as

$$A'_K(D(\boldsymbol{\kappa})) = \frac{1}{K}\sum_{k=1}^{K}(-1)^{k-1}D(T^{k-1}(\boldsymbol{\kappa})). \tag{2.D.8}$$

Then for a symmetric prior $g_{\boldsymbol{\theta}}(\cdot)$ for $\boldsymbol{\theta}$, we have

$$B_{g_{\boldsymbol{\theta}}}(A'_k(D(\boldsymbol{\kappa}))) \le B_{g_{\boldsymbol{\theta}}}(D(\boldsymbol{\kappa})) \tag{2.D.9}$$

where

$$B_{g_{\boldsymbol{\theta}}}(D(\boldsymbol{\kappa})) = \int_{\boldsymbol{\Theta}} \mathrm{E}_{\boldsymbol{\theta}}(L(D(\boldsymbol{\kappa})|\boldsymbol{\theta}))\, g_{\boldsymbol{\theta}}(\boldsymbol{\theta})\, d\boldsymbol{\theta} \tag{2.D.10}$$

is the Bayes risk.

Proof. We first note that

$$f_{\boldsymbol{\theta}}(T^2(\boldsymbol{\kappa})) = f_{-\theta}(T(\boldsymbol{\kappa})) = f_{\theta}(\boldsymbol{\kappa}). \tag{2.D.11}$$

By convexity of $L(\cdot|\boldsymbol{\theta})$, we have

$$L\big(A'_K(D(\boldsymbol{\kappa}))|\boldsymbol{\theta}\big) \le \frac{1}{K}\sum_{k=1}^{K} L\big((-1)^{k-1}D(T^{k-1}(\boldsymbol{\kappa}))|\boldsymbol{\theta}\big). \tag{2.D.12}$$

Thus by linearity of expectation, we have

$$B_{g_\rho}\big(A'_K(D(\boldsymbol{\kappa}))\big) \le \frac{1}{K}\sum_{k=1}^{K} B_{g_\rho}\big((-1)^{k-1}D(T^{k-1}(\boldsymbol{\kappa})\big). \tag{2.D.13}$$

If k is an odd number, then

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{\theta}}\big(L((-1)^{k-1}D(T^{k-1}(\boldsymbol{\kappa}))|\boldsymbol{\theta})\big) &= \int_{\mathbb{R}^d} L\big(D(T^{k-1}(\boldsymbol{\kappa}))|\boldsymbol{\theta}\big)\, f_{\boldsymbol{\theta}}(\boldsymbol{\kappa})\, d\boldsymbol{\kappa} \\
&= \int_{\mathbb{R}^d} L\big(D(T^{k-1}(\boldsymbol{\kappa}))|\boldsymbol{\theta}\big)\, f_{\boldsymbol{\theta}}(T^{k-1}(\boldsymbol{\kappa}))\, d\boldsymbol{\kappa} \qquad (2.D.14)
\end{aligned}$$

Let $\boldsymbol{\varkappa} = T^{k-1}(\boldsymbol{\kappa})$. Note that as T is measure-preserving, we have $|\det \boldsymbol{J}_{T^{k-1}}| = 1$. Hence we have

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{\theta}}\big(L((-1)^{k-1}D(T^{k-1}(\boldsymbol{\kappa}))|\boldsymbol{\theta})\big) &= \int_{\mathbb{R}^d} L(D(\boldsymbol{\varkappa})|\boldsymbol{\theta})\, f_{\boldsymbol{\theta}}(\boldsymbol{\varkappa})\, |\det \boldsymbol{J}_{T^{k-1}}|^{-1}\, d\boldsymbol{\varkappa} \\
&= \int_{\mathbb{R}^d} L(D(\boldsymbol{\varkappa})|\boldsymbol{\theta})\, f_{\boldsymbol{\theta}}(\boldsymbol{\varkappa})\, d\boldsymbol{\varkappa}, \\
&= \mathrm{E}_{\boldsymbol{\theta}}\big(L(D(\boldsymbol{\kappa})|\boldsymbol{\theta})\big) \qquad (2.D.15) \\
B_{g_{\boldsymbol{\theta}}}\big((-1)^{k-1}D(T^{k-1}(\boldsymbol{\kappa}))\big) &= B_{g_{\boldsymbol{\theta}}}(D(\boldsymbol{\kappa})). \qquad (2.D.16)
\end{aligned}$$

This follows exactly from Garman and Klass (1980) who proved the invariant transformation. If K is an even number, then follows the similar argument, we have

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{\theta}}\big(L((-1)^{k-1}D(T^{k-1}(\boldsymbol{\kappa}))|\boldsymbol{\theta})\big) &= \int_{\mathbb{R}^d} L(D(T^{k-1}(\boldsymbol{\kappa}))|\boldsymbol{\theta})\; f_{-\boldsymbol{\theta}}(T^{k-1}(\boldsymbol{\kappa}))\; d\boldsymbol{\kappa} \\
&= \mathrm{E}_{-\boldsymbol{\theta}}(L(D(\boldsymbol{\kappa})|-\boldsymbol{\theta})) \qquad (2.D.17) \\
B_{g_{\boldsymbol{\theta}}}((-1)^{i-1}D(T^{i-1}(\boldsymbol{\kappa}))) &= \int_{\boldsymbol{\Theta}} \mathrm{E}_{-\boldsymbol{\theta}}(L(D(\boldsymbol{\kappa})|-\boldsymbol{\theta}))\; g_{\boldsymbol{\theta}}(\boldsymbol{\theta})\; d\boldsymbol{\theta}. \qquad (2.D.18)
\end{aligned}$$

Let $\boldsymbol{\vartheta} = -\boldsymbol{\theta}$. By the symmetry of $g_\rho(\cdot)$ we have

$$\begin{aligned}
B_{g_{\boldsymbol{\theta}}}((-1)^{i-1}D(T^{i-1}(\boldsymbol{\kappa}))) &= \int_{\boldsymbol{\Theta}} \mathrm{E}_{\boldsymbol{\vartheta}}(L(D(\boldsymbol{\kappa})|\boldsymbol{\vartheta}))\; g_{\boldsymbol{\theta}}(-\boldsymbol{\vartheta})\; |\det \boldsymbol{J}_{-I}|\; d\boldsymbol{\vartheta} \\
&= \int_{\boldsymbol{\Theta}} \mathrm{E}_{\boldsymbol{\vartheta}}(L(D(\boldsymbol{\kappa})|\boldsymbol{\vartheta}))\; g_{\boldsymbol{\theta}}(\boldsymbol{\vartheta})\; d\boldsymbol{\vartheta}. \qquad (2.D.19) \\
&= B_{g_{\boldsymbol{\theta}}}(D(\boldsymbol{\kappa})). \qquad (2.D.20)
\end{aligned}$$

By (2.D.13), (2.D.16), (2.D.20),

$$B_{g_{\boldsymbol{\theta}}}(A'_K(D(\boldsymbol{\kappa}))) \leq \frac{1}{K}\sum_{k=1}^{K} B_{g_{\boldsymbol{\theta}}}((D(\boldsymbol{\kappa})) = B_{g_{\boldsymbol{\theta}}}((D(\boldsymbol{\kappa})) \qquad (2.D.21)$$

as required. □

Remark 2.D.4. Similar to the case of invariant transform, if $T^M = I$ and M is an even number, we have

$$\begin{aligned}
A'_M(D(T(\boldsymbol{\kappa}))) &= \frac{1}{M}\sum_{m=1}^{M}(-1)^{m-1}D(T^m(\boldsymbol{\kappa})) \\
&= -\frac{1}{M}\left[\sum_{m=2}^{M}(-1)^{m-1}D(T^{k-1}(\boldsymbol{\kappa})) + (-1)^M D(T^M(\boldsymbol{\kappa}))\right] \\
&= -\frac{1}{M}\sum_{k=1}^{M} D(T^{k-1}(\boldsymbol{\kappa})) \\
&= -A'_M(D(\boldsymbol{\kappa})), \qquad (2.D.22) \\
A'_M(A_k(D(\boldsymbol{\kappa}))) &= \frac{1}{M}\sum_{m=1}^{M}(-1)^{m-1}\frac{1}{K}\left[\sum_{k=1}^{K}(-1)^{k-1}D(T^{m+k-2}(\boldsymbol{\kappa}))\right] \\
&= \frac{1}{K}\sum_{k=1}^{K}(-1)^{k-1}\frac{1}{M}\left[\sum_{m=1}^{M}(-1)^{m-1}D(T^{m-1}(T^{k-1}(\boldsymbol{\kappa})))\right] \\
&= \frac{1}{K}\sum_{k=1}^{K}(-1)^{k-1}A_M(D(T^{M-1}(\boldsymbol{\kappa}))) \\
&= \frac{1}{K}\sum_{k=1}^{K}(-1)^{2k-2}A_M(D(\boldsymbol{\kappa})) \qquad (2.D.23) \\
&= A'_M(D(\boldsymbol{\kappa})). \qquad (2.D.24)
\end{aligned}$$

That is the best estimator among the class of averaging estimators generated by $D(\boldsymbol{\kappa})$ will be $A_M(D(\boldsymbol{\kappa}))$. What would happen if M is an odd number? We have

$$\begin{aligned} A'_{2M}(D(T(\boldsymbol{\kappa}))) &= \frac{1}{2M}\sum_{m=1}^{2M}(-1)^{m-1}D(T^{m-1}(\boldsymbol{\kappa})) \\ &= \frac{1}{2M}\sum_{m=1}^{M}(-1)^{m-1}D(T^{m-1}(\boldsymbol{\kappa})) + \frac{1}{2M}\sum_{m=1}^{M}(-1)^{m-1+M}D(T^{m-1}(\boldsymbol{\kappa})) \\ &= \frac{1}{2M}\sum_{m=1}^{M}(-1)^{m-1}D(T^{m-1}(\boldsymbol{\kappa})) - \frac{1}{2M}\sum_{m=1}^{M}(-1)^{m-1}D(T^{m-1}(\boldsymbol{\kappa})) \\ &= 0. \end{aligned} \tag{2.D.25}$$

That is all estimators will have higher Bayes risk compared to trivial decisions $D(\boldsymbol{\kappa}) = 0$. By (2.D.7), we have

$$f_{\boldsymbol{\theta}}(\boldsymbol{\kappa}) = f_{\boldsymbol{\theta}}(T^M(\boldsymbol{\kappa})) = f_{(-1)^M\boldsymbol{\theta}}(\boldsymbol{\kappa}) = f_{-\boldsymbol{\theta}}(\boldsymbol{\kappa}), \tag{2.D.26}$$

that is, the posterior distribution will always be symmetric. We can hardly get an estimator for $\boldsymbol{\theta}$. From here we can also see why an estimator D consisting of $\boldsymbol{h}_1 - \boldsymbol{l}_1, \boldsymbol{h}_2 - \boldsymbol{l}_2$ only will fail, as $D(T_2(\boldsymbol{\kappa})) = D(\boldsymbol{\kappa})$.

2.E. Property of Poisson process

Proposition 2.E.1. *Let N_t be a Poisson process with rate λ. Define*

$$\mathsf{T}_0 = 0,\ \mathsf{T}_n = \inf\{s > 0 : N_s = n\},\ \ n = 1, 2, ... \tag{2.E.1}$$

Then given $N_t = k$, $(\mathsf{T}_1, \mathsf{T}_2, ..., \mathsf{T}_k)|N_t = k$ has the same distribution as $(tU_{(1)}, tU_{(2)}, ..., tU_{(k)})$, where $(U_{(1)}, U_{(2)}, ..., U_{(k)})$ are the order statistics of k i.i.d random variable uniformly distributed over $[0, 1]$.

Proposition 2.E.2. *Let $E_1, ..., E_{k+1} \overset{i.i.d}{\sim} \mathrm{Exp}(\lambda)$. Let $(U_{(1)}, U_{(2)}, ..., U_{(k)})$ be the order statistics of k i.i.d random variable uniform distributed over $[0, 1]$. Then we have*

$$\left(\frac{E_1}{\sum_{i=1}^{k+1}E_i}, \frac{E_1 + E_2}{\sum_{i=1}^{k+1}E_i}, ..., \frac{\sum_{i=1}^{k}E_i}{\sum_{i=1}^{k+1}E_i}\right) \overset{d}{=} (U_{(1)}, U_{(2)}, ..., U_{(k)}). \tag{2.E.2}$$

Proposition 2.E.3. *Let $N_t, t \geq 0$, be a Poisson process with rate λ. Suppose that each arrival of N_t can be independently classified into 'type I' or 'type II' arrival with probability p and $1 - p$, respectively. Denote by $N_t(1)$ or $N_t(2)$ the number of 'type I' or 'type II' arrivals, respectively. Then $N_t = N_t(1) + N_t(2)$, where*

1. *$N_t(1)$ is a Poisson process with rate λp ;*
2. *$N_t(2)$ is a Poisson process with rate $\lambda(1 - p)$;*
3. *$N_t(1)$ and $N_t(2)$ are independent.*

2.F. Summary of polynomial regression of f_μ

Coefficients:	Estimate	Std. Error	t-value	P-value
(Intercept)	−4.091E-17	1.782E-04	0.000	1.0000
poly(m, 10)1	4.857E+00	1.563E-03	3106.857	< 2e-16
poly(m, 10)2	−4.238E-01	1.563E-03	−271.084	< 2e-16
poly(m, 10)3	2.522E-02	1.563E-03	16.134	< 2e-16
poly(m, 10)4	−1.059E-02	1.563E-03	−6.777	4.03E-09
poly(m, 10)5	3.057E-03	1.563E-03	1.955	0.0548
poly(m, 10)6	−1.626E-03	1.563E-03	−1.040	0.3020
poly(m, 10)7	−6.913E-05	1.563E-03	−0.044	0.9649
poly(m, 10)8	−7.852E-04	1.563E-03	−0.502	0.6172
poly(m, 10)9	1.19E-03	1.563E-03	0.761	0.4494
poly(m, 10)10	2.239E-04	1.563E-03	0.143	0.8866

2.G. Asymptotic property of MSE

Assume that the true correlation between asset 1 and asset 2 is ρ. By construction we have

$$\mathrm{E}(s_{1,t}) = f_\mu^{-1}(\rho). \tag{2.G.1}$$

While the joint distribution of $(\boldsymbol{h}_1, \boldsymbol{h}_2)$ is unknown, Muirhead and Waternaux (1980) suggested that as long as the joint distribution of $(\boldsymbol{h}_1, \boldsymbol{h}_2)$ has finite fourth moments, then the asymptotic distribution of $s_{1,t} = \mathrm{cor}(\boldsymbol{h}_1, \boldsymbol{h}_2)$ is given by

$$\frac{n\big(s_{1,t} - f_\mu^{-1}(\rho)\big)}{\sigma_\rho} \xrightarrow{d} \mathrm{N}(0,1), \tag{2.G.2}$$

where σ_ρ is a constant. Furthermore, for any $\rho > -1$, f_μ is differentiable at $f_\mu^{-1}(\rho)$. Thus by the delta method, we have

$$\frac{n\big(f_\mu(s_{1,t}) - \rho\big)}{f'_\mu(f_\mu(\rho))\sigma_\rho} \xrightarrow{d} \mathrm{N}(0,1). \tag{2.G.3}$$

Hence the MSE of point estimation using f_μ decays asymptotically at the rate of n^{-1}. Similarly, for the mode estimator, we have

$$\frac{n(f_\mu(s_{1,t}) - \rho)}{f'_\mu(f_\mu(\rho))\sigma_\rho} + \frac{n(\rho - f_m(f_\mu^{-1}(\rho)))}{f'_\mu(f_\mu(\rho))\sigma_\rho} \xrightarrow{d} \mathrm{N}(0,1). \tag{2.G.4}$$

Assume that

$$\frac{\rho - f_m(f_\mu^{-1}(x))}{f'_\mu(f_\mu(\rho))\sigma_\rho} \ll 0.$$

Hence the MSE of point estimation, using f_μ, is asymptotically decayed at the rate of $o(n^{-1})$, then remain constant . One can notice that f_μ seems to have better performance as it does not introduce an imperishable bias term in 2.G.4. We should note that we do not derive the f_μ theoretically. When ρ is getting closer to 1, the sample mean may not accurately reflect the theoretical mean of the data as the distribution is heavily skewed, which introduces a larger error in the fitting f_μ. In such cases, our parametric mode function would yield a better performance.

2.H. The range of ρ in band matrix structure

Cholecskey decomposition of a correlation matrix $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^\top$, where $\boldsymbol{L}$ is a lower triangular matrix, can be computed using a recurrent relation defined as

$$l_{i_1,i_2} = \begin{cases} 1, & i_1 = i_2 = 1, \\ \frac{1}{l_{i_1,i_1}}\left(\rho_{i_1,i_2} - \sum_{i_2'=1}^{i_1-1} l_{i_1,i_2'}^2\right), & i_2 < i_1, \\ \sqrt{1 - \sum_{i_2'=1}^{i_1-1} l_{i_1,i_2'}^2}, & i_1 = i_2 > 1, \\ 0, & i_2 > i_1. \end{cases}$$

where l_{i_1,i_2} is the (i_1, i_2) entry of matrix $\boldsymbol{L}$. For the one-band correlation matrix $\boldsymbol{Q}$, the recurrent relation can be simplified to

$$l_{i,j} = \begin{cases} 0, & j < i-1, \\ \frac{\rho}{l_{i,i}}, & j = i-1, \\ \sqrt{1 - l_{i,i-1}^2}, & i = j, \\ 0, & j > i. \end{cases}$$

For instance, if $d = 3$, then the Choleskey decomposition $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^\top$ would be

$$\begin{bmatrix} 1 & 0 & 0 \\ \rho & \sqrt{1-\rho^2} & 0 \\ 0 & \frac{\rho}{\sqrt{1-\rho^2}} & \sqrt{1 - \frac{\rho^2}{1-\rho^2}} \end{bmatrix} \begin{bmatrix} 1 & \rho & 0 \\ 0 & \sqrt{1-\rho^2} & \frac{\rho}{\sqrt{1-\rho^2}} \\ 0 & 0 & \sqrt{1 - \frac{\rho^2}{1-\rho^2}} \end{bmatrix} = \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}.$$

Now we define the following process

$$P_{-1} = 0, P_0 = 1, P_{i+1} = P_i - \rho^2 P_{i-1}. \tag{2.H.1}$$

Furthermore, we have

$$l_{1,1}^2 = 1 = \frac{P_1}{P_0}. \tag{2.H.2}$$

Assume that $l_{i,i}^2 = \frac{P_i}{P_{i-1}}$ for particular i, then we have

$$l_{i+1,i+1}^2 = 1 - \frac{\rho^2}{l_{i,i}^2} = 1 - \frac{\rho^2 P_{i-1}}{P_i} = \frac{P_{i+1}}{P_i}. \tag{2.H.3}$$

By 2.H.2, 2.H.3, we have $l_{i,i}^2 = \frac{P_i}{P_{i-1}}$ for all i. Then a necessary and sufficient condition for $\boldsymbol{Q}$ to be positive definite is $P_i > 0$ for all $i \leq d$.
Equation (2.H.1) can be written into the following matrix equation

$$\begin{pmatrix} P_i \\ P_{i-1} \end{pmatrix} = \boldsymbol{A}\begin{pmatrix} P_{i-1} \\ P_{i-2} \end{pmatrix} = \boldsymbol{A}^2\begin{pmatrix} P_{i-2} \\ P_{i-3} \end{pmatrix} = ... = \boldsymbol{A}^i\begin{pmatrix} P_0 \\ P_{-1} \end{pmatrix} = \boldsymbol{A}^i\begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ where } \boldsymbol{A} = \begin{pmatrix} 1 & -\rho^2 \\ 1 & 0 \end{pmatrix}$$

and the eigenvalues of $\boldsymbol{A}$ are given by

$$\lambda_1 = \frac{1 - \sqrt{1-4\rho^2}}{2} \quad \text{and} \quad \lambda_2 = \frac{1 + \sqrt{1-4\rho^2}}{2}.$$

When $\rho = 0$, the correlation matrix $\boldsymbol{Q}$ is clearly positive definite. When $-1/2 < \rho < 1/2$ and $\rho \neq 0$, we have $\lambda_1, \lambda_2 > 0$. Hence matrix A is positive definite, so as A^i.Thus the matrix $\boldsymbol{Q}$ is positive definite since

$$P_i = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} P_i \\ P_{i-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix} \boldsymbol{A}^i \begin{pmatrix} 1 \\ 0 \end{pmatrix} > 0.$$

When $\rho^2 > 1/4$, we have two complex eigenvalues. We set $\lambda_1 = |\rho| e^{-i\theta}$ where $\tan\theta = \sqrt{4\rho^2 - 1}, 0 < \theta < \frac{\pi}{2}$. Then $\lambda_2 = \rho e^{i\theta}$ and $\lambda_1 + \lambda_2 = \rho e^{-i\theta} + \rho e^{i\theta} = 1$. Define $v_1 = \begin{pmatrix} |\rho| \\ e^{i\theta} \end{pmatrix}$. Then we have

$$\boldsymbol{A} v_1 = \begin{pmatrix} 1 & -\rho^2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} |\rho| \\ e^{i\theta} \end{pmatrix} = \begin{pmatrix} |\rho| - \rho^2 e^{i\theta} \\ |\rho| \end{pmatrix} = \begin{pmatrix} |\rho|(1 - |\rho| e^{i\theta}) \\ \rho \end{pmatrix} = \begin{pmatrix} \rho^2 e^{-i\theta} \\ |\rho| \end{pmatrix} = \lambda_1 v_1,$$

Hence v_1 is an eigenvector of $\boldsymbol{A}$ with respect to eigenvalue λ_1. Apply the Rotation-Scaling Theorem, we have

$$\boldsymbol{C} = \begin{pmatrix} \mathrm{Re}(v_1) & \mathrm{Im}(v_1) \end{pmatrix} = \begin{pmatrix} |\rho| & 0 \\ \cos\theta & \sin\theta \end{pmatrix}, \ \boldsymbol{C}^{-1} = \frac{1}{|\rho| \sin\theta} \begin{pmatrix} \sin\theta & 0 \\ -\cos\theta & |\rho| \end{pmatrix}$$

$$\boldsymbol{B} = \begin{pmatrix} \mathrm{Re}(\lambda_1) & \mathrm{Im}(\lambda_1) \\ \mathrm{Im}(\lambda_1) & -\mathrm{Re}(\lambda_1) \end{pmatrix} = |\rho| \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

$$\boldsymbol{A} = \boldsymbol{C}\boldsymbol{B}\boldsymbol{C}^{-1}.$$

Thus we have

$$\begin{aligned} P_i &= \begin{pmatrix} 1 & 0 \end{pmatrix} \boldsymbol{A}^i \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \end{pmatrix} \boldsymbol{C}\boldsymbol{B}^i\boldsymbol{C}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} |\rho| & 0 \\ \cos\theta & \sin\theta \end{pmatrix} |\rho|^i \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}^i \frac{1}{|\rho| \sin\theta} \begin{pmatrix} \sin\theta & 0 \\ -\cos\theta & |\rho| \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \frac{|\rho|^i}{\sin\theta} \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}^i \begin{pmatrix} \sin\theta \\ -\cos\theta \end{pmatrix} \end{aligned} \tag{2.H.4}$$

Note that $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ defines a counterclockwise rotation around origin at angle θ and $\begin{pmatrix} \sin\theta \\ -\cos\theta \end{pmatrix} = \begin{pmatrix} \cos(\theta - \pi/2) \\ \sin(\theta - \pi/2) \end{pmatrix}$ is a unit vector in the fourth quadrant with the angle between the x-axis being $\pi/2 - \theta$. Thus we have

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}^i \begin{pmatrix} \sin\theta \\ -\cos\theta \end{pmatrix} = \begin{pmatrix} \cos\left[(i+1)\theta - \pi/2\right] \\ \sin\left[(i+1)\theta - \pi/2\right] \end{pmatrix}. \tag{2.H.5}$$

A graphical illustration of the action is given in Figure $2.H.1$. From (2.H.4), we have

$$P_i = \frac{|\rho|^i}{\sin\theta} \cos\left[(i+1)\theta - \pi/2\right] \quad \text{and} \quad l_{i,i} = \sqrt{Q_i} = \sqrt{|\rho| \frac{\cos\left[(i+1)\theta - \pi/2\right]}{\cos\left[i\theta - \pi/2\right]}}.$$

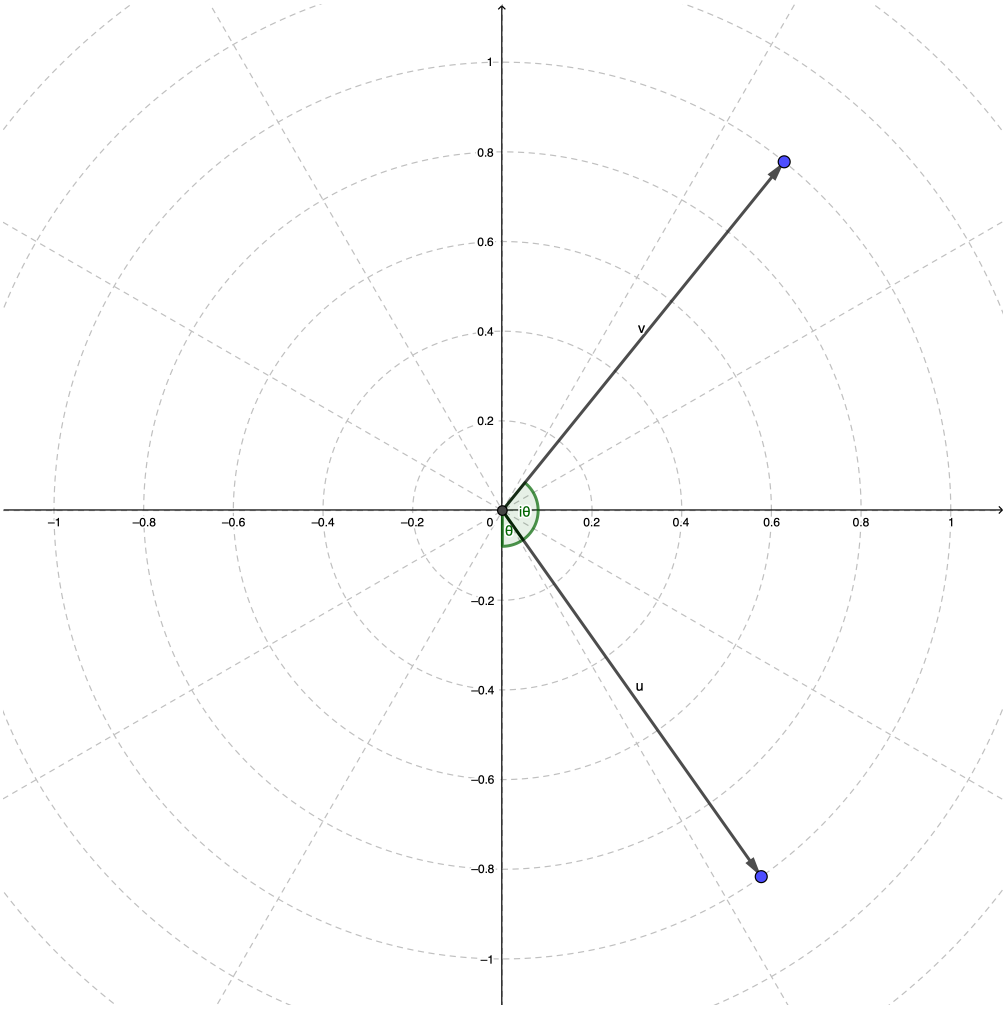


FIGURE 2.H.1. Action of rotation map act on u.

With dimension d, we require the existence of $l_{i,i}$ for all $1 \geq i \leq d$, that is

$$\left[(d+1)\theta - \pi/2\right] \leq \pi/2 \to \theta \leq \frac{\pi}{d+1} \to \sqrt{4\rho^2 - 1} \leq \tan\frac{\pi}{d+1} \to \rho^2 \leq \frac{\tan^2\frac{\pi}{d+1} + 1}{4}.$$

In summary, one-band correlation matrix $\boldsymbol{Q}$ is positive definite if and only if

$$-\frac{\sqrt{\tan^2\frac{\pi}{d+1} + 1}}{2} < \rho < \frac{\sqrt{\tan^2\frac{\pi}{d+1} + 1}}{2}. \tag{2.H.6}$$

When $\rho = \pm\frac{\sqrt{\tan^2\frac{\pi}{d+1} + 1}}{2}$, one-band correlation matrix $\boldsymbol{Q}$ is semi-positive definite.

2.I. Asymptotic behaviour of simple average covariance estimator

We aim to apply the test in (2.89) with p-value in (2.90) to test the hypotheses in (2.86) using sample covariance and simple average mode estimators which correspond to $\mathfrak{m} = 1, 2.2$ respectively. Since (2.89) is developed for using sample covariance estimates, we aim to show that

the distribution of $\frac{\hat{\rho}_{2.2,i,j}-\rho_{i,j}}{\sqrt{\text{Var}(\hat{\rho}_{2.1,i,j})}}$ for the correlation pair (i,j) using simple average mode estimator ($\mathfrak{m}=2.2$) converges to the distribution using sample covariance estimator ($\mathfrak{m}=1$) where $V_{i,j}=\text{Var}(\hat{\rho}_{2.2,i,j})$. Specifically, we aim to show

$$\frac{\hat{\sigma}_{i,j,\varsigma}-\sigma_{i,j,\varsigma}}{\sqrt{\hat{\theta}_{i,j,\tau}/n}} \xrightarrow{D} \frac{\hat{\sigma}'_{i,j,\varsigma}-\sigma'_{i,j,\varsigma}}{\sqrt{V_{i,j,\varsigma}}}, \quad \varsigma=1,2 \text{ sample}, \tag{2.I.1}$$

where $\sigma_{i,j,\varsigma}=\rho_{1,i,j,\varsigma}$, $\sigma'_{i,j,\varsigma}=\rho_{2.1,i,j,\varsigma}$ assuming all variances are one for simplicity. We then drop subscript ς referring to any one of the two samples. For sample covariance estimator, $\hat{\theta}_{i,j}$ is given in (2.88) whereas for simple average mode estimator, $\hat{\theta}_{2.2}$ is estimated by taking the sample variance of $\hat{\sigma}_{2.2}$ over $N=2000$ repeats (days). Simulation studies in Section 2.4.2 are performed with $n=60,300,600$, $N=2000$ and $\rho=0,0.3,0.6$. Figure 2.I.1 shows the distributions of standardised variables in (2.I.1) in which the black line refers to simple average mode estimates ($\mathfrak{m}=2.2$) whereas the grey line refers to sample covariance estimates ($\mathfrak{m}=1$). The plots show some biases when $\rho_{2.2,i,j}$ is higher. The biases can be seen in Figure 5d in which the simple average mode estimator shows minimum MSE but when $\rho\to 1$, the biases using simple average mode estimator enlarge. In this case, the sample correlation ($\mathfrak{m}=1$) or variance weighted mode estimator ($\mathfrak{m}=2.4$) are preferred.

After checking the convergence from the simulation study, we prove the convergence analytically. Assume that asymptotically the variance of the simple average mode estimate $\hat{\sigma}'_{i,j}$ decrease at rate $\frac{1}{n}$, i.e. $n\widehat{V}_{i,j} \xrightarrow{D} \theta'_{i,j}$, where the n is the sample size. Our simulation study shows that for different ρ values, the assumption is not violated. Table 2.I.1 summarise simulated value of the $n\hat{V}_{i,j}$ where $\sigma_i^2=0.00001$ and $\sigma_j^2=0.00002$ with the number repeats being $N=5000$.

TABLE 2.I.1. Values of nV by levels of true ρ and window size n

	60	100	200	300	400	500	750	1000
0	1.1794	1.1704	1.1557	1.1540	1.1567	1.1862	1.1622	1.1859
0.1	1.1582	1.1717	1.1675	1.1400	1.1330	1.1450	1.1506	1.1500
0.2	1.1690	1.1562	1.1354	1.1499	1.1419	1.1387	1.1580	1.1436
0.3	1.1344	1.1247	1.1025	1.1251	1.1153	1.1144	1.1282	1.1197
0.4	1.1105	1.1116	1.0876	1.0961	1.0804	1.0790	1.0813	1.1057
0.5	1.0606	1.0628	1.0767	1.0667	1.0534	1.0635	1.0643	1.0674

All figures are multiplied by $\times 10^{-5}$.

Now define

$$\mathcal{M}_{i,j}=\frac{(\hat{\sigma}_{i,j,1}-\hat{\sigma}_{i,j,2})^2}{\hat{\theta}_{i,j,1}/n_1+\hat{\theta}_{i,j,2}/n_2} \text{ and } \mathcal{M}'_{i,j}=\frac{(\hat{\sigma}'_{i,j,1}-\hat{\sigma}'_{i,j,2})^2}{\hat{V}_{i,j,1}+\hat{V}_{i,j,2}}$$

in (2.92). Under H_0, we have $\sigma_{i,j,1} = \sigma_{i,j,2}$. If we have $n_1 = n_2 = n$, then

$$\begin{aligned}\mathcal{M}_{i,j} &= \left(\frac{\sqrt{n}\hat{\sigma}_{i,j,1} - \sqrt{n}\hat{\sigma}_{i,j,2}}{\sqrt{\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2}}}\right)^2 \\ &= \left[\frac{\sqrt{n}(\hat{\sigma}_{i,j,1} - \sigma_{i,j,1})}{\sqrt{\hat{\theta}_{i,j,1}}}\frac{\sqrt{\hat{\theta}_{i,j,1}}}{\sqrt{\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2}}} - \frac{\sqrt{n}(\hat{\sigma}_{i,j,2} - \sigma_{i,j,2})}{\sqrt{\hat{\theta}_{i,j,2}}}\frac{\sqrt{\hat{\theta}_{i,j,2}}}{\sqrt{\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2}}}\right]^2 \qquad (2.I.2)\\ \mathcal{M}'_{i,j} &= \left(\frac{\sqrt{n}\hat{\sigma}'_{i,j,1} - \sqrt{n}\hat{\sigma}'_{i,j,2}}{\sqrt{n\hat{V}_{i,j,1} + n\hat{V}_{i,j,2}}}\right)^2 \\ &= \left[\frac{\hat{\sigma}'_{i,j,1} - \sigma_{i,j,1}}{\sqrt{\widehat{V}_{i,j,1}}}\frac{\sqrt{n\widehat{V}_{i,j,1}}}{\sqrt{n\widehat{V}_{i,j,1} + n\widehat{V}_{i,j,2}}} - \frac{\hat{\sigma}'_{i,j,2} - \sigma_{i,j,2}}{\sqrt{\widehat{V}_{i,j,2}}}\frac{\sqrt{n\widehat{V}_{i,j,2}}}{\sqrt{n\hat{V}_{i,j,1} + n\hat{V}_{i,j,2}}}\right]^2. \qquad (2.I.3)\end{aligned}$$

Under H_0, we have $\theta_{i,j,1} = \theta_{i,j,2}$ and $\theta'_{i,j,1} = \theta'_{i,j,2}$. Hence we have

$$\frac{\sqrt{\hat{\theta}_{i,j,\varsigma}}}{\sqrt{\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2}}} \xrightarrow{P} \sqrt{\frac{\theta_{i,j,\varsigma}}{\theta_{i,j,1} + \theta_{i,j,2}}} = \frac{1}{\sqrt{2}} \qquad (2.I.4)$$

$$\frac{\sqrt{n\widehat{V}_{i,j,\varsigma}}}{\sqrt{n\widehat{V}_{i,j,1} + n\widehat{V}_{i,j,2}}} \xrightarrow{P} \sqrt{\frac{\theta'_{i,j,\varsigma}}{\theta'_{i,j,1} + \theta'_{i,j,2}}} = \frac{1}{\sqrt{2}} \qquad (2.I.5)$$

where $\varsigma = 1, 2$ samples. Then by (2.I.1) to (2.I.5), we have $\mathcal{M}_{i,j} \xrightarrow{D} \mathcal{M}'_{i,j}$ Hence $\mathcal{M} \xrightarrow{D} \mathcal{M}'$. To see whether the assumption (2.I.1), we compare the empirical distribution of our simulated data. Note that

$$\frac{\bar{\sigma}_{i,j,\varsigma} - \sigma_{i,j,\varsigma}}{\sqrt{r\hat{\theta}_{i,j,\varsigma}/n}} \xrightarrow{P} 0$$

as $n/N \to 0$ where $\bar{\sigma}_{i,j,\varsigma}$ is the mean of n realisations of $\hat{\sigma}_{i,j,\varsigma}$. Hence we have

$$\frac{\hat{\sigma}_{i,j,\varsigma} - \bar{\sigma}_{i,j,\varsigma}}{\sqrt{\hat{\theta}_{i,j}/n}} = \frac{\hat{\sigma}_{i,j,\varsigma} - \sigma_{i,j,\varsigma}}{\sqrt{\hat{\theta}_{i,j}/n}} - \frac{\bar{\sigma}_{i,j,\varsigma} - \sigma_{i,j,\varsigma}}{\sqrt{\hat{\theta}_{i,j,\varsigma}/n}} \xrightarrow{D} \frac{\hat{\sigma}_{i,j,\varsigma} - \sigma_{i,j,\varsigma}}{\sqrt{\hat{\theta}_{i,j}/n}} \qquad (2.I.6)$$

as $n \to \infty, n/N \to 0$. A similar argument also applied to our range estimators. Thus we only need to show the distribution of standardised estimators is asymptotically the same, given the number of repeats $N \gg n$. Figure 2.I.1 shows kernel density estimations of standardised estimators under different simulation schemes. Table 2.I.2 summarises p-value of 2-sample Kolmogorov-Smirnov Tests between standardised range and sample covariance estimators with the number of repeats being $N = 5000$. Under our simulation scheme, the return follows i.i.d normal. Hence the sample covariance follows normal asymptotically. That is, we expect that the standardized estimator follows $\mathrm{N}(0, 1)$. Table 2.I.3 summarises the p-values of Kolmogorov-Smirnov tests of standardised range against standard normal distribution.

TABLE 2.I.2. p-values of 2-sample Kolmogorov-Smirnov test by levels of true ρ and window size n

	60	100	200	300	400	500	750	1000
0	0.9572	0.9124	0.9987	0.9228	0.9800	0.9981	0.9639	0.9639
0.1	0.8772	0.9124	0.5441	0.8508	0.9572	0.9753	0.9992	0.5770
0.2	0.5936	0.6272	0.8896	0.9903	0.9013	0.9800	0.9926	0.9800
0.3	0.1486	0.4209	0.5770	0.7604	0.7112	0.8643	0.9800	0.7442
0.4	0.3399	0.3657	0.9415	0.5936	0.7920	0.7604	0.9124	0.1700
0.5	0.0582	0.1857	0.2392	0.4067	0.3927	0.9013	0.7278	0.9700

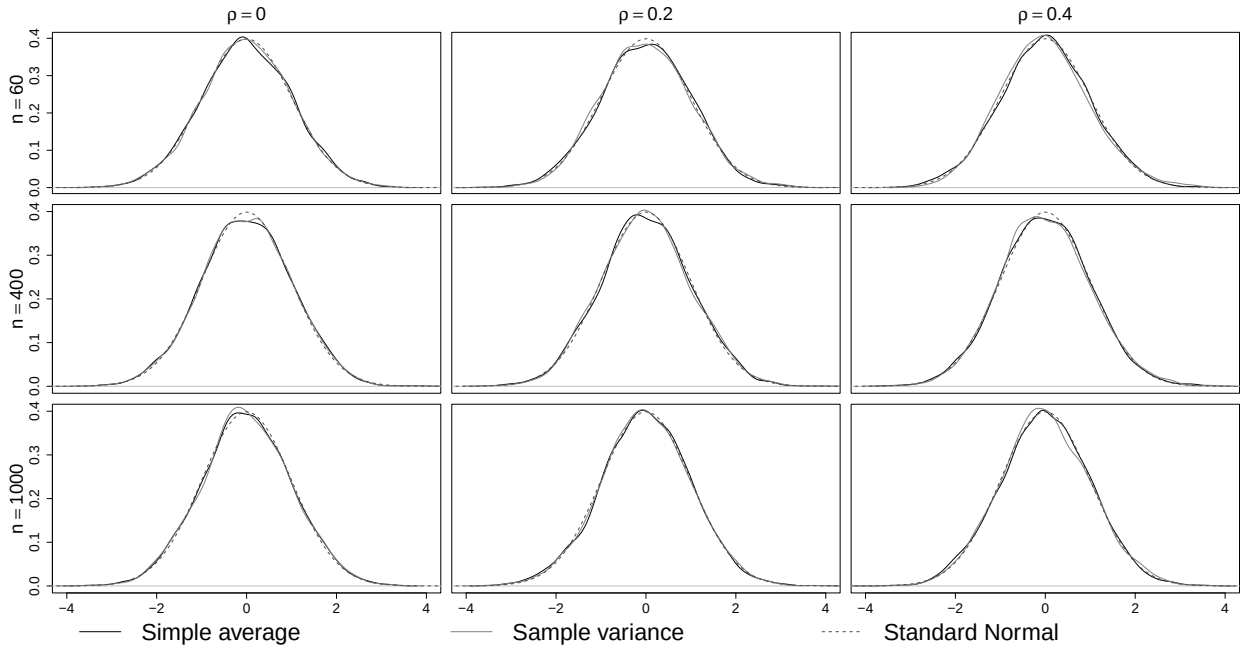


FIGURE 2.I.1. The distribution of standardized estimators

TABLE 2.I.3. p-values of one-sample Kolmogorov-Smirnov test by levels of true ρ and window size n

	60	100	200	300	400	500	750	1000
0	0.8875	0.7244	0.9975	0.9827	0.9230	0.7859	0.9446	0.8552
0.1	0.8327	0.9355	0.5086	0.8406	0.8619	0.9742	0.6248	0.8749
0.2	0.9082	0.9975	0.8571	0.9885	0.8956	0.9261	0.9318	0.8175
0.3	0.8956	0.7319	0.1854	0.8799	0.9962	0.8297	0.7191	0.6922
0.4	0.8578	0.9352	0.9935	0.5434	0.9465	0.7015	0.9176	0.7394
0.5	0.7711	0.9073	0.9957	0.8096	0.9228	0.9877	0.7190	0.9816

CHAPTER 3

Multivariate Conditional Autoregressive Range Model

3.1. Introduction

This Chapter aims to extend CARR models to multivariate CARR (MCARR) models to provide a direct and efficient modelling of covariance matrix series. By managing the inherent uncertainty and unpredictability, the models will ultimately contribute to more efficient and robust financial systems.

The development of volatility models has evolved substantially over the years. The ARCH model (Engle, 1982), introduced the concept of volatility clustering, capturing the phenomenon of high volatility periods followed by relative calm periods. Building upon this, the GARCH model (Bollerslev, 1986) incorporated lagged values of both returns and volatility, offering improved forecasting capabilities. The EGARCH model of Nelson (1991) and the GJR-GARCH model of Glosten et al. (1993) are extensions of the GARCH model to capture the asymmetric volatility caused by positive and negative returns. Further advancements include SV models that allow volatility itself to be a stochastic process (Bergomi, 2015).

A simple multivariate extension to the GARCH (MGARCH) model is the vector autoregressive GARCH (VAR-GARCH) model and its VEC-GARCH model extension, which estimates the covariance matrix of multiple time series by the autoregressive and moving average components of contemporaneous and lagged covariance matrices (Bollerslev et al., 1988). However, the covariance matrix processes using scalar coefficients may not be positive definite. Another popular multivariate GARCH model is the BEKK model (Engle and Kroner, 1995), which used quadratic forms of coefficients to preserve the positive definite property of the correlation matrix. A more general form of the volatility model separates the volatility dynamics into two components: the diagonal entries of the covariance matrix and the off-diagonal entries of the correlation matrix. These MGARCH models include the CCC-GARCH model (**?**) with constant correlation over time and the DCC-GARCH model (Engle, 2002) with dynamic correlation using a combination of past values of correlation and error terms.

However, these GARCH and MGARCH models model returns and treat volatility as a hidden process of returns. Instead, MCARR models directly the covariance matrix measures to smooth out the random noise and identify patterns of the covariance processes. Compared to MGARCH models, MCARR models take the volatility matrices, not return vectors, as a stochastic instead of a deterministic process. The fitted covariance matrices can be applied to multivariate return models to capture the heteroskedasticity and cross-dependency of returns. Hence, this extended two-stage MCARR-return model is different from the MGARCH and MSV models that use only the information of return vectors. Instead, it uses both returns and volatility information and models them as two stochastic processes. Moreover, CARR-return models also differ from the realised GARCH models (Gerlach and Wang, 2016) since the latent volatility equation is omitted, although the realised GARCH model also has two stochastic processes for returns and realised volatilities.

An extension to the MCARR model was considered by Fernandes et al. (2005). The it was not covariance matrix but the variances of individual components and their pairwise sums that modelled directly. Then, the covariances of the pairwise components can be recalculated based on the variance information. However, the effects on the variances of the pairwise sum may not reflect the effects on the covariance. The objective of this chapter to fill up this research gap is three-folded. Firstly, we propose the MCARR model to model the positive definite covariance matrix series using the Wishart distribution and suggest two approaches for modelling the mean of the Wishart distribution, borrowing the ideas of BEKK-GARCH and DCC-GARCH models. Then, the fitted volatilities are applied to the multivariate return model to forecast multiple returns. This model is called the two-stage MCARR-return model. Secondly, we conduct simulation experiments to study a number of issues to fine-tune the models. The models are implemented using a Bayesian approach via the user-friendly `RStan` package. Lastly, we apply the best-proposed model to model and forecast returns, volatilities, VaRs and correlations of three stock market indices.

The remaining chapter is organised as follows. Section 3.2 reviews the return and volatility measures extending from Section 1.3. Section 3.3 reviews the MCARR-return models extending from Section 1.4.2 and proposes new MCARR-return models. Section 3.4 describes the Bayesian implementation of proposed models. Section 3.5 conducts simulation studies to evaluate model performances across model choices, levels of correlation, stability of $\boldsymbol{Q}$-matrix, and the impact of the shape parameter of Wishart distribution and risk metrics. Lastly, Section 3.6 concludes the chapter and discusses future research. An application of the proposed models will be given in Section 4.2.

3.2. Review of return and volatility measures

Let $\boldsymbol{\kappa}_{i,t} = (\boldsymbol{o}_{i,t}, \boldsymbol{h}_{i,t}, \boldsymbol{l}_{i,t}, \boldsymbol{c}_{i,t})^\top$ defined in (1.5) be the collection of normalised open, high, low and close price of asset i on day t, where $\boldsymbol{o}_{i,t}, \boldsymbol{h}_{i,t}, \boldsymbol{l}_{i,t}$ and $\boldsymbol{c}_{i,t}$ vectors consist of normalised intraday open, high, low and close with length n_t. We consider d return series, and denote $\mathcal{F}_t = \sigma\{(\boldsymbol{\kappa}_{1,t}, ..., \boldsymbol{\kappa}_{d,t})^\top\}$ be the natural σ-algebra generated by the multivariate price process. Assume that the price within a day follows a geometric Brownian motion with constant drift and volatility in (1.4). We can use the intraday data to conduct a covariance matrix measure $\boldsymbol{V}_t$. However, the data size will be very large if we use the entire intraday OHLC data to build a model. Furthermore, the majority of the financial markets are not trading 24 hours per day and 7 days per week. Especially, the off-market hour is longer than the trading hour, which results in a large proportion of missing. Thus we use the summary statistic of intraday data, such as intraday open, high, low and close, to build a model for intraday return and volatility. Intraday return using closing and opening (CO) prices is given by

$$r_{i,t}^{\texttt{CO}} = \sum_{\tau=1}^{n_t} c_{i,t,\tau}. \tag{3.1}$$

We adopt the CO measure in all subsequent analyses and drop the superscript `CO` to simplify our notation.

3.3. Two-stage volatility-return models

In this section, we review the development of volatility models, from univariate to multivariate settings and the extension to multivariate two-stage volatility-return models.

3.3.1. Review of univariate volatility-return model. Let v_t be some volatility measures of an asset price series using some existing methods. Chou (2005) proposed the CARR model of order (p, q) given by (1.53) to (1.55). The difference between daily high and low is equivalent to a daily volatility measure. By utilising intraday transactions, the daily range can be replaced by a realised intraday range volatility measure to reduce the measuring error.

Then the second stage return model for the return series $\{R_1, R_2, ..., R_N\}$ is

$$R_t|\mathcal{F}_{t-1} \sim \mathrm{F}_r(\mu_t, \sigma_t^2, \gamma, \nu) \tag{3.2}$$

where $\mathrm{F}_r(\mu_t, \sigma_t^2, \nu)$ is certain distribution such as normal, Student-t and variance gamma (Madan and Seneta, 1990) with real support, mean parameter μ_t, scale parameter σ_t^2, skewness parameter γ, and shape parameter ν. The mean function can adopt an AR(1) structure such as

$$\mu_t = \psi_0 + \psi_1 r_{t-1} \tag{3.3}$$

and the scale parameter is estimated by

$$\sigma_t^2 = \omega\lambda_t \tag{3.4}$$

where λ_t is the fitted volatility in (1.54) and ω is a scale adjustment parameter to ensure unbiasedness. Since the two-stage model in (1.53) to (1.55) and (3.2) to (3.4) takes two sources of information from the volatility measures v_t and returns r_t, it is generally known to be more efficient than the GARCH model in (1.32) to (1.34) which considers only returns r_t.

3.3.2. Review of marginal multivariate volatility model. In portfolio formulation, several asset returns $\boldsymbol{r}_{i,1:N}, i = 1, \ldots, d$ are considered. To model the dependence between returns $\boldsymbol{r}_t = \boldsymbol{r}_{1:d,t}$ at time t, a simple extension of the two-stage model in (1.55) and (3.2) to model $\boldsymbol{r}_t$ is

$$\text{Model I: } V_{i,t}|\mathcal{F}_{t-1} \sim \mathrm{F}_v(\lambda_{i,t}, \phi_i), \quad \boldsymbol{R}_t|\mathcal{F}_{t-1} \sim \mathrm{F}_{\boldsymbol{r}}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t, \boldsymbol{\gamma}, \nu), \quad \boldsymbol{\Sigma}_t = \boldsymbol{\Lambda}_t^{1/2}\boldsymbol{\Upsilon}\boldsymbol{\Lambda}_t^{1/2}$$

where $V_{i,t}$ are some volatility measures, $\mathrm{F}_v(\lambda_{i,t}, \phi_i)$ is some distributions with positive support such as $\text{Weibull}\Big(\phi_i, \lambda_{i,t}\big[\Gamma(1 + 1/\phi_i)\big]^{-1}\Big)$, $\lambda_{i,t}$ is given in (1.54), $\mathrm{F}_{\boldsymbol{r}}(\cdot)$ is some d-dimensional multivariate distribution with real supports such as multivariate Student-t distribution, the mean vector

$$\boldsymbol{\mu}_t = (\mu_{1,t}, \ldots, \mu_{d,t}),\ \mu_{i,t} = \psi_{0i} + \psi_{1i} r_{i,t-1},\ \boldsymbol{\Lambda}_t = \text{diag}(\lambda_{1,t}, \ldots, \lambda_{d,t}),$$

the skewness vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)$ and $\boldsymbol{\Upsilon}$ is a time invariant positive definite matrix. However, the model models the dynamics of the variance processes in (1.54) but assumes the correlation in $\boldsymbol{\Upsilon}$ to be time-invariant similar to the CCC-GARCH in (1.63) to (1.65). This model was proposed in Nitithumbundit and Chan (2022) as a model extension.

Alternatively, Tan et al. (2022) considered the multivariate CARR (MCARR) volatility-return model using the covariance formula

$$\text{Cov}(R_{i,t}, R_{j,t}) = [\text{Var}(R_{i,t} + R_{j,t}) - \text{Var}(R_{i,t}) - \text{Var}(R_{j,t})]/2.$$

The model is given by

$$\text{Model II: } V_{ij,t}|\mathcal{F}_{t-1} \sim \mathrm{F}_v(\lambda_{ij,t}, \phi_{ij}), \boldsymbol{R}_t|\mathcal{F}_{t-1} \sim \mathrm{F}_{\boldsymbol{r}}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t, \boldsymbol{\gamma}, \nu), \quad \boldsymbol{\Sigma}_t = \boldsymbol{\Omega}^{1/2}\boldsymbol{\Lambda}_t\boldsymbol{\Omega}^{1/2} \tag{3.5}$$

where the observed values $v_{ii,t} = v_{i,t}$ is the variance measure for return series i, $v_{ij,t}$ is the variance measures of the pairwise sum of return series i and j,

$$\mathbf{\Lambda}_t = \begin{pmatrix} \lambda_{1,t} & \lambda_{21,t} & \dots & \lambda_{d1,t} \\ \lambda_{21,t} & \lambda_{2,t} & \dots & \lambda_{d2,t} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{d1,t} & \lambda_{d2,t} & \dots & \lambda_{d,t} \end{pmatrix},$$

$\lambda_{i,t} = \alpha_{0i} + \alpha_{1i} v_{i,t-1} + \beta_{1i} \lambda_{i,t-1}$ is given in (1.54) for order (1,1) and adding subscript i to indicate components for parameters α_0, β_1 and $\mathbf{\Omega} = \text{diag}(\omega_1^{1/2}, \dots, \omega_d^{1/2})$ is a diagonal matrix of scale adjustment factors ω_i. The covariance estimates are given by

$$\begin{aligned} \lambda_{ij,t} &= (\lambda^*_{ij,t} - \lambda_{i,t} - \lambda_{j,t})/2, \\ \lambda^*_{ij,t} | \mathcal{F}_{t-1} &= \alpha^*_{0ij} + \alpha^*_{1ij} v^*_{ij,t-1} + \beta^*_{1i} \lambda^*_{ij,t-1}, \end{aligned} \tag{3.6}$$

where $v^*_{ij,t}$ is a variance measure such as (1.17) for $\text{Var}(R_{i,t} + R_{j,t})$. Note that $\mathbf{\Sigma}_t$ in (3.5) is a multivariate extension of (3.4).

However, the MCARR model models the covariances indirectly by modelling the variances of the pairwise sums of two price series. Another issue is that volatility is measured as the variance of return instead of price. Returns of the sum of two price series are different from the sum of two return series in general. Alternatively, Hansen et al. (2014) proposed the BGARCH model to model correlation directly in (1.72).

We propose a direct modelling of the observed covariance matrices $\boldsymbol{V}_t = (V_{ij,t}) = \widehat{\mathbf{\Sigma}}_t$ in (2.91) using some distributions for positive definite matrices.

3.3.3. Wishart distribution. We assign Wishart distribution to the observed $d \times d$ covariance matrices $\boldsymbol{V}_t = (V_{ij,t})$ directly, that is,

$$\boldsymbol{V}_t | \mathcal{F}_{t-1} \quad \sim \quad \text{Wishart}(\mathbf{\Lambda}_t, \nu), \tag{3.7}$$

where the pdf of Wishart (W_d) distribution is

$$f_{\boldsymbol{V}_t,d}(\mathbf{v}_t | \mathbf{\Lambda}_t) = \frac{|\mathbf{v}_t|^{(\nu-d-1)/2} \exp(-\text{tr}(\mathbf{\Lambda}_t{}^{-1} \mathbf{v}_t)/2)}{2^{\nu d/2} |\mathbf{\Lambda}_t|^{\nu/2} \Gamma_d(\nu/2)},$$

where $\text{tr}(\cdot)$ is the trace function and $\Gamma_d(\frac{\nu}{2}) = \pi^{d(d-1)/4} \prod_{i=1}^{d} \Gamma(\frac{n+i-1}{2})$, is a matrix generalisation of the chi-square distribution. The degree of freedom ν has to be greater than $d-1$, while in practice we often set $\nu > d+1$, which allows the mode of the distribution to exist. $\Gamma_d(\cdot)$ is a multivariate gamma function. The mean and variance of Wishart distribution are given by

$$\text{E}(\boldsymbol{V}_t | \mathbf{\Lambda}_t) = \nu \mathbf{\Lambda}_t \quad \text{and} \quad \text{Var}(V_{ij,t} | \mathbf{\Lambda}_t) = \nu(\lambda^2_{ij,t} + \lambda_{ii,t} \lambda_{jj,t}), \tag{3.8}$$

respectively.

Essentially, assume that $\boldsymbol{R}_1, ..., \boldsymbol{R}_N \overset{i.i.d}{\sim} \text{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\boldsymbol{R}_t$ is a $d \times 1$ column vector and $\mathbf{\Sigma}$ is $d \times d$ positive definite matrix. Then we have

$$\sum_{t=1}^{N} \boldsymbol{R}_t \boldsymbol{R}_t^\top \sim \text{W}_d(\mathbf{\Sigma}, \nu). \tag{3.9}$$

Furthermore, let $\bar{\boldsymbol{R}} = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{R}_t$. Then just like in univariate cases, we have

$$\sum_{t=1}^{N}(\boldsymbol{R}_t - \bar{\boldsymbol{R}})(\boldsymbol{R}_t - \bar{\boldsymbol{R}})^\top \sim \mathrm{W}_d(\boldsymbol{\Sigma}, \nu - 1). \tag{3.10}$$

There are many other important properties of Wishart distribution, for example, the partitioning of $\boldsymbol{\Sigma}$. See Page Jr (1984) for more details.

3.3.4. Intuitions of two-stage model. Let $\boldsymbol{r}_{t,\tau} = (r_{1,t,\tau}, ..., r_{d,t,\tau})^\top$ be the vector of return in day t at time τ. The daily return on day t is given by

$$\boldsymbol{r}_t = \sum_{\tau=1}^{n_t}\boldsymbol{r}_{t,\tau}. \tag{3.11}$$

By assumption, we have

$$\boldsymbol{r}_{t,\tau}|\mathcal{F}_{t-1} \overset{i.i.d}{\sim} \mathrm{N}(\boldsymbol{\mu}_t/n_t, \boldsymbol{\Sigma}_t/n_t) \quad \text{for } \tau \in \{1, ..., n_t\}. \tag{3.12}$$

Now consider the sample covariance estimators as a daily volatility measure, that is

$$\boldsymbol{V}_t^* = \frac{n_t}{n_t - 1}\sum_{\tau=1}^{n_t}(\boldsymbol{r}_{t,\tau} - \boldsymbol{r}_t/n_t)(\boldsymbol{r}_{t,\tau} - \boldsymbol{r}_t/n_t)^\top. \tag{3.13}$$

Note that to estimate $\boldsymbol{\Sigma}_t$, we need to rescale sample covariance estimators by n_t. Then by (3.10), we know that

$$\boldsymbol{V}_t^* \sim \mathrm{W}_d(\boldsymbol{\Sigma}_t, n_t - 1). \tag{3.14}$$

Furthermore, by normality assumption defined in (3.12), we can conclude that $\boldsymbol{r}_t = \sum_{\tau=1}^{n}\boldsymbol{r}_{t\tau}$ and $\boldsymbol{V}_t$ are independent. The result leads to the following joint model

$$\begin{aligned} \boldsymbol{V}_t^* &\sim \mathrm{W}_d(\boldsymbol{\Sigma}_t, n_t - 1), & (3.15)\\ \boldsymbol{R}_t &\sim \mathrm{N}_d(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t). & (3.16) \end{aligned}$$

Such a joint model utilises the intraday information $\boldsymbol{r}_{t,\tau}$ to estimate $\boldsymbol{V}_t^*$ as compared to the MGARCH model which only relies on (3.16). Furthermore, the autoregressive effect of the covariance matrix can now be captured by lagged matrices $\boldsymbol{V}_{t-k}^*$, which is more efficient compared to the MGARCH model using lagged pointwise return-squares $r_{i,t-k}^2$ in (1.68). Moreover, the covariance matrix measures $\boldsymbol{V}_t = (V_{ij,t}) = \widehat{\boldsymbol{\Sigma}}_t$ in (2.91) is more efficient than $\boldsymbol{V}_t^*$ in measuring covariance matrices. The return model in (3.16) can also be extended to more general distributions such as multivariate Student-t (ST). Compared to the MCARR model in (3.5) and (3.6), we now specify a distribution for the covariance matrix measure $\boldsymbol{V}_t$ whereas the MCARR models of Chou et al. (2009) only specify the marginal distribution of volatility for each asset.

3.3.5. Proposed joint multivariate volatility model. There are two approaches, the covariance model(model A) and correlation model(model B), to model the unscaled mean function $\boldsymbol{\Lambda}_t$. The first approach takes the BEKK-GARCH (Engle and Kroner, 1995) model approach in (1.61) and (1.62) and is also a multivariate extension of the mean function in (1.54). It models $\boldsymbol{\Lambda}_t$ in terms of the long and short-term persistence of covariance matrices given by

$$\text{Model A: } \boldsymbol{\Lambda}_t = \boldsymbol{A}_0 + \boldsymbol{A}_1^\top\boldsymbol{V}_{t-1}\boldsymbol{A}_1 + \boldsymbol{B}_1^\top\boldsymbol{\Lambda}_{t-1}\boldsymbol{B}_1, \tag{3.17}$$

where $\boldsymbol{\Lambda}_t = (\lambda_{ij,t})$ is a $d\times d$ positive definite matrix, the degrees of freedom parameter is $\nu > d-1$, $\boldsymbol{A}_0$, $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ are $d \times d$ coefficient matrices, $\boldsymbol{A}_0$ is positive definite for the long term mean, $\det(\boldsymbol{A}_1) \neq 0$, and $\det(\boldsymbol{B}_1) \neq 0$. Note that (3.17) is invariant of the sign of $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$, that is, we

still get the same result if we substitute $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ by $-\boldsymbol{A}_1$ and $-\boldsymbol{B}_1$ respectively. To resolve the problem, we introduce a non-unique restriction such as restricting the first entities, $a_{1,11}$ and $b_{1,11}$ of $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$, respectively, to be positive.

The second approach is based on the DCC-GARCH model in (1.66) to (1.70) to model the mean covariance matrix $\boldsymbol{\Lambda}_t$ in terms of the long and short term persistence of each variance component $\sigma_{i,t}^2$ and the correlation matrix $\boldsymbol{Q}_t$ instead of the covariance matrix. The unscaled mean matrix of (3.7) is given by

$$\begin{aligned}
\text{Model B: } \boldsymbol{\Lambda}_t &= \text{diag}(\sigma_{1,t},\ldots,\sigma_{d,t})\boldsymbol{Q}_t\text{diag}(\sigma_{1,t},\ldots,\sigma_{d,t}), && (3.18)\\
\sigma_{i,t}^2 &= a_{0i} + a_i v_{ii,t-1} + b_i\sigma_{i,t-1}^2, && (3.19)\\
\boldsymbol{Q}_t &= \boldsymbol{\Psi} + \alpha(\widehat{\boldsymbol{\mathcal{Q}}}_{t-1} - \boldsymbol{\Psi}) + \beta(\boldsymbol{Q}_{t-1} - \boldsymbol{\Psi}), && (3.20)
\end{aligned}$$

where $a_{0i}, a_i, b_i > 0$, the correlation, long-term correlation, and observed correlation matrices are given by

$$\boldsymbol{Q}_t = \begin{pmatrix} 1 & \rho_{21,t} & \cdots & \rho_{d1,t} \\ \rho_{21,t} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d1,t} & \cdots & \cdots & 1 \end{pmatrix}, \quad \boldsymbol{\Psi} = \begin{pmatrix} 1 & \psi_{21} & \cdots & \psi_{d1} \\ \psi_{21} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{d1} & \cdots & \cdots & 1 \end{pmatrix}, \quad \widehat{\boldsymbol{\mathcal{Q}}}_t = \begin{pmatrix} 1 & \varrho_{21,t} & \cdots & \varrho_{d1,t} \\ \varrho_{21,t} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \varrho_{d1,t} & \cdots & \cdots & 1 \end{pmatrix}, \tag{3.21}$$

respectively, the stationary condition for variances $\sigma_{i,t}^2$ is $a_i + b_i < 1$, the stationary condition for correlation $\rho_{ij,t}$ is $\alpha + \beta < 1$, and $\widehat{\boldsymbol{\mathcal{Q}}}_t$ is correlation matrix estimate on day t in (2.57). We replace $r_{i,t-1}^2$ in the GARCH model with $v_{ii,t-1}$ in (3.19) where $\boldsymbol{V}_t = (v_{ij,t}) = \widehat{\boldsymbol{\Sigma}}_t$ in (2.91). Furthermore, we model the process directly via (3.20), while $\boldsymbol{Q}_t$ is a latent process which lacks physical interpretation in DCC-GARCH. For the special case of the CCC-GARCH-type Wishart volatility model, $\boldsymbol{Q}_t$ in (3.20) becomes time-invariant with $\alpha = \beta = 0$.

3.3.6. Two stage MCARR-return model. We propose the return model in MCARR-return model as following

$$\begin{aligned}
\boldsymbol{R}_t|\mathcal{F}_{t-1} &\sim \text{F}_{\boldsymbol{r}}(\boldsymbol{\mu}_t, \nu\boldsymbol{\Lambda}_t, \boldsymbol{\gamma}, \upsilon) && (3.22)\\
\boldsymbol{\mu}_t|\mathcal{F}_{t-1} &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{R}_{t-1}, && (3.23)
\end{aligned}$$

where $\text{F}_{\boldsymbol{r}}(\cdot)$ is a multivariate normal, multivariate Student-t or multivariate variance gamma distributions with various levels of skewness $\boldsymbol{\gamma}$ and kurtosis υ, and ν is the degrees of freedom of Wishart distribution.

We recognise that this joint model is computationally intensive even in lower dimensions. Meanwhile, the independence between $\boldsymbol{r}_t = \sum_{\tau=1}^{n} \boldsymbol{r}_{t\tau}$ and $\boldsymbol{V}_t$ are only applied for normal assumption, which rises difficulty in generalising the model to other distribution since the joint density of $(\boldsymbol{V}_t, \boldsymbol{R}_t)$ is not straightforward. Thus instead of having a joint model, we propose a 2-stage model defined as follows:

$$\begin{aligned}
\text{Stage-1 volatility model: } \boldsymbol{V}_t &\sim \text{G}(\boldsymbol{\Lambda}_t, \nu), && (3.24)\\
\text{Stage-2 return model: } \boldsymbol{R}_t &= \boldsymbol{\mu}_t + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \text{F}_r(\boldsymbol{0}, \boldsymbol{\Omega}\widehat{\boldsymbol{\Lambda}}_t\boldsymbol{\Omega}, \upsilon), && (3.25)
\end{aligned}$$

where $\widehat{\boldsymbol{\Lambda}}_t$ is the estimate of $\boldsymbol{\Lambda}_t$ in stage-1 model, $\text{G}(\boldsymbol{\Lambda}_t, \nu)$ is a $d \times d$ covariance matrix distribution with mean matrix $\boldsymbol{\Lambda}_t$, $\text{F}_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a d-dimensional distribution with location parameter $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{\Omega}$ is a diagonal matrix of scale adjustment factors, which captures the variability of

covariance matrix in the second stage model. Furthermore, $\boldsymbol{\Omega}$ can also adjust the potential scaling bias for each asset in our volatility measure in the first stage model.

While our two-stage model may not specify the exact distribution of $(\boldsymbol{V}_t, \boldsymbol{R}_t)$, our two-stage model does have some advantages. In general, MCMCs take a shorter time to converge compared to the joint model. The two-stage model allows feasibility in choosing the marginal distribution of $\boldsymbol{V}_t$ and $\boldsymbol{R}_t$, while the joint density of $(\boldsymbol{V}_t, \boldsymbol{R}_t)$ is more complicated due to the dependence of $\boldsymbol{V}_t$ and $\boldsymbol{R}_t$ when the normality assumption is dropped. Furthermore, the covariance matrix measure $\boldsymbol{V}_t$ can be any kind of volatility measure, allowing the adoption of range volatility measure as in (2.91) and discussed in Section 1.3.2. Here, we choose $\mathrm{G}(\boldsymbol{\Lambda}, \nu)$ to be the Wishart distribution with unknown degrees of freedom ν and $\mathrm{F}_r(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ can be the multivariate Student-t distribution with unknown degrees of freedom υ. Our two-stage model is given by:

$$\text{Stage-1 Volatility model: } \boldsymbol{V}_t \sim \text{Wishart}_d(\boldsymbol{\Lambda}_t, \nu), \tag{3.26}$$

$$\text{Stage-2 Return model: } \boldsymbol{R}_t = \boldsymbol{\mu}_t + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathrm{ST}_d(\boldsymbol{0}, \boldsymbol{\Omega}\widehat{\nu\boldsymbol{\Lambda}}_t\boldsymbol{\Omega}, \upsilon). \tag{3.27}$$

As the variance of multivariate Student-t distribution is given by $\frac{\upsilon}{\upsilon-2}\omega_i^2\nu\lambda_{i,t}$ where $\lambda_{i,t} = \mathrm{Diag}(\boldsymbol{\Lambda}_t)$, we expect that $\frac{\upsilon}{\upsilon-2}\omega_i^2 \approx 1$ for all i if $\widehat{\nu\boldsymbol{\Lambda}}_t$ is an unbiased estimate of the true covariance matrix $\frac{\upsilon}{\upsilon-2}\boldsymbol{\Sigma}_t$.

In summary, the multivariate two-stage volatility-return model or the Wishart(1,1)-T(1,0) model is given by

$$\text{Volatility model: } \boldsymbol{V}_t \sim \text{Wishart}_d(\boldsymbol{\Lambda}_t, \nu), \tag{3.28}$$

$$\boldsymbol{\Lambda}_t = \begin{cases} \boldsymbol{A}_0 + \boldsymbol{A}_1^\top \boldsymbol{V}_{t-1}\boldsymbol{A}_1 + \boldsymbol{B}_1^\top \boldsymbol{\Lambda}_{t-1}\boldsymbol{B}_1, & \text{if Model A,} \\ \mathrm{diag}(\sigma_{1,t}, \ldots, \sigma_{d,t})\boldsymbol{Q}_t\mathrm{diag}(\sigma_{1,t}, \ldots, \sigma_{d,t}), & \\ \sigma_{i,t}^2 = a_{0i} + a_i v_{ii,t-1} + b_i\sigma_{i,t-1}^2, & \text{if Model B,} \\ \boldsymbol{Q}_t = \boldsymbol{\Psi} + \alpha(\widehat{\boldsymbol{\mathcal{Q}}}_{t-1} - \boldsymbol{\Psi}) + \beta(\boldsymbol{Q}_{t-1} - \boldsymbol{\Psi}), & \end{cases}$$

$$\text{Return model: } \boldsymbol{R}_t = \boldsymbol{\mu}_t + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathrm{F}_{\boldsymbol{r}}(\boldsymbol{0}, \nu\boldsymbol{\Lambda}_t, \boldsymbol{\gamma}, \upsilon), \tag{3.29}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{R}_{t-1} \tag{3.30}$$

such that

$$\mathrm{E}(V_{ii,t}) = \nu\lambda_{ii,t}, \quad \mathrm{E}(\varrho_{ij,t}) = \rho_{ij,t}, \quad \text{and} \quad \mathrm{E}(R_{i,t}) = \mu_{i,t} \tag{3.31}$$

where $\boldsymbol{V}_t = (V_{ij,t})$, $\boldsymbol{\Lambda}_t = (\lambda_{ij,t})$, $\widehat{\boldsymbol{\mathcal{Q}}}_t = (\varrho_{ij,t})$ and $\boldsymbol{Q}_t = (\rho_{ij,t})$. Equivalently, (3.28) can be written as

$$\boldsymbol{V}_t = \boldsymbol{\Lambda}_t^{1/2}\boldsymbol{\Upsilon}_t\boldsymbol{\Lambda}_t^{1/2}, \quad \boldsymbol{\Upsilon}_t \sim \mathrm{Wishart}(\boldsymbol{I}_d, \nu).$$

3.4. Bayesian inference

In recent years, the Bayesian MCMC method has been extensively used because it is one of the most efficient estimation tools since there is no need for the numerical optimisation of the intractable likelihood function. In addition, the method allows one to incorporate apriori beliefs on the hierarchical model structures. Therefore, we develop a Bayesian modelling framework for the analysis instead of using the frequentist approach. The Bayesian approach treats probability as a measure of belief or certainty about an event, incorporating prior knowledge or beliefs for the model parameters along with the sample data to update the estimates using Bayes' theorem. Bayesian methods allow model parameters to be treated as random variables and provide a more

flexible framework for incorporating external information or subjective beliefs into the analysis. The analytical derivation of the posterior distribution is not necessary in recent years as samples from the posterior distribution can be stimulated using the Markov Chain Monte Carlo (MCMC) method. The implementation of MCMC in the Bayesian approach has become increasingly popular, making it feasible to compute more complex hierarchical models. We will show in the next section that this is naturally very beneficial for obtaining posterior predictive distributions that provide distributional forecast summaries for the proposed hierarchical models.

3.4.1. Posterior distribution. Given a set of observed values $\boldsymbol{y}_{1:T} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T), \boldsymbol{y}_t \in \mathbb{R}^d$ and the vector of model parameters $\boldsymbol{\vartheta}$, the posterior distribution for $\boldsymbol{\vartheta}$ conditional on $\boldsymbol{y}_{1:d,1:T}$ is given by

$$\pi(\boldsymbol{\vartheta}|\boldsymbol{y}_{1:T}) = \frac{f(\boldsymbol{y}_{1:d,1:T}|\boldsymbol{\vartheta})\ \pi(\boldsymbol{\vartheta})}{\int f(\boldsymbol{y}_{1:d,1:T}|\boldsymbol{\vartheta})\ \pi(\boldsymbol{\vartheta})\ d\boldsymbol{\vartheta}} \propto f(\boldsymbol{y}_{1:d,1:T}|\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta}),$$

which is proportional to the likelihood function $f(\boldsymbol{y}_{1:d,1:T}|\boldsymbol{\vartheta})$ and the prior densities $\pi(\boldsymbol{\vartheta})$ estimated using available information or past data. Even if there is no prior information, we can still apply non-informative or reference priors. Furthermore, the Bayesian approach generates posterior distributions for all parameters of interest and hence allows the construction of creditable intervals for these parameters.

The posterior distribution of parameters for the Wishart volatility model A in (3.17) and B in (3.18) to (3.20) are given by

$$\pi(\boldsymbol{\theta}_{1m}|\mathcal{F}_T) = \frac{\prod_{t=1}^{T} f_{\mathrm{W}}(\mathbf{v}_t|\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_{1m}), \nu, \mathcal{F}_T)\pi_{\boldsymbol{\theta}_{1m}}}{\int \prod_{i=1}^{T} f_{\mathrm{W}}(\mathbf{v}_t|\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_{1m}), \nu, \mathcal{F}_T)\pi_{\boldsymbol{\theta}_{1m}} d\boldsymbol{\theta}_{1m}},\ m = A, B$$

where $f_{\mathrm{W}}(\cdot)$ is the density of Wishart distribution, $\boldsymbol{\theta}_{1A} = (\boldsymbol{A}_0, a_{1,1}, ..., a_{d,d}, b_{1,1}, ..., b_{d,d}, \nu)$, and $\boldsymbol{\theta}_{1B} = (a_{01}, ..., a_{0d}, a_1, ..., a_d, b_1, ..., b_d, \alpha, \beta, \psi_{21}, \ldots, \psi_{d,d-1}, \nu)$. The priors are

$$\pi_{\boldsymbol{\theta}_{1A}} = \pi_{\boldsymbol{A}_0}\pi_{a_{11}}\pi_{b_{11}}\Big[\prod_{(i,j)\setminus(1,1)}^{(d,d)} \pi_{a_{ij}}\pi_{b_{ij}}\Big]\pi_\nu, \quad \pi_{\boldsymbol{A}_0} = \mathrm{W}(\bar{\boldsymbol{V}}/\nu, \nu),\ \pi_{a_{11}}, \pi_{a_{11}} = \mathrm{HN}(0, 1), \tag{3.32}$$

$$\pi_{a_{ij}}, \pi_{b_{ij}} = \mathrm{N}(0, 1), i, j = 1, \ldots, d, (i, j) \neq (1, 1),\ \pi_\nu = \gamma(15, 1). \tag{3.33}$$

$$\pi_{\boldsymbol{\theta}_{1B}} = \Big[\prod_{i=1}^{d} \pi_{a_{0i}}\pi_{a_i}\pi_{b_i}\pi_{\zeta_i}\pi_{\beta_i}\Big]\Big[\prod_{i,j=2,1}^{d,d-1} \pi_{\psi_{ij}}\Big]\pi_\nu,\ \pi_{a_{0i}} = \gamma(0.5, 1),\ \pi_{a_i}, \pi_{b_i} = \beta(2, 5), \tag{3.34}$$

$$\pi_{\psi_{ij}} = \mathrm{N}(0.6, 0.2)I(-1, 1), i, j = 1, \ldots, d, i > j,\ \pi_\alpha, \pi_\beta = \beta(2, 20),\ \pi_\nu = \gamma(15, 1). \tag{3.35}$$

where $\mathrm{W}(\cdot,\cdot)$, $\mathrm{HN}(\cdot,\cdot)$, $\mathrm{N}(\cdot,\cdot)$, $\gamma(\cdot,\cdot)$ and $\beta(\cdot,\cdot)$ denote Wishart, half normal, normal, gamma and beta distributions respectively. For model A in (3.17), we restrict the prior distributions for a_{11} in $\boldsymbol{A}_1$ and b_{11} in $\boldsymbol{B}_1$ to be positive using a half normal distribution to avoid the redundancy. We set the prior of the long-term mean $\boldsymbol{A}_0 \sim \mathrm{Wishart}(\bar{\boldsymbol{V}}/\nu, \nu)$, where $\bar{\boldsymbol{V}}$ is the sample mean of covariance matrices $\boldsymbol{V}_t$. The priors for a_i, b_i, α, β have a higher density for values < 0.5 to ensure stationarity conditions for variances and correlations.

For the return model, $\boldsymbol{\theta}_2 = (\boldsymbol{\phi}_0, \boldsymbol{\Phi}_1, \boldsymbol{\gamma}, \upsilon)$, and the whole vector of parameters is $\boldsymbol{\theta} = (\boldsymbol{\theta}_{1m}, \boldsymbol{\theta}_2), m = A, B$. The posterior distributions for $\boldsymbol{\theta}_{1m}, m = A, B$ and $\boldsymbol{\theta}$ are given by

$$\pi(\boldsymbol{\theta}_2|\mathcal{F}_N, \boldsymbol{\theta}_{1m}) = \frac{\prod_{t=1}^{N} f_r(\boldsymbol{r}_t|\boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_{1m}), \boldsymbol{\gamma}, \upsilon, \mathcal{F}_N)\pi_{\boldsymbol{\theta}_2}}{\int \prod_{t=1}^{N} f_r(\boldsymbol{r}_t|\boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_{1m}), \boldsymbol{\gamma}, \upsilon, \mathcal{F}_N)\pi_{\boldsymbol{\theta}_2} d\boldsymbol{\theta}_2},$$

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathcal{F}_N) &= \pi(\boldsymbol{\theta}_{1m}|\mathcal{F}_N)\pi(\boldsymbol{\theta}_2|\mathcal{F}_N, \boldsymbol{\theta}_{1m}) \\ &= \frac{\prod_{t=1}^{N} f_{\mathrm{W}}(\mathbf{v}_t|\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_{1m}), \nu, \mathcal{F}_N) f_r(\boldsymbol{r}_t|\boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_{1m}), \boldsymbol{\gamma}, \upsilon, \mathcal{F}_N)\pi_{\boldsymbol{\theta}}}{\int \prod_{t=1}^{N} f_{\mathrm{W}}(\mathbf{v}_t|\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_{1m}), \nu, \mathcal{F}_N) f_r(\boldsymbol{r}_t|\boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_{1m}), \boldsymbol{\gamma}, \upsilon, \mathcal{F}_N)\pi_{\boldsymbol{\theta}} d\boldsymbol{\theta}}.\end{aligned}$$

respectively where $f_r(\cdot)$ can be $f_{\mathrm{ST},d}(\cdot)$, the density of multivariate Student-t distribution,

$$\begin{aligned}&\pi_{\boldsymbol{\theta}_2} = \pi_{\phi_0}\pi_{\boldsymbol{\gamma}}\pi_{\upsilon}\prod_{i,j=1}^{d}\pi_{\phi_{i,j}}, \\ &\pi_{\boldsymbol{\phi}_0} = \mathrm{N}_d(\boldsymbol{\phi}_0|\mathbf{0}, 0.000001\boldsymbol{I}),\ \pi_{\phi_{i,j}} = \mathrm{N}(0,1),\ i,j = 1,...,d,\ \pi_{\boldsymbol{\gamma}} = \mathrm{N}_d(\mathbf{0}, \boldsymbol{I}),\ \pi_{\upsilon} = \mathrm{N}(0,1), \\ &\pi_{\boldsymbol{\theta}} = \pi_{\boldsymbol{\theta}_{1m}}\pi_{\boldsymbol{\theta}_2},\ m = 1, 2.\end{aligned}$$

The predictive distributions for 1-step ahead forecast $\boldsymbol{V}_{t+1}|\mathcal{F}_N$ and $\boldsymbol{r}_{t+1}|\mathcal{F}_N$ are given by

$$\pi(\boldsymbol{V}_{N+1}|\mathcal{F}_N) = \int f_{\mathrm{W}}(\boldsymbol{V}_{N+1}|\boldsymbol{\Lambda}_{N+1}, \nu, \mathcal{F}_N)\pi(\boldsymbol{\theta}_{1m})\, d\boldsymbol{\theta}_{1m}, \tag{3.36}$$

$$\begin{aligned}\pi(\boldsymbol{r}_{N+1}, \boldsymbol{V}_{N+1}|\mathcal{F}_N) &= \int f_r(\boldsymbol{r}_{N+1}|\boldsymbol{\mu}_{\boldsymbol{N+1}}, \boldsymbol{\Lambda}_{N+1}, \boldsymbol{\gamma}, \upsilon, \mathcal{F}_N)\pi(\boldsymbol{V}_{N+1}|\boldsymbol{\theta}_{1m}, \mathcal{F}_N)\pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_{1m})\, d\boldsymbol{\theta}_2 \\ &= \iint f_r(\boldsymbol{r}_{N+1}|\boldsymbol{\mu}_{N+1}, \boldsymbol{\Lambda}_{N+1}, \boldsymbol{\gamma}, \upsilon, \mathcal{F}_N) f_{\mathrm{W}}(\boldsymbol{V}_{N+1}|\boldsymbol{\Lambda}_{N+1}, \nu, \mathcal{F}_N)\pi(\boldsymbol{\theta}_{1m})\pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_{1m})\, d\boldsymbol{\theta}_{1m} d\boldsymbol{\theta}_2,\end{aligned} \tag{3.37}$$

where $m = A, B$.

3.4.2. Model implementation. The Bayesian models are implemented via the user-friendly Bayesian software `RStan` package as the `R` interface of `Stan` (Team et al., 2016) which utilises the `Stan` program within `R` developed in the `C++` language. To draw posterior samples, the package applies the no-U-turn sampler (NUTS) (Hoffman and Gelman, 2014), an adepted Hamiltonian Monte Carlo (HMC) (Duane et al., 1987; Neal, 1994; Betancourt and Girolami, 2013) which converges to target distribution more efficiently compared to Random Walk Metropolis Hasting (Metropolis et al., 1953) and Gibbs sampler (Geman and Geman, 1984) via utilising the full marginal variance even if there is a high local correlation of target distribution in hierarchical models. This is possible also because the HMC sampler transforms the simulation of high-dimensional target distribution into the simulation of Hamiltonian dynamics (Neal, 2011). By using the gradient information, Hamiltonian dynamics can produce distant proposals for the Metropolis algorithm, thereby avoiding the slow exploration of the state space that results from the diffusive behaviour of random-walk proposals. The Hamiltonian Monte Carlo generate posterior sampling for a vector of parameters $\boldsymbol{\theta}$. The system introduces a movement variable $\boldsymbol{\xi}$ and $(\boldsymbol{\theta}, \boldsymbol{\xi})$ draws from

$$\pi(\boldsymbol{\xi}, \boldsymbol{\theta}) = \pi(\boldsymbol{\xi}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

In most application, we set $\boldsymbol{\xi}$ to be independent from $\boldsymbol{\theta}$ and $\boldsymbol{\xi} \sim \mathrm{N}(0, \boldsymbol{M})$. By default `Stan` sets $\boldsymbol{M}$ to be the inverse of a diagonal estimate of the covariance computed during the warm-up. The $f(\boldsymbol{\theta}, \boldsymbol{\xi})$ defines a Hamiltonian

$$\begin{aligned} H(\boldsymbol{\xi}, \boldsymbol{\theta}) &= -\log \pi(\boldsymbol{\xi}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\ &= -\log \pi(\boldsymbol{\xi}|\boldsymbol{\theta}) - \log \pi(\boldsymbol{\theta}) \\ &= T(\boldsymbol{\xi}|\boldsymbol{\theta}) + V(\boldsymbol{\theta}) \end{aligned}$$

A transition to a new state is generated following steps. First, the momentum $\boldsymbol{\xi}$ is drawn independently for the current iteration. Next, the joint system is evaluated by

$$\frac{d\boldsymbol{\theta}}{dt} = \frac{\partial H}{\partial \boldsymbol{\xi}} = \frac{\partial T}{\partial \boldsymbol{\xi}} \tag{3.38}$$

$$\frac{d\boldsymbol{\xi}}{dt} = -\frac{\partial H}{\partial \boldsymbol{\xi}} = -\frac{\partial T}{\partial \boldsymbol{\theta}} - \frac{\partial V}{\partial \boldsymbol{\theta}} = -\frac{\partial V}{\partial \boldsymbol{\theta}}. \tag{3.39}$$

Note $\frac{\partial T}{\partial \boldsymbol{\theta}} = 0$ as the momentum density is independent of the target density. Then leapfrog integrator is used to solve the two-state differential equation numerically to get $(\boldsymbol{\xi}^*, \boldsymbol{\theta}^*)$. Finally, we have a metropolis accept step where the acceptance probability is defined as

$$\min(1, e^{H(\boldsymbol{\xi},\boldsymbol{\theta}) - H(\boldsymbol{\xi}^*,\boldsymbol{\theta}^*)}). \tag{3.40}$$

3.4.3. Convergence diagnostics. To assess the dependence, precision and convergence of the posterior sample, two measures are reported in `Rstan`. The first measure is the *number of effective samples* n_{eff} which indicates dependence within a Monte Carlo sample. To monitor the convergence for $k > 2$ chains of length $2n$ each, Gelman and Rubin (1992) proposed $\widehat{\mathcal{R}}$ which is defined as

$$\widehat{\mathcal{R}} = \frac{\widehat{\mathcal{V}}}{\mathcal{W}} \cdot \frac{\mathrm{df}}{\mathrm{df} - 2},$$

where

$$\widehat{\mathcal{V}} = \frac{n-1}{n}\mathcal{W} + \frac{k+1}{kn}\mathcal{B}, \quad \mathcal{W} = \sum_{i=1}^{k} \frac{s_i^2}{k}, \quad \mathcal{B} = n\sum_{i=1}^{k} \frac{(\overline{\vartheta}_{i.} - \overline{\vartheta}_{..})^2}{k-1}, \quad \mathrm{df} = \frac{2\widehat{\mathcal{V}}^2}{\widehat{\mathrm{Var}}(\widehat{\mathcal{V}})},$$

$$\begin{aligned} \widehat{\mathrm{Var}}(\widehat{\mathcal{V}}) &= \left(\frac{n-1}{n}\right)^2 \frac{1}{k}\widehat{\mathrm{Var}}(s_i^2) + \left(\frac{k+1}{kn}\right)^2 \frac{2}{k-1}\mathcal{B}^2 + \\ &\quad 2\frac{(k+1)(n-1)}{kn^2} \cdot \frac{n}{k}\Big[\widehat{\mathrm{Cov}}(s_i^2, \overline{\vartheta}_{i.}^2) - 2\overline{\vartheta}_{..}\widehat{\mathrm{Cov}}(s_i^2, \overline{\vartheta}_{i.})\Big], \end{aligned}$$

s_i is the within-chain variance and ϑ_{ij} is the j-th parameter in chain i. If $\widehat{R}$ is close to 1, the parameter ϑ has converged.

We run a single Markov chain of 10,000 iterations in total, discard the first 5,000 iterations as burn-in and sample the subsequence of $L = 5,000$ iterations. We check $\widehat{\mathcal{R}}$ and the history plot carefully to ensure that all parameters meet the convergence condition. The number of effective samples ranges from 1727 to 5008 across parameters and the range of $\widehat{\mathcal{R}}$ is between 0.9979 and 1.0000 indicating moderate dependency and clear convergence.

3.4.4. Model selection. The capability to fit complicated models raises the requirement to compare different model formulations. The Deviance Information Criterion(DIC), defined as

$$\text{DIC} = D(\bar{\boldsymbol{\theta}}) + 2p_D = \overline{D(\boldsymbol{\theta})} + p_D, \tag{3.41}$$

$$p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}), \tag{3.42}$$

is widely used by researchers as the goodness of fit measure. A smaller deviance, defined as $D(\boldsymbol{\theta}) = -2\log(p(\boldsymbol{y}|\boldsymbol{\theta})) + C$, where C is an arbitrary constant, suggests a higher log-likelihood of data $\boldsymbol{y}$ given parameter $\boldsymbol{\theta}$. The p_D term can be viewed as the penalty of the model dimension. Furthermore, p_D is strictly non-negative if the log-likelihood is a concave function in $\boldsymbol{\theta}$. But it can still be negative in general cases. Studies suggested that a negative p_D indicates an improper choice of the prior (Spiegelhalter et al., 2002).

Furthermore, if we compare the DIC of multiple models, the decision is invariant of the constant C, which could be helpful for the likelihood function containing a complicated normalisation constant since we can simply drop it. Note that DIC, as well as the decision of model comparison, is not invariant under parametrisation. Suppose we have a parametrization of $\boldsymbol{\theta}$ such that $f(\boldsymbol{\phi}) = \boldsymbol{\theta}$. We choose an appropriate prior for $\boldsymbol{\phi}$ such that the posterior distributions of $f(\boldsymbol{\phi})$ (and $\boldsymbol{\theta}$ are identical. We do have $\overline{D(\boldsymbol{\theta})} = \overline{D(f(\boldsymbol{\phi}))}$ in this case. But $\bar{\boldsymbol{\theta}} \neq f(\bar{\boldsymbol{\phi}})$ in general, which leads to the discrepancy in p_D. Replacing $\bar{\boldsymbol{\theta}}$ by posterior median does lead to an invariant p_D. But it not only introduces additional computational difficulty but also violates some approximate properties (Spiegelhalter et al., 2002). Hence we should pay extra attention to the parametrisation of model parameters when we compare DIC across models.

3.5. Simulation studies

Between the proposed models A and B, we find model A describes correlation indirectly through covariance. Hence, the accuracy of the correlation estimate is affected by the accuracy of both variances in the pair, often leading to inaccurate correlation estimates. Because of this, we perform our first simulation study to compare the performance of models A and B. Results show a preference for model B. The second simulation study assesses the accuracy of parameter estimates for model B under two choices of priors for a subset of parameters. Results show that the performance of model B is affected by the shape parameter ν of the Wishart distribution and the correlation level given by ψ_{ij} in the long-term mean $\boldsymbol{\Psi}$ of the correlation model in (3.20). We perform the five simulation studies to assess their effects.

3.5.1. Study 1: Comparing model A on covariance and model B on correlation. We simulate $M = 1$ bivariate ($d = 2$) covariance matrix series from model A in (3.7) and (3.17) of size $N = 1000$ and fit the simulated series to model A and B (cases AA and AB). The true parameters of model A are set to be

$$\text{True A: } \boldsymbol{A}_0 = \begin{pmatrix} 4.0 & 3.9 \\ 3.9 & 6.0 \end{pmatrix} \times 10^{-6}\ \boldsymbol{A}_1 = \begin{pmatrix} 0.29 & 0.01 \\ 0.00 & 0.30 \end{pmatrix},\ \boldsymbol{B}_1 = \begin{pmatrix} 0.32 & -0.13 \\ 0.05 & 0.37 \end{pmatrix},\ \nu = 10. \tag{3.43}$$

and the initial condition the mean function for is $\boldsymbol{\Lambda}_0 = \boldsymbol{A}_0$. We then fit the simulated data to model A and model B. Parameter estimates from fitting to model A are given by

$$\text{Est AA: } \boldsymbol{A}_0 = \begin{pmatrix} 3.8 & 3.5 \\ 3.5 & 5.7 \end{pmatrix} \times 10^{-6},\ \boldsymbol{A}_1 = \begin{pmatrix} 0.29 & 0.15 \\ 0.05 & 0.25 \end{pmatrix},\ \boldsymbol{B}_1 = \begin{pmatrix} 0.24 & 0.17 \\ -0.19 & -0.13 \end{pmatrix},\ \nu = 10. \tag{3.44}$$

All true parameter values in (3.43) are contained in the 95% credible intervals of the estimates in (3.44). Comparing the DIC -58626.01 and -58624.27 for cases AA and AB respectively, we

find the DIC of case AA is marginally better than that of case AB as expected. Similarly, we also simulate one covariance matrix series from model B in (3.7) and (3.18) to (3.20). The simulated series is fitted to models A and B (cases BA and BB). The true parameters of model B are set to be

$$\text{True B:}\quad a_{01} = 4\times10^{-6}, a_1 = 0.75, b_1 = 0.2, a_{02} = 6\times10^{-6}, a_2 = 0.85, b_2 = 0.1, \alpha = 0.8, \beta = 0.1,$$

$$\psi = 0.8, \nu = 10, \tag{3.45}$$

and the parameter estimates from fitting to model B are given by

$$\text{Est BB:}\quad a_{01} = 3.6\times10^{-6}, a_1 = 0.08, b_1 = 0.21, a_{02} = 5.2\times10^{-6}, a_2 = 0.09, b_2 = 0.18, \alpha = 0.07, \beta = 0.12,$$

$$\psi = 0.78, \nu = 10.11. \tag{3.46}$$

Comparing the DICs -55965.75 and -58756.51 for cases BA and BB, we find the DIC of case BB is much better than that of case BA confirming the better performance of model B than model A. Plots of observed (dark grey), fitted (black) and 95% credible intervals for volatilities, covariance and correlation for the four cases are presented in Figure 3.1. Results show that there is no obvious pattern in variance and covariance processes (and so for correlation) when data are simulated from model A although we have tried a few sets of true parameter values. Clearly, the model fit for the BB case is better than the BA case as the former has more precise confidence intervals. As model B shows some obvious preference in terms of precisely capturing patterns in the variance and correlation processes, we proceed with the simulation studies using model B and apply model B in real application.

3.5.2. Study 2: The effect of ν and prior sensitivity of parameter estimates for model B. In this study, we simulate $M = 100$ bivariate covariance matrix series $\boldsymbol{V}_t$ from model B similar to study 1. The true parameters are given by

$$a_{01} = 4 \times 10^{-6},\ a_1 = 0.3,\ b_1 = 0.2,\ \alpha = 0.2, \tag{3.47}$$

$$a_{02} = 6 \times 10^{-6},\ a_2 = 0.4,\ b_2 = 0.1,\ \beta = 0.1. \tag{3.48}$$

with the stationarity constraints $a_i + b_i < 1$ and $\alpha + \beta < 1$ for the variance and correlation processes respectively fulfilled. We compare two sets of shape parameters: the smaller set $\nu = 5, 10, 15, 20, 25$ and the larger set $\nu = 10, 30, 50, 100$ which affects the variance of $V_{ij,t}$ in (3.8). For the correlation level, we set $\psi_{21} = \psi = 0.8$ to reflect the often observed highly positive correlation between asset indices in financial markets. Figures 3.2 and 3.3 dispatch violin plots of all parameters for the smaller and larger sets of shape parameters respectively when the priors for a_i, b_i are $\beta(2, 5)$ (with mean 0.29, variance 0.025, and higher density below 0.5). Priors for other parameters are the same between the two figures and they are given in (3.34) and (3.35). Results show that true values lie closer to the median of violin plots showing that estimates are more accurate when ν is lower in general. Moreover, all true values lie within the 95% confidence intervals for smaller ν in Figure 3.2.

Next, we study the accuracy of parameter estimates under two choices of priors for a_i, b_i: $\beta(2, 5)$ and $\beta(2, 2)$ (with mean 0.5, variance 0.05, and symmetric density). Figures 3.3 and 3.4 dispatch violin plots of all parameters for the two choices of priors respectively when $\nu = 10, 30, 50, 100$. Results show that true values lie closer to the median of violin plots when $\beta(2, 5)$ prior is used. Except for two parameters, all true values lie within the 95% confidence intervals in Figure 3.3. When $\beta(2, 2)$ prior is used, the accuracy of parameter estimates gets worse with increasing ν.

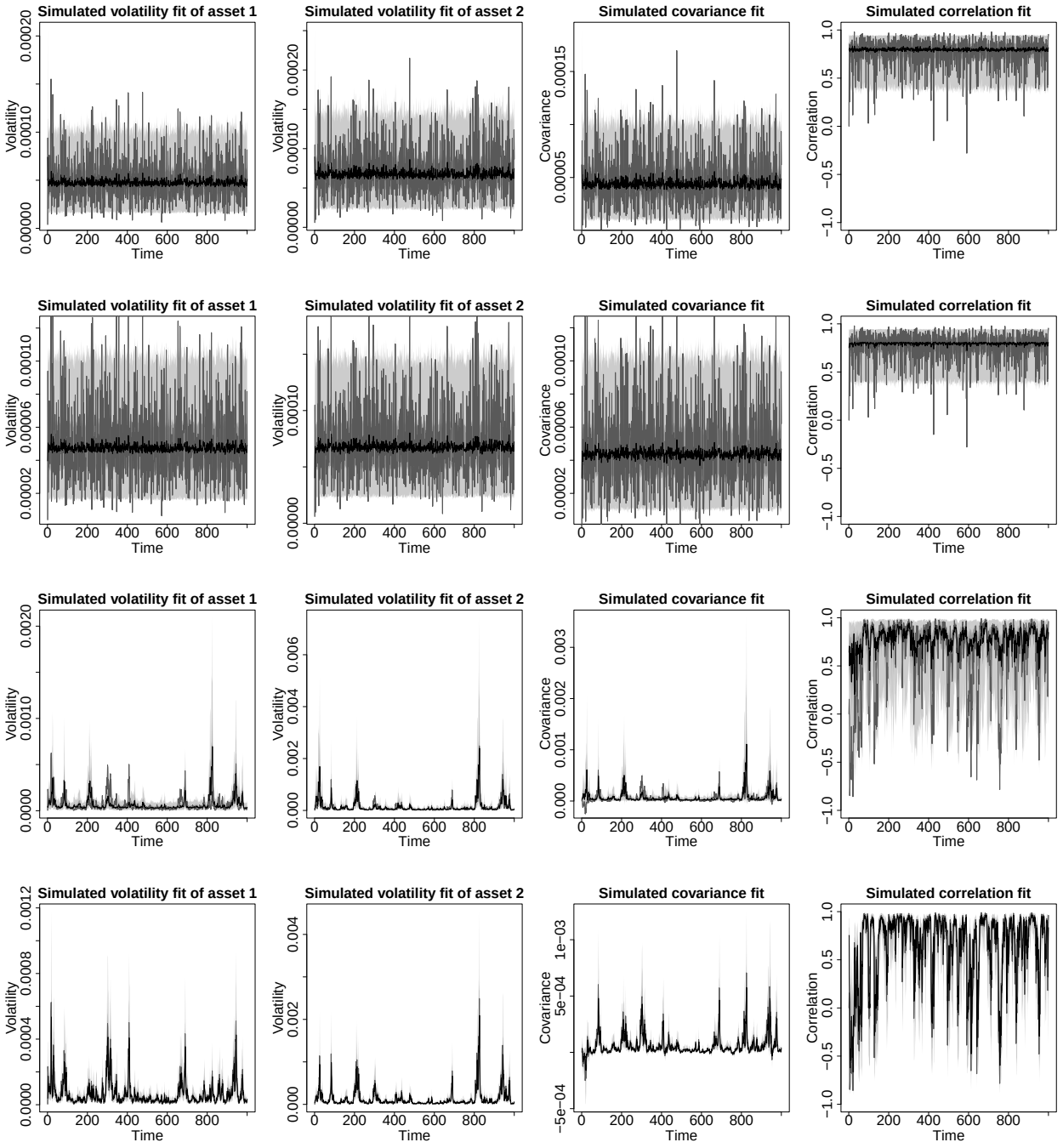


FIGURE 3.1. Observed (dark grey), fitted median (black), 50% credible interval (grey) and 95% credible intervals (light grey) for volatilities (column 1 and 2), covariance (column 3) and correlation (column 4) for cases AA, AB, BA and BB (row 1 to 4 respectively) for simulation study 1.

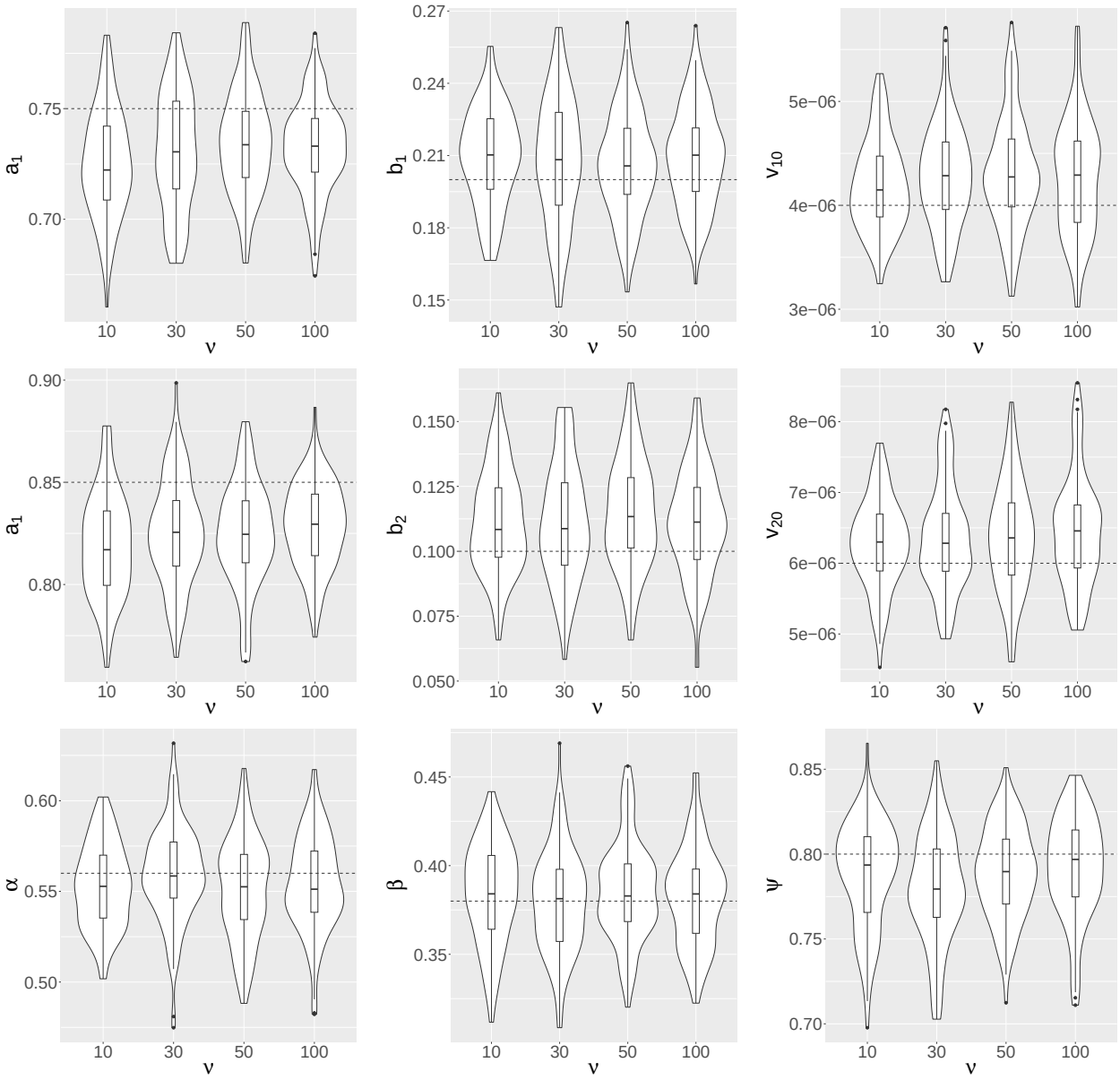


FIGURE 3.2. Violin plots of parameter estimates for $M = 100$ simulated covariance matrix series using model B for ν = 5,10, 15, 20, 25 and $\beta(2, 5)$ prior for a_i, b_i in simulation study 2. `v10` and `v20` are long term volatilities $a_{0i}\nu$ for two return series $\boldsymbol{r}_{1,t}$ and $\boldsymbol{r}_{2,t}$. Dash lines represent true parameters.

3.5.3. Study 3: Effect of long term correlation parameter ψ and shape parameter ν of Wishart distribution on model performance. Based on the $M = 1$ set of simulated covariance matrix series, this study evaluates the model performance at various levels of $\nu = 10, 30, 50, 100$ when $\psi = 0.8$ as in study 2.

Figures 3.5 plots the simulated $v_{ii,t}$ and fitted $\nu\lambda_{ii,t}$ variances as well as the simulated ϱ_t and fitted ρ_t correlations in (3.31) using the posterior median. As ν increases, results show that the width of the credible interval for correlation ρ_t decreases while those for variance increases. We also observe the severity of spikes for variances also decreases with increasing ν. Hence, ν reveals

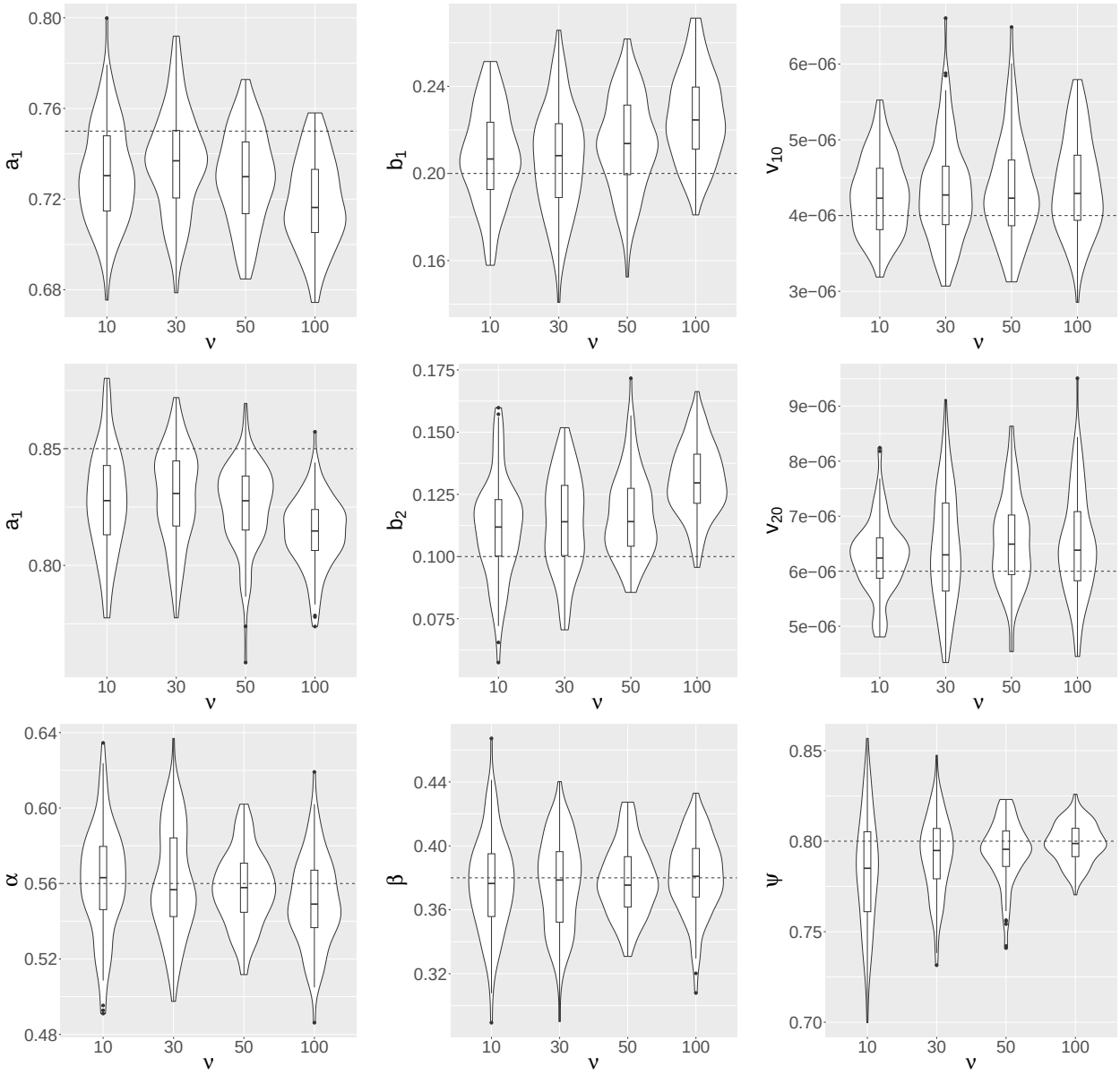


FIGURE 3.3. Violin plots of parameter estimates for $M = 100$ simulated covariance matrix series using model B for ν = 10, 30, 50, 100 and $\beta(2, 5)$ prior for a_i, b_i in simulation study 2. `v10` and `v20` are long term volatilities $a_{0i}\nu$ for two return series $\boldsymbol{r}_{1,t}$ and $\boldsymbol{r}_{2,t}$. Dash lines represent true parameters.

the characteristics of the variance and correlation series showing improvement of estimation accuracy on increasing ν. In summary, parameter ν acts like a scaling factor for the credible intervals of correlation in particular.

3.5.4. Study 4: Impact of $\alpha,\ \beta$, and $\alpha + \beta$ on the ρ_t process in matrix $\boldsymbol{Q}_t$. In this section, we use simulation to get insight into the dynamics of the $\boldsymbol{Q}_t$ matrix we define in (3.20). We first notice that

$$\boldsymbol{Q}_t|\mathcal{F}_{t-1} = (1 - \alpha - \beta)\boldsymbol{\Psi} + \alpha\widehat{\boldsymbol{\mathcal{Q}}}_{t-1} + \beta\boldsymbol{Q}_{t-1}. \tag{3.49}$$

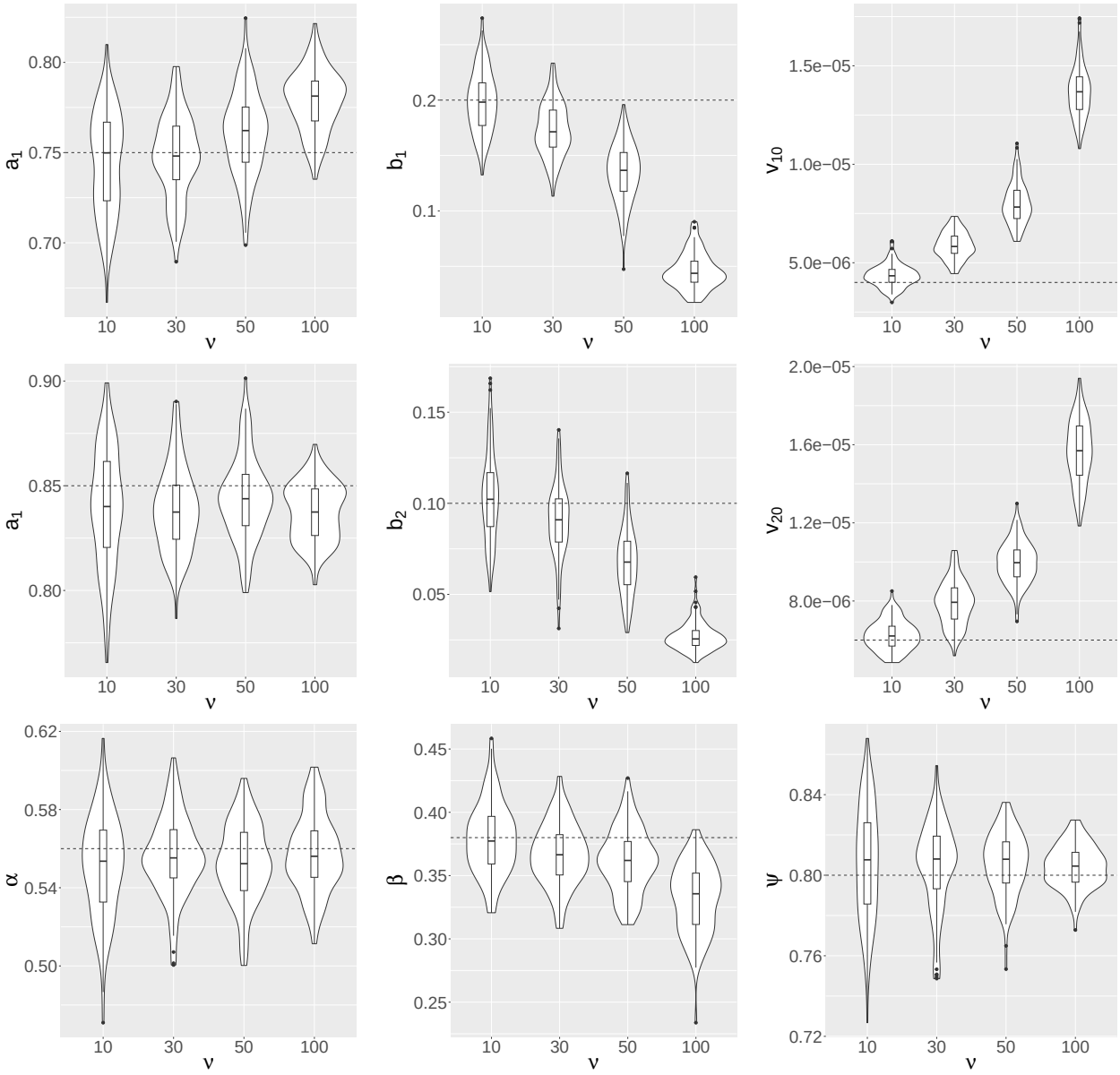


FIGURE 3.4. Violin plot of parameter estimates for $M = 100$ simulated covariance matrix series using model B with ν = 10, 30, 50, 100 and $\beta(2,2)$ prior for a_i, b_i in simulation study 2. `v10` and `v20` are long term volatilities $a_{0i}\nu$ for two return series $\boldsymbol{r}_{1,t}$ and $\boldsymbol{r}_{2,t}$. Dash lines represent true parameters.

Assume that $\mathrm{E}(\widehat{\boldsymbol{\mathcal{Q}}}_t)|\mathcal{F}_{t-1} \approx \boldsymbol{Q}_t$ and the process $\boldsymbol{Q}_t$ are stationary. As $\boldsymbol{\Psi}, \widehat{\boldsymbol{\mathcal{Q}}}_{t-1}$ and $\boldsymbol{Q}_{t-1}$ are positive definite matrices, a sufficient condition for process $\{\boldsymbol{Q}_t\}$ to be positive definite will be

$$\alpha, \beta \geq 0, \;\; \alpha + \beta \leq 1. \tag{3.50}$$

Furthermore, we have

$$\mathrm{E}(\boldsymbol{Q}_t)|\mathcal{F}_{t-2} = \boldsymbol{\Psi} + \alpha\big(\mathrm{E}(\widehat{\boldsymbol{\mathcal{Q}}}_{t-1})|\mathcal{F}_{t-2} - \boldsymbol{\Psi}\big) + \beta(\boldsymbol{Q}_{t-1} - \boldsymbol{\Psi}). \tag{3.51}$$

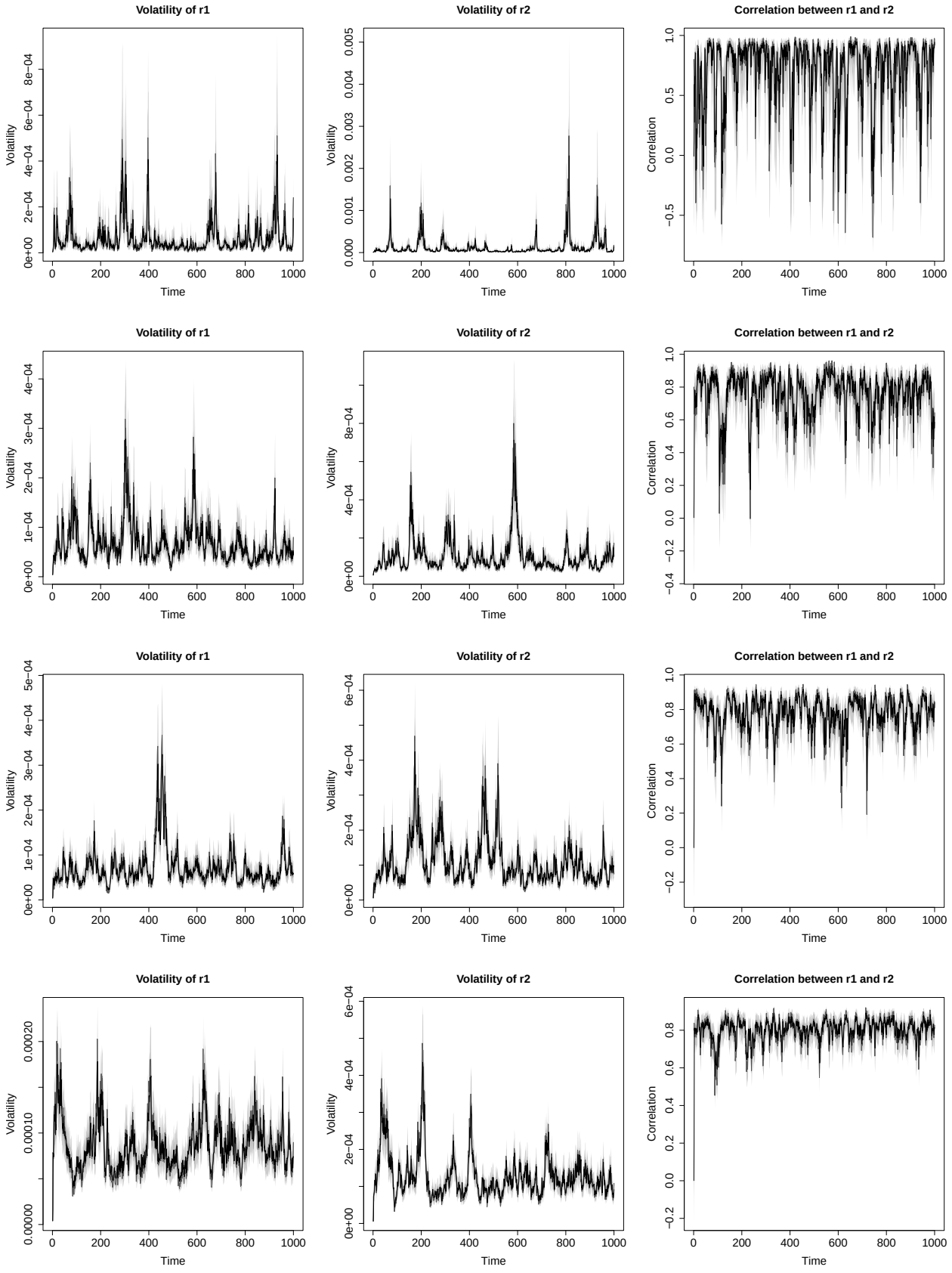


FIGURE 3.5. Observed (dark grey), fitted median (black) and 95% credible intervals (grey) for volatilities (column 1 and 2) and correlation (column 3) at $\nu = 10, 30, 50, 100$ (row 1 to 4) and $\psi_{21} = 0.8$ for simulation study 3.

Taking expectation on both side and let $\mathrm{E}(\boldsymbol{Q}_t) = \mathrm{E}(\boldsymbol{Q}_{t-1})$ as $t \to \infty$, we have

$$\mathrm{E}(\boldsymbol{Q}_t) = \boldsymbol{\Psi} \quad \text{as } t \to \infty. \tag{3.52}$$

That is, $\boldsymbol{\Psi}$ is the long-term mean of the $\boldsymbol{Q}_t$ process. But in practice, this may not be the case. Equation (3.49) suggests that $\boldsymbol{Q}_t$ is a weighted mean of $\boldsymbol{\Psi}$, $\boldsymbol{\mathcal{Q}}_{t-1}$ and $\boldsymbol{Q}_{t-1}$. When $\alpha + \beta \to 1$, the weight of $\boldsymbol{\Psi}$ is approximately 0. Thus $\boldsymbol{\Psi}$ will have little impact on the $\boldsymbol{Q}_t$ process. We will explore the dynamics of $\{\boldsymbol{Q}_t\}$ under this scenario. For simplicity, we consider a bivariate case, where $\boldsymbol{Q}_t = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}$ and $\boldsymbol{\Psi} = \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix}$. The dynamic of $\boldsymbol{Q}$ will be described by the off-diagonal term $\rho_t = \psi + \alpha(\varrho_{t-1} - \psi) + \beta(\rho_{t-1} - \psi)$, where ϱ_{t-1} is the off-diagonal entry of $\boldsymbol{\mathcal{Q}}_{t-1}$.

Simulation scheme

- Set hyperparameters of the model as below:

$$\begin{aligned} a_{01} &= 4 \times 10^{-6}, \quad a_1 = 0.3, \quad b_1 = 0.2, \\ a_{02} &= 6 \times 10^{-6}, \quad a_2 = 0.4, \quad b_2 = 0.1, \\ \psi &= -0.5, \quad \nu = 10. \end{aligned}$$

- Set the initial value at $t = 1$ of the process as below:

$$\begin{aligned} v_{11,1} &= 4 \times 10^{-5}, \quad \sigma^2_{1,1} = 4 \times 10^{-6}, \quad \varrho_1 = 0.8, \\ v_{22,1} &= 6 \times 10^{-5}, \quad \sigma^2_{2,1} = 6 \times 10^{-6}, \quad \rho_1 = 0.5. \end{aligned}$$

- For $t = 2$, we calculate $\boldsymbol{\Lambda}_t$ definded in (3.19). Then we sample a $\boldsymbol{V}_t$ from $\mathrm{W}_2(\boldsymbol{\Lambda}_t, 10)$.
- Repeat previous process for $t = 3, ..., 1000$.

Figure 3.6(a)-(c) displays the dynamic of $\{\rho_t\}$ when $\nu = 10$ for different sum $\alpha+\beta = 0.8, 0.95, 0.9999$ respectively which is gradually approaching 1. Across each row, the effect of increasing α is shown. Figure 3.6(a) demonstrates the $\{\rho_t\}$ process where $\alpha + \beta = 0.8$ is not close to 1. Given that our initial values ($\rho_1 = 0.5, \varrho_1 = 0.8$) are far way from the long-term means (-0.5), the ρ_t stations around ψ after a few transitions. As $\boldsymbol{Q}_{t-1}$ is $\mathcal{F}_{t-2}$ measurable, we have

$$\mathrm{Var}(\rho_t|\mathcal{F}_{t-2}) = \alpha^2 \mathrm{Var}(\varrho_{t-1}|\mathcal{F}_{t-2}). \tag{3.53}$$

Thus α determines the level of local variability. The higher the α, the larger the variability of ρ_t across each row. When the sum $\alpha + \beta$ is closer to 1, Figure 3.6(b) shows that the process $\{\rho_t\}$ gradually becomes more volatile due to the higher sum of 0.95. Figures 3.6(a)-(b) show that given the same sum $\alpha + \beta$ not being too close to 1, the $\{\rho_t\}$ can still station around ψ if α is small with little weight for ϱ_t in (3.49).

We are interested in the case of $\alpha + \beta \approx 1$. Figure 3.6(c) presents 3 extreme cases. In the left plot, β is more dominating and the initial period of the process is determined by the initial value ρ_1. As the proportion of α increases, $\{\rho_t\}$ becomes more volatile and no longer centres around ψ. Instead, ρ_t becomes chaotic and shows a random walk pattern. When α is large, we can observe a few periods where a sudden change in correlation occurs. We interpret this pattern as high systematic risk. While the correlation is stationary for a period, we could have a sudden change in correlation. Such a dramatic change can remain for a long period. When α is dominating in the right plot, suppringly $\{\rho_t\}$ will either moving near 1 or -1. The initial value does not carry the information for a long period as the persistent term β is small. This is a most extreme

case of systematic risk. While the correlation is quite stationary for a long period, it can reverse instantaneously. Hence, it is clear that the $\{\rho_t\}$ is non-stationary overall when the sum $\alpha + \beta = 0.9999 \approx 1$ and the local variability α is also high.

The total or unconditional variability of $\{\rho_t\}$ is also affected by β. Figure 3.6(d) and Figure 3.7(d) illustrate how β impacts the dynamics of $\{\rho_t\}$. Three processes for fixed $\alpha = 0.2$ and $\beta = 0.3, 0.6, 0.7999$ in Figure 3.6(d) share a similar scale of local variability, as they have the same α. The increasing in β contributes to the total variability of $\{\rho_t\}$, and when $\alpha + \beta \approx 1$, the $\{\rho_t\}$ process is dominated by the initial value instead of ψ. Moreover, comparing Figure 3.7(a)-(c) with Figure 3.6(a)-(c), the former has relatively larger β and so relatively smaller α given the same sum $\alpha + \beta$. These larger β and smaller α give relatively more stable $\{\rho_t\}$.

The simulation studies illustrate that our model is capable of capturing the potential systematic risk via α and β. If $\alpha + \beta$ is estimated to be close to 1, this might suggest that there is a high systematic risk. Furthermore, the larger the α, the higher the risk given that $\alpha + \beta \approx 1$.

3.5.5. Study 5: Impact of ν on the ρ_t process in matrix $\boldsymbol{Q}_t$. The impact of ν can be demonstrated by comparing Figure 3.8(a)-(d) when $\nu = 100$ and Figure 3.9(a)-(d) when $\nu = 5$ to Figure 3.6(a)-(c) when $\nu = 10$ while α, β and so their sum remain the same. With higher ν, Figures 3.8(a)-(d) display relatively more stable $\{\rho_t\}$. Conversely, with smaller $\nu = 5$, Figure 3.9(a)-(d) display more volatile dynamics of $\{\rho_t\}$ compared to Figure 3.6(a)-(d). Studies 4 and 5 lead to the study of the property of ρ_t and the proposal of ergodicity metrics.

3.5.6. Property study: The loss of ergodicity metrics. Equation (3.20) can also be rewritten in the following form:

$$\boldsymbol{Q}_t|\mathcal{F}_{t-1} = (1-\alpha-\beta)\boldsymbol{\Psi} + \alpha(\widehat{\boldsymbol{\mathcal{Q}}}_{t-1} - \boldsymbol{Q}_{t-1}) + (\alpha+\beta)\boldsymbol{Q}_{t-1}. \tag{3.54}$$

The term $\widehat{\boldsymbol{\mathcal{Q}}}_{t-1} - \boldsymbol{Q}_{t-1}$ is the random error term. Thus α and $\alpha + \beta$ determine the level of short-term and persistent term effects respectively. Note that by (3.10), $\widehat{\boldsymbol{\mathcal{Q}}}_{t-1}|\mathcal{F}_{t-2}$ follows the same distribution of sample covariance matrix of $\{\boldsymbol{R}_1, ..., \boldsymbol{R}_{\nu+1}\}$, where $\boldsymbol{R}_t \overset{i.i.d}{\sim} \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. In bivariate case, Hotelling (1953) showed that asymptotically as $\nu \to \infty$ we have

$$\mathrm{E}(\varrho_{t-1} - \rho_{t-1})|\mathcal{F}_{t-2} \sim -\frac{(1-\rho_{t-1}^2)\rho_{t-1}}{2\nu}, \tag{3.55}$$

$$\mathrm{E}(\varrho_{t-1} - \rho_{t-1})^2|\mathcal{F}_{t-2} \sim \frac{(1-\rho_{t-1}^2)^2}{\nu}. \tag{3.56}$$

Hence we can find that the local variability of ρ_t is given by

$$\mathrm{Var}(\varrho_{t-1} - \rho_{t-1}|\mathcal{F}_{t-2}) = \mathrm{E}(\varrho_{t-1} - \rho_{t-1})^2|\mathcal{F}_{t-2} - [\mathrm{E}(\varrho_{t-1}) - \rho_{t-1}|\mathcal{F}_{t-2}]^2 \sim \frac{(1-\rho_{t-1}^2)^2}{\nu}, \tag{3.57}$$

$$\mathrm{Var}(\rho_t|\mathcal{F}_{t-2}) \sim \frac{\alpha^2(1-\rho_{t-1}^2)^2}{\nu}. \tag{3.58}$$

The coefficient $\frac{\alpha}{\sqrt{\nu}}$ takes the variance of the error distribution into account, thus, it is a more adequate indicator of the local variability of process $\{\rho_t\}$.

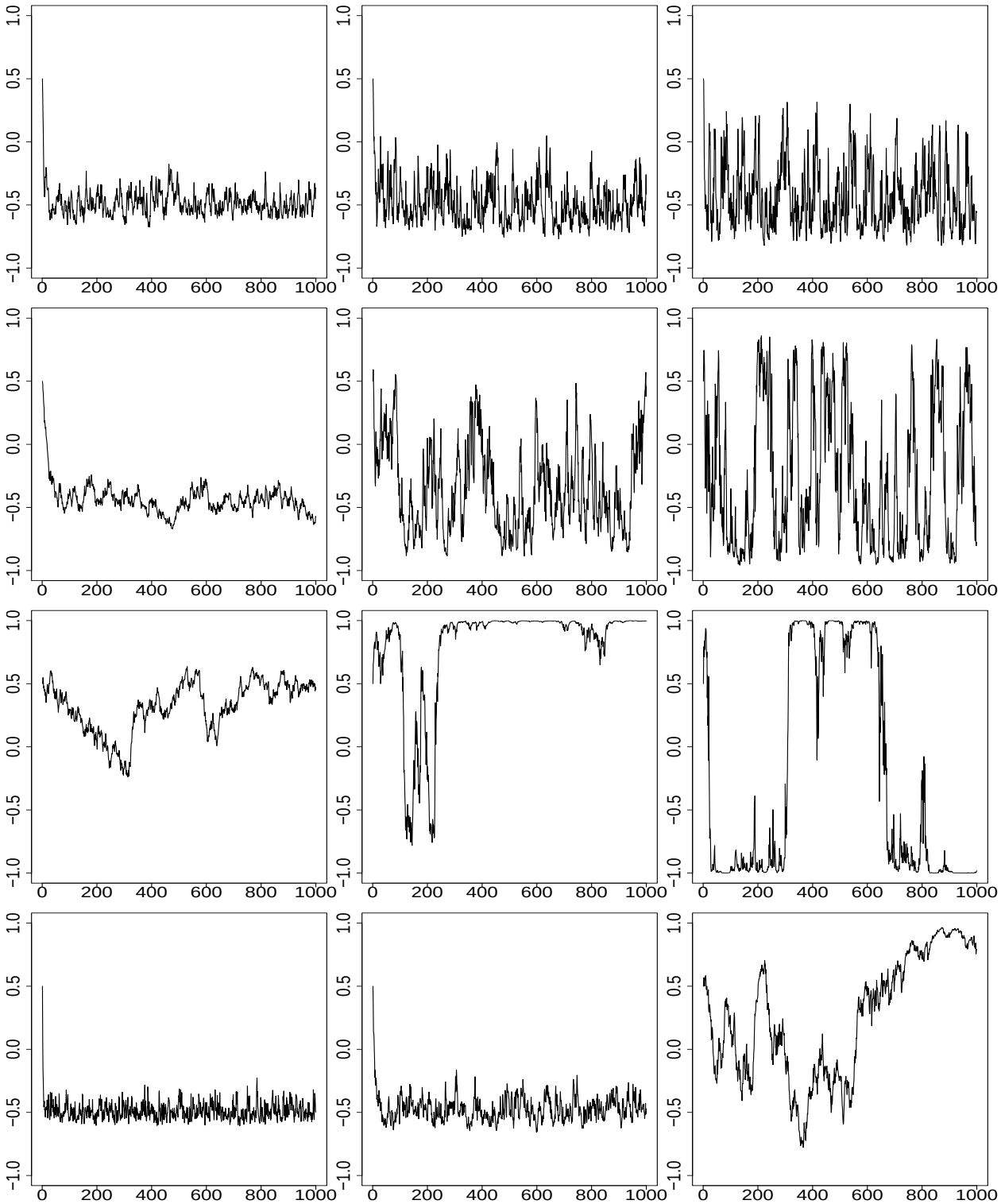


FIGURE 3.6. Dynamics of ρ_t when $\nu = 10$ and
(a) Row 1: $\alpha = 0.2, 0.4, 0.6$ and $\beta = 0.6, 0.4, 0.2$ summing to 0.8;
(b) Row 2: $\alpha = 0.1, 0.475, 0.85$ and $\beta = 0.85, 0.475, 0.1$ summing to 0.95;
(c) Row 3: $\alpha = 0.1, 0.49995, 0.8999$ and $\beta = 0.8999, 0.49995, 0.1$ summing to 0.9999;
(d) Row 4: $\alpha = 0.2, 0.2, 0.2$ and $\beta = 0.3, 0.6, 0.7999$ with increasing sum from 0.5 to 0.9999 and increasing β while $\alpha = 0.2$ is fixed.

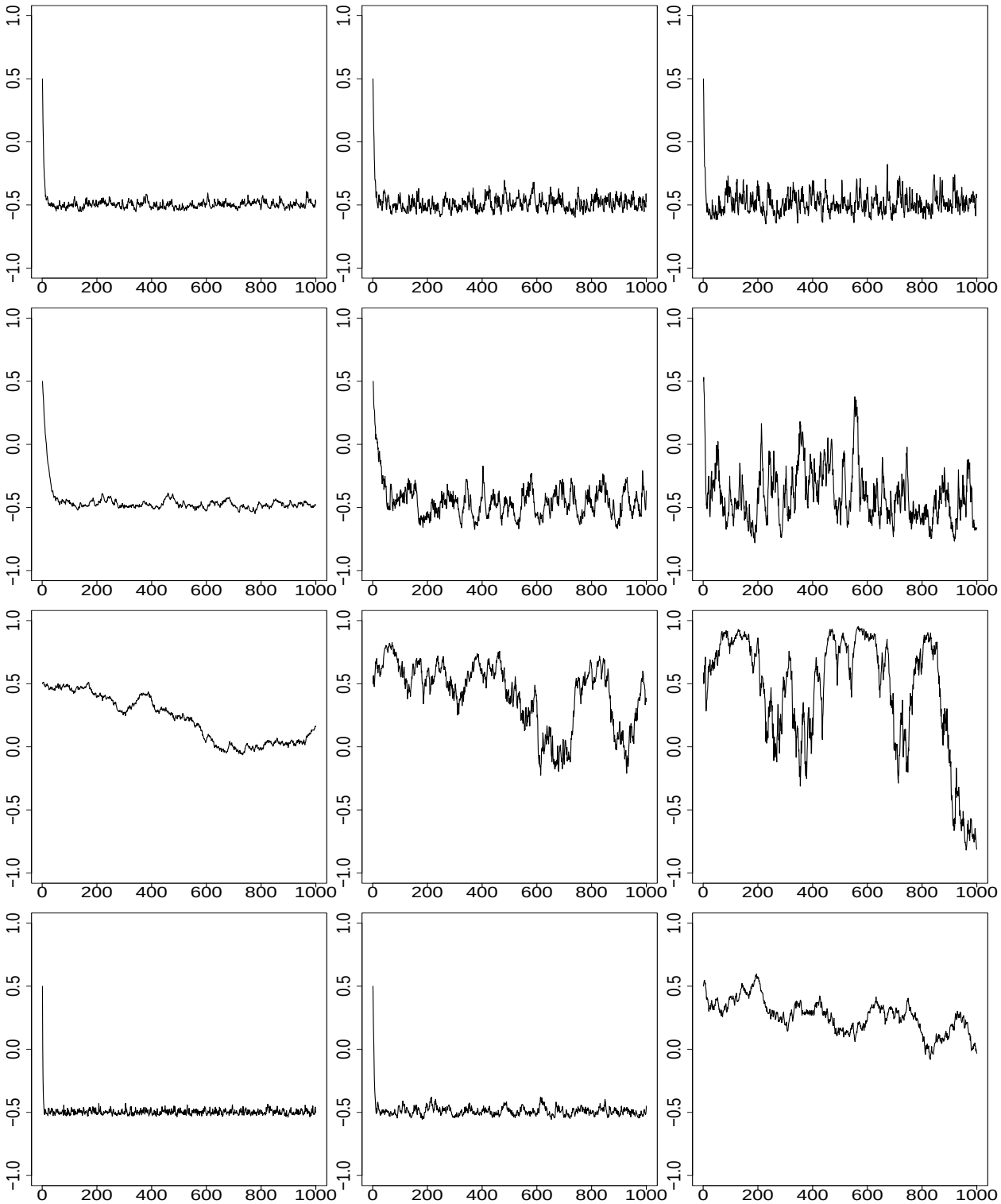


FIGURE 3.7. Dynamics of ρ_t with $\nu = 10$ when
(a) Row 1: $\alpha = 0.0632, 0.1265, 0.1897$ and $\beta = 0.7368, 0.6735, 0.6103$ summing to 0.8;
(b) Row 2: $\alpha = 0.0316, 0.1502, 0.2688$ and $\beta = 0.9184, 0.7998, 0.6812$ summing to 0.95;
(c) Row 4: $\alpha = 0.1581, 0.2846, 0.0316$ and $\beta = 0.8418, 0.7153, 0.9683$ suming to 0.9999;
(d) Row 3: $\alpha = 0.06325, 0.06325, 0.06325$ and $\beta = 0.4368, 0.7368, 0.93674$ with increasing sum of 0.5, 0.8 and 0.9999 and increasing β while $\alpha = 0.06325$ is fixed.

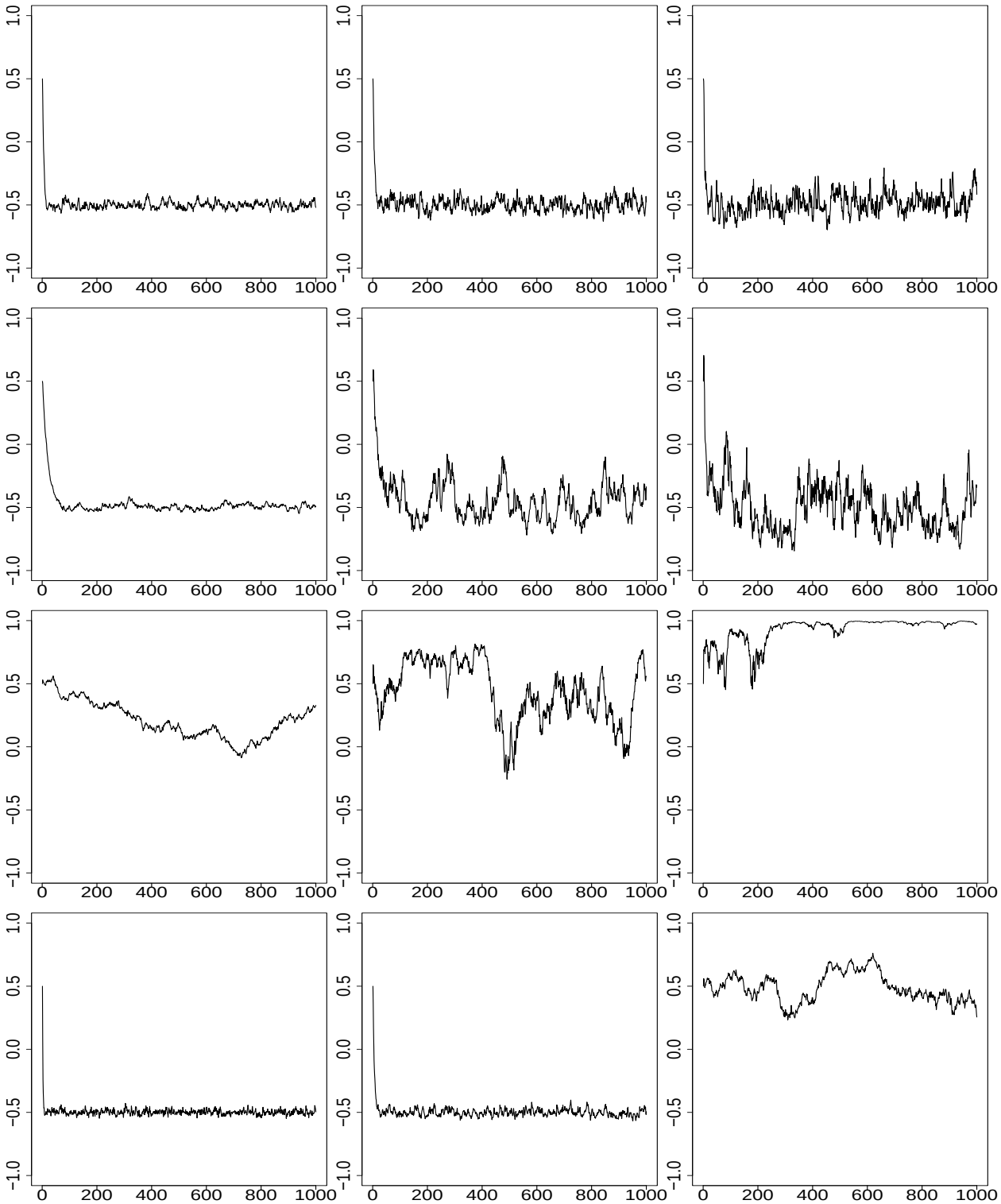


FIGURE 3.8. Dynamics of ρ_t when $\nu = 100$ and
(a) Row 1: $\alpha = 0.2, 0.4, 0.6$ and $\beta = 0.6, 0.4, 0.2$ summing to 0.8;
(b) Row 2: $\alpha = 0.1, 0.475, 0.85$ and $\beta = 0.85, 0.475, 0.1$ summing to 0.95;
(c) Row 3: $\alpha = 0.49995, 0.8999, 0.1$ and $\beta = 0.49995, 0.1, 0.8999$ summing to 0.9999;
(d) Row 3: $\alpha = 0.2, 0.2, 0.2$ and $\beta = 0.3, 0.6, 0.7999$ with increasing sum from 0.5 to 0.9999 and increasing β while $\alpha = 0.2$ is fixed.

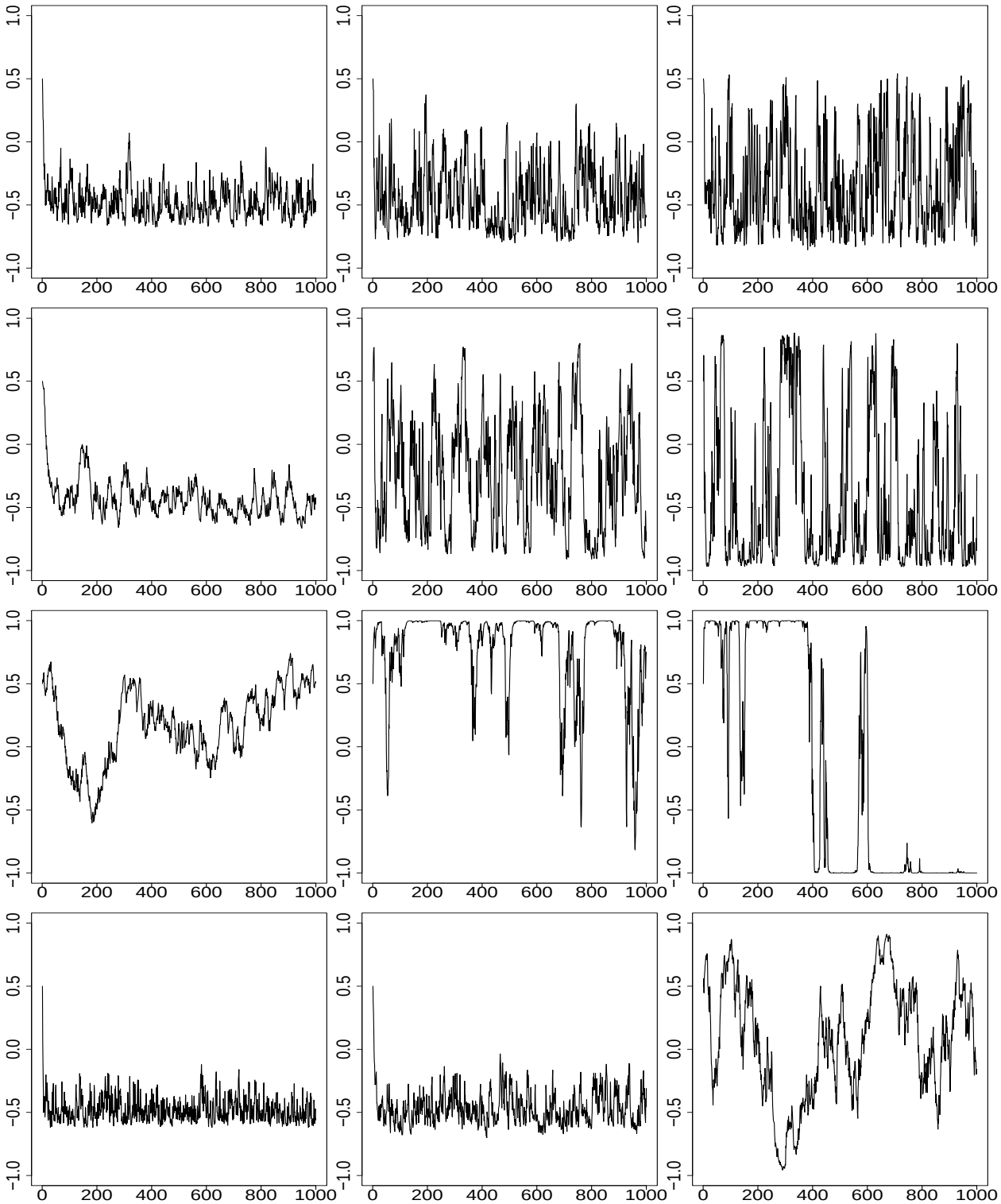


FIGURE 3.9. Dynamics of ρ_t when $\nu = 5$ and
(a) Row 1: $\alpha = 0.2, 0.4, 0.6$ and $\beta = 0.6, 0.4, 0.2$ summing to 0.8;
(b) Row 2: $\alpha = 0.1, 0.475, 0.85$ and $\beta = 0.85, 0.475, 0.1$ summing to 0.95;
(c) Row 3: $\alpha = 0.49995, 0.8999, 0.1$ and $\beta = 0.49995, 0.1, 0.8999$ summing to 0.9999;
(d) Row 3: $\alpha = 0.2, 0.2, 0.2$ and $\beta = 0.3, 0.6, 0.7999$ with increasing sum from 0.5 to 0.9999 and increasing β while $\alpha = 0.2$ is fixed.

Now lets consider two processes with hyperparameters $\{..., \alpha_1, \beta_1, ..., \nu_1\}$ and $\{..., \alpha_2, \beta_2, ..., \nu_2\}$. We expect that two processes have similar patterns if

$$\frac{\alpha_1}{\sqrt{\nu_1}} = \frac{\alpha_2}{\sqrt{\nu_2}} \tag{3.59}$$

$$\alpha_1 + \beta_1 = \alpha_2 + \beta_2. \tag{3.60}$$

Note that the square of bias decays at a rate of $\frac{1}{\nu^2}$, which is much faster than the variance ($\frac{1}{\nu}$) and justifies our approximation of $\mathrm{E}(\varrho_t)|\mathcal{F}_{t-1} \approx \rho_t$.

We propose a loss of ergodicity (LE) metric defined as

$$\mathrm{LE} = \frac{\alpha}{\sqrt{\nu}(1-\alpha-\beta)}. \tag{3.61}$$

While the metric does not apply to the case where $\alpha \approx 0, \beta \approx 1$, we find that LE is an adequate indicator for the ergodicity. A $\mathrm{LE} > 10$ suggests that the process is quite chaotic and ψ will have little impact on the dynamics of $\{\rho_t\}$. Figure 3.7 illustrates that, with a lower ν value of 10, the dynamics of $\boldsymbol{Q}$ exhibit a strong resemblance to the case of a higher ν value of 100, as shown in Figure 3.8, provided that the LE metric remains similar. This indicates that the overall correlation dynamics are not significantly impacted by the magnitude of ν when the ergodicity, as measured by the LE metric, is consistent.

3.6. Conclusion

This chapter extends the two-stage MCARR-return models of Tan et al. (2022) and Nitithumbundit and Chan (2022) (in discussion) to model the covariance matrix series directly using Wishart distribution. We study the choices of modelling the covariance matrix (model A) or the variances and correlation matrix (model B) as the mean process in Wishart distribution through simulation study 1 and find that model B provides better performance.

In study 2 using model B, we explore the choice of priors for parameters a_i, b_i to ensure stationary condition ($a_i + b_i < 1$) for variances and to provide accurate estimates. Results show that $\beta(2, 5)$ prior with lower mean, lower variance, and higher density below 0.5 is preferred over $\beta(2, 2)$ prior in providing more reasonable variance and, hence, correlation estimates. The prior for α, β is taken to be $\beta(2, 20)$ to ensure stationary condition for correlation.

In study 3, we explore the impact of the shape parameter ν of the Wishart distribution on the simulated variance and correlation series and their model performance. Results show that on increasing ν, correlation can be more precisely estimated, and variance shows fewer spikes. These three studies guide the choice of model and priors as well as reveal the interpretation of ν estimate in the way that larger ν implies systematic variation of variance with fewer spikes and less variation of correlation, in general.

Moreover, studies 4 and 5 explore the impact of α, β, their sum and ν on the stability of the correlation estimates ρ_t. Results find that the stability of ρ_t is inversely related to α and $\alpha + \beta$, and directly related ν. Hence, we propose the LE metric in (3.61), which indicates ergodicity.

Lastly, since we model the correlation directly using an ARMA process but do not specify a distribution for the correlation, the variability of the correlation is sourced from the variability of the marginal distribution for that component within the Wishart distribution. This distribution is unknown to us. For future research, we will explore some appropriate distributions to model the correlation process apart from the return process and variance process, incorporating all distinct sources of information. This direct modelling of correlation will allow us to estimate the confidence interval for the correlation more efficiently.

CHAPTER 4

Application of range-based correlation measure and two-stage MCARR-return models

4.1. Introduction

In this Chapter, we demonstrate the applicability of our proposed range-based covariance matrix measures in (2.91) and our proposed two-stage MCARR-return models in (3.26) to (3.27) and (3.28) to (3.30) in two applications. The first application focuses on US stock market indices and compares the in-sample performance of model A in (3.17) and model B in (3.18) to (3.20). The efficiency of the range measure is assessed by comparing the model fit to the sample covariance measure. Finally, we perform a 1-step-ahead VaR forecast of stock market indices, volatilities and correlations for 100 time points. The forecast performance is assessed through some performance metrics and compared to the DCC-MGARCH models implemented using the R package `rmgarch`.

Understanding economic development in stressful situations such as pandemics is a crucial task for academic research. Our second application aims to study COVID-19's impact on the financial market, with a similar aim to Nitithumbundit and Chan (2022), but they focused only on the cryptocurrency market, whereas this paper also studies indices of gold and crude oil to provide a more complete picture of the impact on financial markets. Cepoi (2020) studied the nonlinear impact of gold price dynamic on equity markets (Dow Jones Industrial Average (DJIA), Financial Times Stock Exchange (FTSE) 100, Deutscher Aktienindex (DAX), Cotation Assistée en Continu French (CAC) 40, Madrid Stock Exchange General Index (IGBM), and Milano Indice di Borsa (MIB)) using quantile regression. However, he only considered a short period, from 3 February 2020 to 17 April 2020. Our analyses consider three periods representing before, during and post-COVID-19 pandemic.

Other studies of pandemic effects include Baur (2020) who studied the impact of COVID-19 on the fine art market based on five financial markets, namely DAX, DJIA, Japanese Nikkei, UK FTSE and South Africa Johannesburg Stock Exchange (JSE), using standard time series models such as ARIMA and GARCH which are known to be less efficient than our proposed multivariate range-based models. Moreover, Zhang et al. (2020) applied eight weekly correlation analyses during February and March 2020 between twelve countries such that analysis within each week uses daily returns of that week. However, correlations based on relatively short observed returns contain larger noises. These noises can be smoothed out by our proposed MCARR models revealing more clearly the underlying dynamics of the correlation process. We show that our proposed models outperform DCC-GARCH models.

The remaining chapter is organised as follows. Section 4.2 provides an application of the proposed two-stage MCARR-return models to analyse US stock market indices. It provides exploratory data analysis, comparison between models A and B defined in Section 3.3.6, out-of-sample forecasting and the comparison of forecasts with DCC-GARCH. Similarly, Section 4.3 applies the proposed model to study the COVID-19 impact on crytocurrency, gold and oil markets

and study the change of patterns for returns, volatilities and correlation across the pre, during and post COVID-19 periods. Lastly, Section 4.4 concludes the analyses and discusses further research work.

4.2. Application to stock market indices for forecasting

4.2.1. Data. The data employed in this research consists of three prominent stock market indices: the National Association of Securities Dealers Automated Quotations (NASDAQ; denoted as index 1), the Standard and Poor's 500 (S&P 500; denoted as index 2), and the Dow Jones Industrial Average (DJIA; denoted as index 3). These indices are among the most widely recognised and closely monitored indicators of global financial markets, providing valuable insights into market dynamics and economic conditions.

The NASDAQ is an electronic stock exchange index comprising over 3,300 publicly listed companies, with a particular emphasis on technology and growth-oriented firms. It serves as a vital benchmark for the performance of high-growth technology sectors, including companies in software, semiconductors, and biotechnology. The S&P 500, on the other hand, is a market capitalisation-weighted index that tracks the performance of the 500 largest publicly traded companies in the United States. Its constituents span a broad range of industries, including healthcare, technology, finance, and consumer goods, making it a widely accepted measure of the overall health and diversity of the U.S. economy. In contrast, the DJIA is a price-weighted index that includes 30 large-cap, blue-chip companies that are considered leaders in their respective industries. Iconic corporations such as Coca-Cola, Apple, and IBM are among their constituents, and the index serves as a bellwether for the performance of established industry-leading firms. Together, these indices serve as crucial indicators of market performance and are extensively used by investors, analysts, and policymakers to evaluate economic trends and investor sentiment.

For our analyses, we collect daily data for these indices from the beginning of 2019 to the end of 2022, using the TradingView platform. The missing of daily return data in our analysis is primarily due to market closures during weekends and public holidays. To address this, we exclude these missing data points from the dataset. After this preprocessing step, a total of 1,258 time points remain. From this dataset, we designate the final 100 observations as the forecast period to evaluate the out-of-sample performance of the models.

In addition, it was observed that 57 correlation matrices estimated using the range-based correlation measure are not positive definite, which is a necessary property for valid covariance matrices in financial time series modelling. To rectify this issue, we apply the NPD projection method in (2.55) and (2.56) to ensure that these correlation matrices satisfy the required mathematical properties. This adjustment guarantees the stability and consistency of subsequent analyses involving these matrices. Within each trading day, 15-minute OHLC data were recorded at multiple instances, providing detailed intraday market movements. We operate under the assumption that volatility remains stationary within a single day, which enables the evaluation of daily covariance matrices using aggregated intraday data.

To examine the relationships among the three indices, we compute the sample correlations of daily returns. The mean sample correlations are calculated as 0.89, 0.71, and 0.91 for the index pairs (1,2), (1,3), and (2,3), respectively. These values reflect a high degree of co-movement between the indices, particularly between NASDAQ and DJIA (0.91) and between NASDAQ and S&P 500 (0.89), indicating significant market integration and interdependence. Based on these findings, we choose to utilise the weighted average correlation estimator of order 2 (that is, $p = 2$ and $\mathfrak{m} = 2.4$) in (2.50) for the range-based correlation estimator, as this approach offers a more

refined measure of correlation dynamics. Then $\boldsymbol{V}_t = \widehat{\boldsymbol{\Sigma}}_t$ for the MCARR model is again given by (2.91), (2.53) and (2.57). For robustness, we also compute the sample covariance and sample correlation matrices to serve as benchmarks for comparison.

This detailed methodology ensures a comprehensive and accurate representation of daily volatility and correlation structures, facilitating a robust analysis of market behaviour across the three indices. By leveraging high-frequency intraday data, we improve the precision of the covariance estimates, which are critical for the subsequent stages of our analysis. Specifically, let $(O_{i,t,\tau}, H_{i,t,\tau}, L_{i,t,tau}, C_{i,t,\tau})$ be the opening, high, low, and close of the stock i on day t at time τ and let $(o_{i,t,\tau}, h_{i,t,\tau}, l_{i,t,\tau}, c_{i,t,\tau})$, $\boldsymbol{o}_{i,t}, \boldsymbol{h}_{i,t}, \boldsymbol{l}_{i,t}, \boldsymbol{c}_{i,t}$, and $\boldsymbol{\kappa}_{i,t}$ be defined in a similar fashion. Let n_t be the number of observations on day t. Then we have the following construction for daily data.

$$v_{ii,t}^2 = \sum_{\tau=1}^{n_t} \frac{(h_{i,t,\tau} - l_{i,t,\tau})^2}{4\log 2}, \qquad v'^2_{ii,t} = \frac{n_t}{n_t - 1} \sum_{\tau=1}^{n_t} (c_{i,t,\tau} - \bar{c}_{i,t,\tau})^2, \tag{4.1}$$

$$\varrho_{ij,t} = \mathrm{cor}_{1.4}(r_{i,t}, r_{j,t}), \qquad \varrho'_{ij,t} = \mathrm{cor}_1(r_{i,t}, r_{j,t}), \tag{4.2}$$

$$\boldsymbol{\mathcal{Q}}_t = \begin{pmatrix} 1 & \varrho_{12,t} & \cdots & \varrho_{1d,t} \\ \varrho_{12,t} & \ddots & \vdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ \varrho_{1d,t} & \cdots & \cdots & 1 \end{pmatrix}, \quad \boldsymbol{\mathcal{Q}}'_t = \begin{pmatrix} 1 & \varrho'_{12,t} & \cdots & \varrho'_{1d,t} \\ \varrho'_{12,t} & \ddots & \vdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ \varrho'_{1d,t} & \cdots & \cdots & 1 \end{pmatrix}. \tag{4.3}$$

We examine the following four estimates for the observed covariance matrix

$$\boldsymbol{V}_t = \mathrm{diag}(v_{11,t}, \ldots, v_{dd,t}) \boldsymbol{\mathcal{Q}}_t \, \mathrm{diag}(v_{11,t}, \ldots, v_{dd,t}), \tag{4.4}$$

$$\boldsymbol{V}_t^{\cdot\prime} = \mathrm{diag}\,(v_{11,t}, \ldots, v_{dd,t}) \boldsymbol{\mathcal{Q}}'_t \, \mathrm{diag}(v_{11,t}, \ldots, v_{dd,t}), \tag{4.5}$$

$$\boldsymbol{V}_t^{\prime\cdot} = \mathrm{diag}(v'_{11,t}, \ldots, v'_{dd,t}) \boldsymbol{\mathcal{Q}}_t \, \mathrm{diag}(v'_{11,t}, \ldots, v'_{dd,t}), \tag{4.6}$$

$$\boldsymbol{V}_t'' = \mathrm{diag}(v'_{11,t}, \ldots, v'_{dd,t}) \boldsymbol{\mathcal{Q}}'_t \, \mathrm{diag}(v'_{11,t}, \ldots, v'_{dd,t}). \tag{4.7}$$

We assess the performance of each estimate via the model fit of the return process in Section 4.2.4.

4.2.2. Exploratory data analysis. Table 4.1 presents the summary statistics for $\boldsymbol{R}_t$ (returns) and $\boldsymbol{V}_t$ (covariance matrices). A detailed examination of the statistics reveals that the sample variance of r_i is approximately equal to the sample mean of v_{ii}. This alignment indicates the consistency of the PK measure in capturing the variance dynamics of the underlying asset returns. Such consistency reinforces the validity of the PK measure as a robust tool for estimating variances in a time-varying financial setting. Furthermore, the sample means of the correlation elements (0.88, 0.66, 0.86) are closely aligned with the sample correlations (0.89, 0.71, 0.91) reported in Section 4.2.1. This observation highlights the efficiency of the proposed range-based correlation measure in accurately capturing the dependence structure between asset pairs, even in the presence of dynamic and potentially volatile market conditions. These findings collectively validate the robustness of the PK measure and the range-based correlation estimators in providing consistent and efficient estimates of variance and correlation.

From the return time series plot presented in Figure 4.1, it can be observed that the mean of the returns is approximately zero. Moreover, the volatility in the return series is considerably more pronounced compared to the drift component, which confirms that the no-drift assumption of the PK measure remains valid in this context. In addition, all return time series plots exhibit clear signs of heteroscedasticity, indicating that the variance of returns is not constant over time. Notably, the NASDAQ index appears to exhibit higher volatility compared to the S&P 500 and the DJIA, suggesting that the NASDAQ is subject to greater variability as shown in Table 4.1, likely

due to its higher concentration of technology and growth-oriented firms. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of returns in Figure 4.2(a) and (b), respectively, reveal minimal evidence of moving average effects in the return series for all three indices. This suggests that the return series for NASDAQ, S&P 500, and DJIA exhibit little serial correlation, which aligns with the widely observed properties of financial time series data.

	Mean	Variance	Min	Max
r_1	4.8264‡	1.8011‡	-0.0888	0.0724
r_2	3.7753‡	1.2455‡	-0.0831	0.0838
r_3	3.0387‡	1.1664‡	-0.0871	0.0820
v_{11}	1.8011‡	0.1151†	4.1637†	0.0043
v_{22}	1.2140‡	0.1092†	2.1203†	0.0046
v_{33}	0.9784‡	0.0873†	2.1976†	0.0042
v_{12}	1.3545‡	0.1010†	0.9160†	0.0044
v_{13}	1.0215‡	0.0788†	-6.4615†	0.0039
v_{23}	0.9880‡	0.0839†	1.9301†	0.0040
ϱ_{12}	0.8827	0.0100	0.0609	0.9965
ϱ_{13}	0.6596	0.0492	-0.3195	0.9866
ϱ_{23}	0.8575	0.0099	0.2499	0.9940

†: $\times 10^{-6}$; ‡: $\times 10^{-4}$.

TABLE 4.1. Summary statistics of the observed returns r_i in $\boldsymbol{R}_t$; variance v_{ii} and covariance v_{ij} in $\boldsymbol{V}_t$; and correlation ϱ_{ij} in $\widehat{\boldsymbol{\mathcal{Q}}}$ for NASDAQ, S&P 500, and DJIA indices.

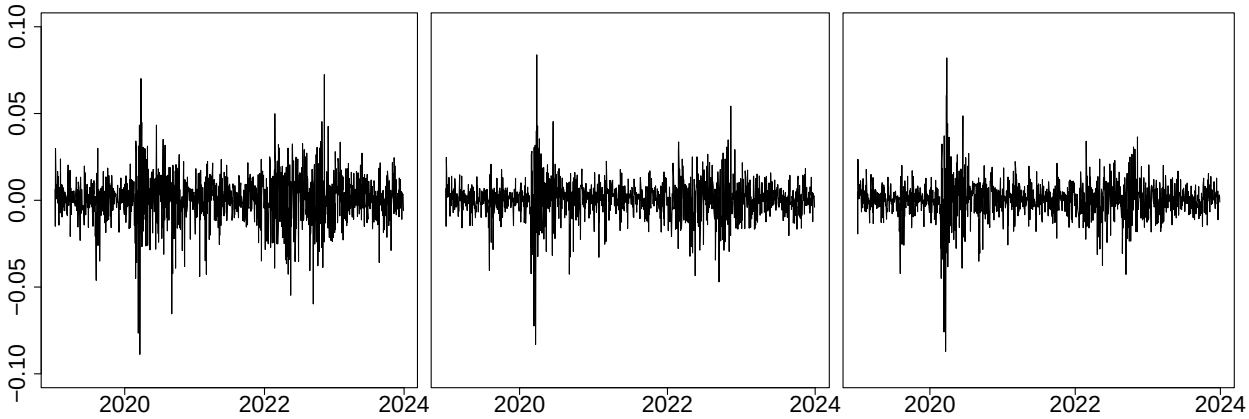


FIGURE 4.1. Plot of return time series for NASDAQ, S&P 500 and DJIA indices respectively.

Comparing the time series of volatility estimates using PK and sample variance in Figure 4.3(a) and (b) respectively, we observe greater noises for sample variance estimates although both two estimates are capable of capturing the volatility, as the jump in Figure 4.3(a) and (b) coincide with the high volatility period shown in Figure 4.1. As for the correlation, Figure 4.4(a) and (b) show that the two correlation estimates are consistent, but the sample correlation tends to display

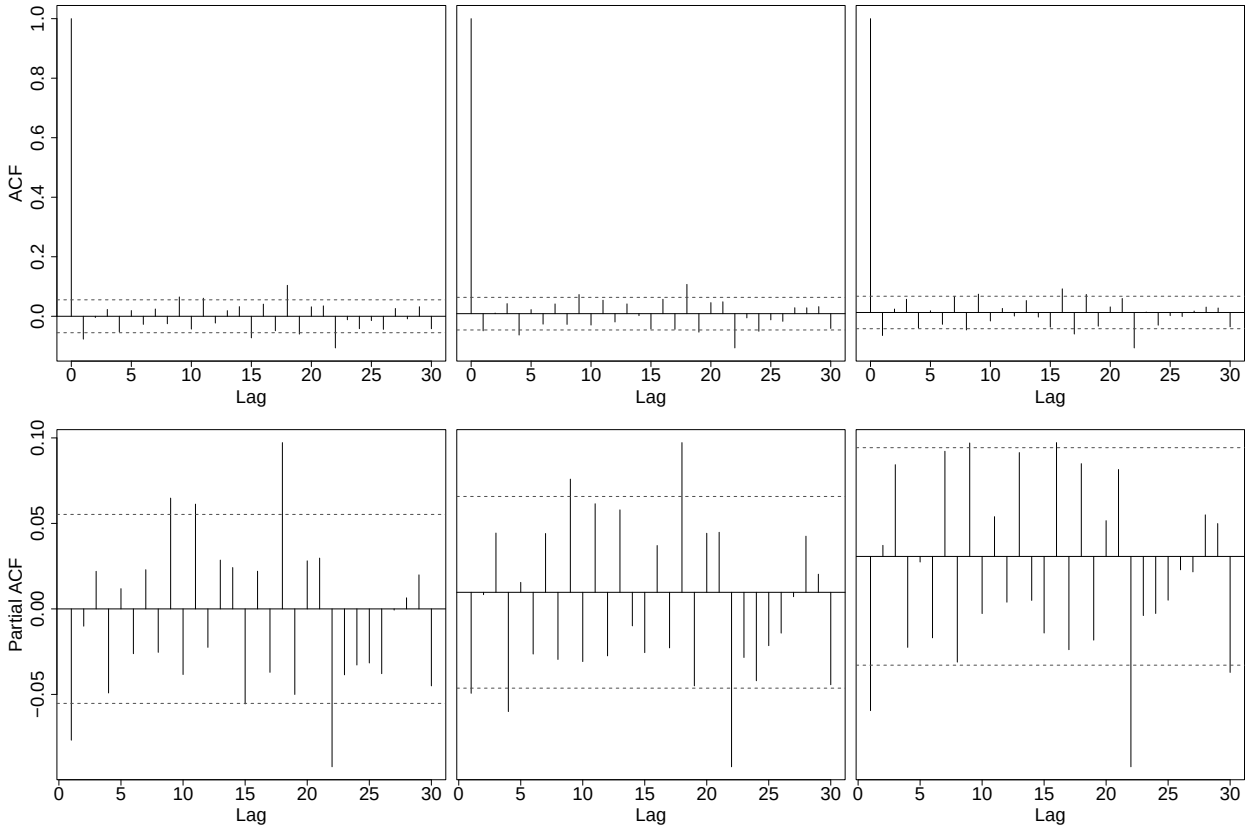


FIGURE 4.2. Plot of (a) ACF and (b) PACF of NASDAQ, S&P 500 and DJIA return time series respectively.

greater noise too. Meanwhile, the dynamics in the two correlation measures are similar. We will further discuss the choice of correlation measure in the next section. The ACF and PACF plots of $\boldsymbol{V}_t$ are summarised in Figure 4.5 and Figure 4.6. Unlike the return series, we observe much stronger persistence in volatilities, correlations and covariances, while the persistence of correlation is relatively lower, and the AR effect is marginal.

To assess the statistical properties of the return series, we conduct a Ljung–Box test to evaluate the independence of the series. This test is particularly important for return data when the ACF and PACF often fail to provide sufficiently conclusive information about the presence of serial correlation due to the inherent noise in the return series. The results of the Ljung–Box test confirm the presence of some degree of dependency in the return series, which necessitates the inclusion of mean functions in the return model that can capture this structure. Accordingly, we selected an AR(1) process to model the return series, as it can adequately account for the observed partial autocorrelation while achieving a parsimonious model. Furthermore, we perform the Augmented Dickey-Fuller (ADF) test (Paparoditis and Politis, 2018) as well as Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test (Kwiatkowski et al., 1992) to formally test for stationarity in a return series. The results of these tests are summarised in Table 4.2. The ADF test results indicate that all time series are stationary, as evidenced by the rejection of the null hypothesis of a unit root. Stationarity is a critical property for valid time series modelling, ensuring that the statistical properties such as mean and variance remain constant over time. To test these similary

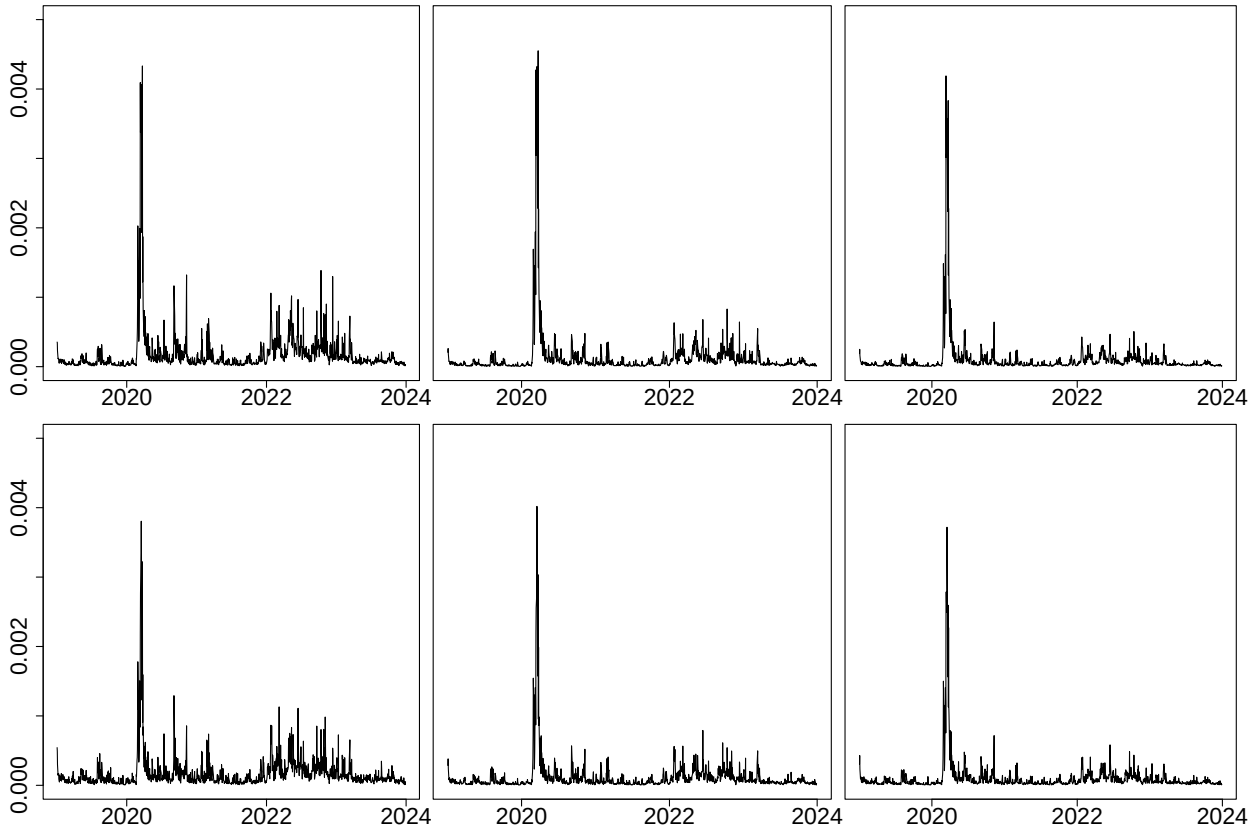


FIGURE 4.3. Plot of (a) PK and (b) sample volatility measure series for NASDAQ, S&P 500 and DJIA indices.

stationarity properties, KPSS tests the nonstationarity of a time series with level stationary (constant mean over time) or trend stationary (a deterministic trend with stationary residuals). Since the p-values of the KPSS test exceed 0.1 for all time series except the correlation series, we conclude that there is no unit root present in the return, volatility, and covariance time series. When modelling volatility, we adopt a GARCH(1,1) specification. Hansen and Lunde (2005) demonstrated that there is no significant evidence to suggest that more complex GARCH models with higher orders, such as GARCH(p, q) for $(p, q > 1)$, consistently outperform the standard GARCH(1,1) model in terms of forecasting volatility. Given its parsimony and robust performance in empirical applications, the GARCH(1,1) specification provides a reliable and computationally efficient framework for modelling the volatility dynamics in our data.

4.2.3. Parametrisation and prior of parameter. We consider two 2-stage MCARR-return models: model A (also called covariance model in this chapter) which models the covariance matrix directly or model B (called correlation model) which models the variances and correlation matrix. A necessary condition for model A to be stationary is that the magnitude of all eigenvalues of $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ in (3.17) is less than 1. Such a condition is not only difficult to implement but also not sufficient for stationarity. The `Rstan` will automatically abort the Markov chain if any eigenvalues in $\boldsymbol{\Lambda}_t$ exceed the range of double-precision floating-point ($\approx 1.8 \times 10^{308}$). While it will take more computational time, we do not set constraints for $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ based on stationarity assumption. As we discuss in Section 3.3.5, matrices $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ are not unique. Especially, the

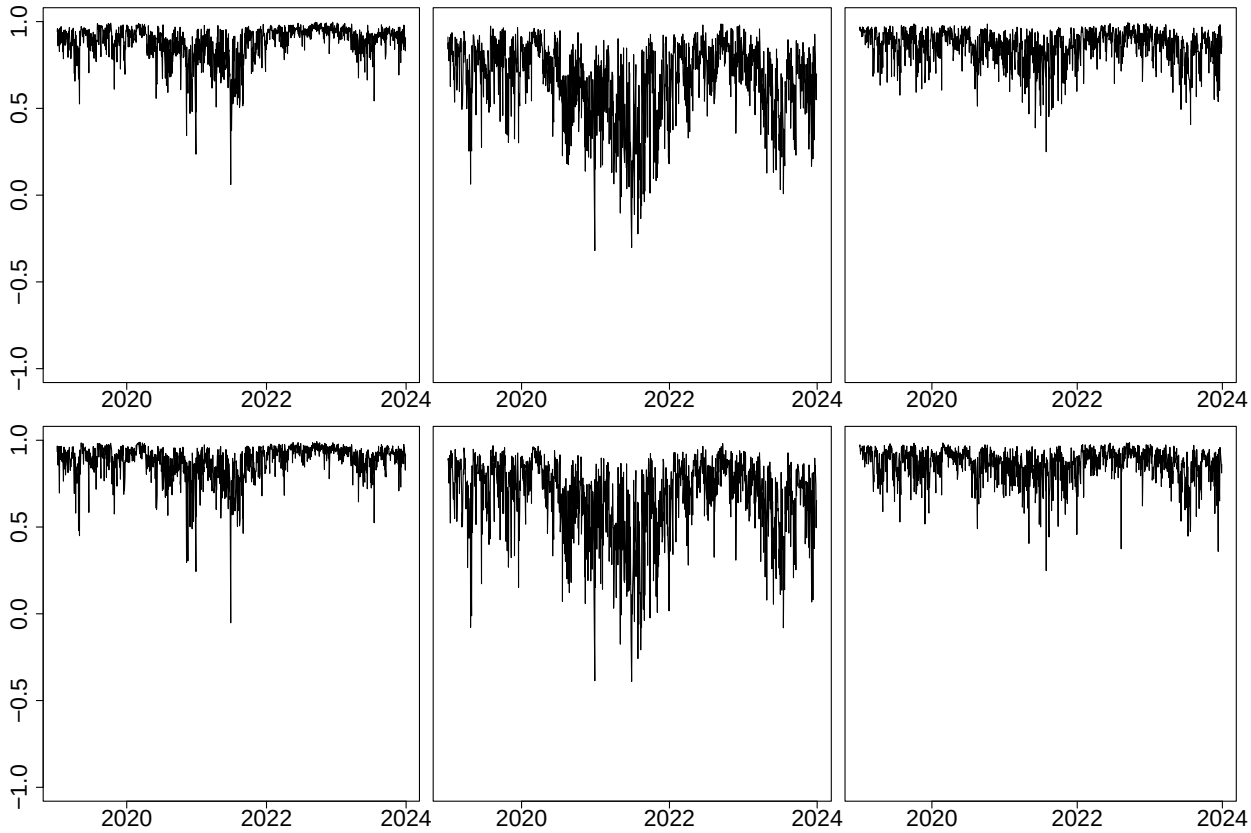


FIGURE 4.4. Plot of (a) range and (b) sample correlation measure series for NASDAQ, S&P 500 and DJIA respectively.

TABLE 4.2. Test results of Ljung–Box test and Augmented Dickey-Fuller test for NASDAQ, S&P 500 and DJIA return series.

	Ljung–Box test		ADF test		KPSS Test	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
r_1	7.4131	0.0065	-10.1673	< 0.01	0.0718	> 0.1
r_2	4.2638	0.0389	-10.4057	< 0.01	0.0644	> 0.1
r_3	7.7234	0.0055	-10.5992	< 0.01	0.0381	> 0.1
v_{11}	792.7418	0.0000	-6.1168	< 0.01	0.2084	> 0.1
v_{22}	880.5685	0.0000	-6.1755	< 0.01	0.2670	> 0.1
v_{33}	893.5316	0.0000	-6.6182	< 0.01	0.3740	> 0.1
v_{12}	824.8397	0.0000	-6.1619	< 0.01	0.2085	> 0.1
v_{13}	815.4499	0.0000	-6.2668	< 0.01	0.2722	> 0.1
v_{23}	857.5917	0.0000	-6.4674	< 0.01	0.3234	> 0.1
ϱ_{12}	350.3004	0.0000	-4.7894	< 0.01	1.6968	< 0.01
ϱ_{13}	329.5272	0.0000	-5.1095	< 0.01	1.3337	< 0.01
ϱ_{23}	138.9016	0.0000	-6.3144	< 0.01	0.8941	< 0.01

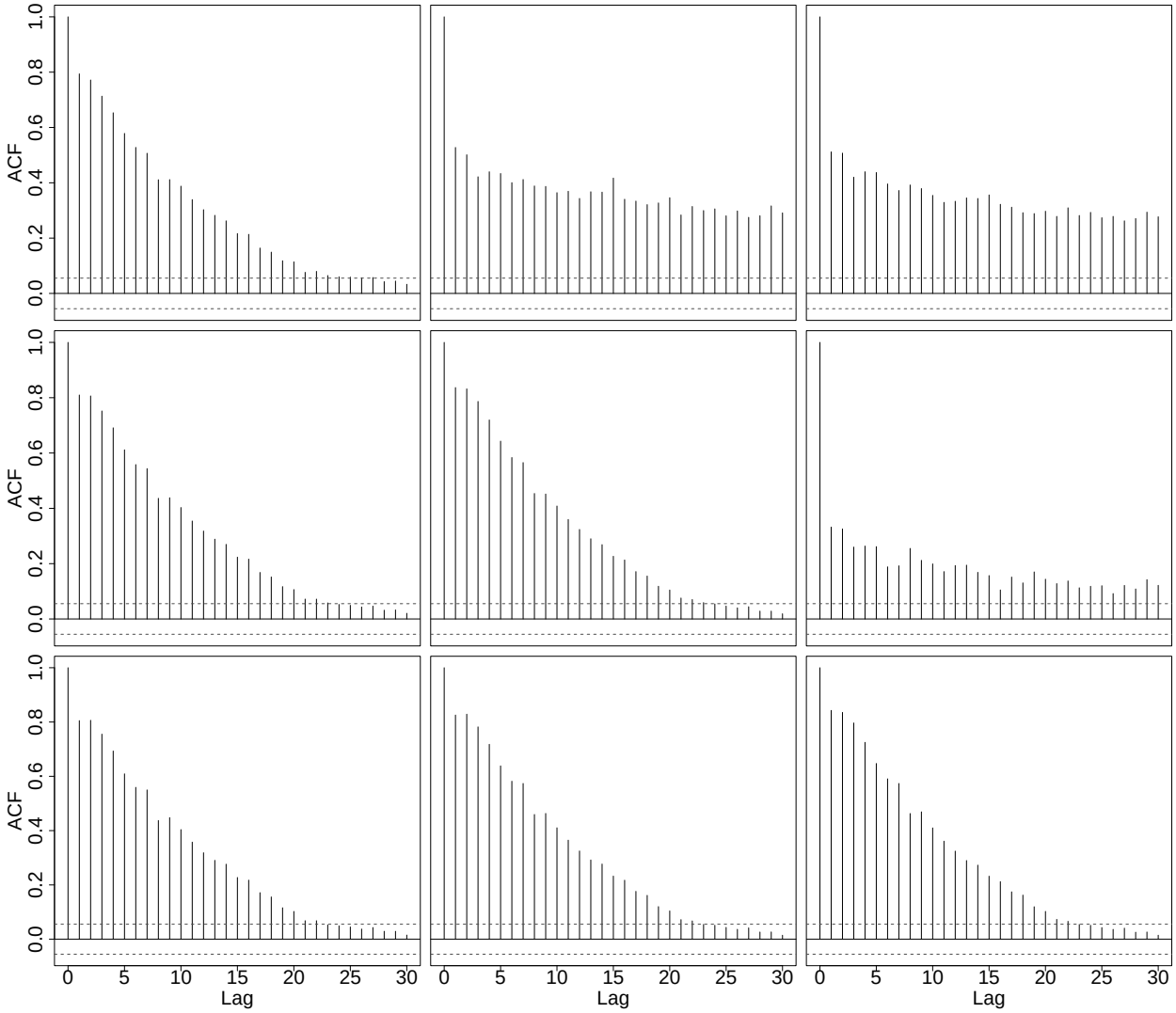


FIGURE 4.5. ACF plots of NASDAQ, S&P 500 and DJIA indices. The diagonals are plots for observed volatilities. The plots in the upper triangle are for observed correlations, while the plots in the lower triangle are for observed covariances.

equation (3.18) is invariant by taking transformation

$$\boldsymbol{A}_1 = -\boldsymbol{A}_1.$$

Thus, if we assume a flat prior or symmetric prior, then the posterior will be symmetric around 0. In such cases, our marginal posterior distribution would be bimodal. To accomplish such an issue, we set the first entry of $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ to be positive. The priors of model A, the covariance model, are given by (3.32) to (3.33) and are restated as below:

$$\begin{aligned}
a_{11} &\sim \text{HN}(0,1), \;\; a_{ij} \sim \text{N}(0,1), \quad 1 \le i,j \le 3, \quad (i,j) \ne (1,1), \\
b_{11} &\sim \text{HN}(0,1), \;\; b_{ij} \sim \text{N}(0,1), \quad 1 \le i,j \le 3, \quad (i,j) \ne (1,1), \\
\boldsymbol{A}_0 &\sim \text{W}(10, \widehat{\boldsymbol{\Lambda}}/10), \quad \nu \sim \gamma(15,1)
\end{aligned}$$

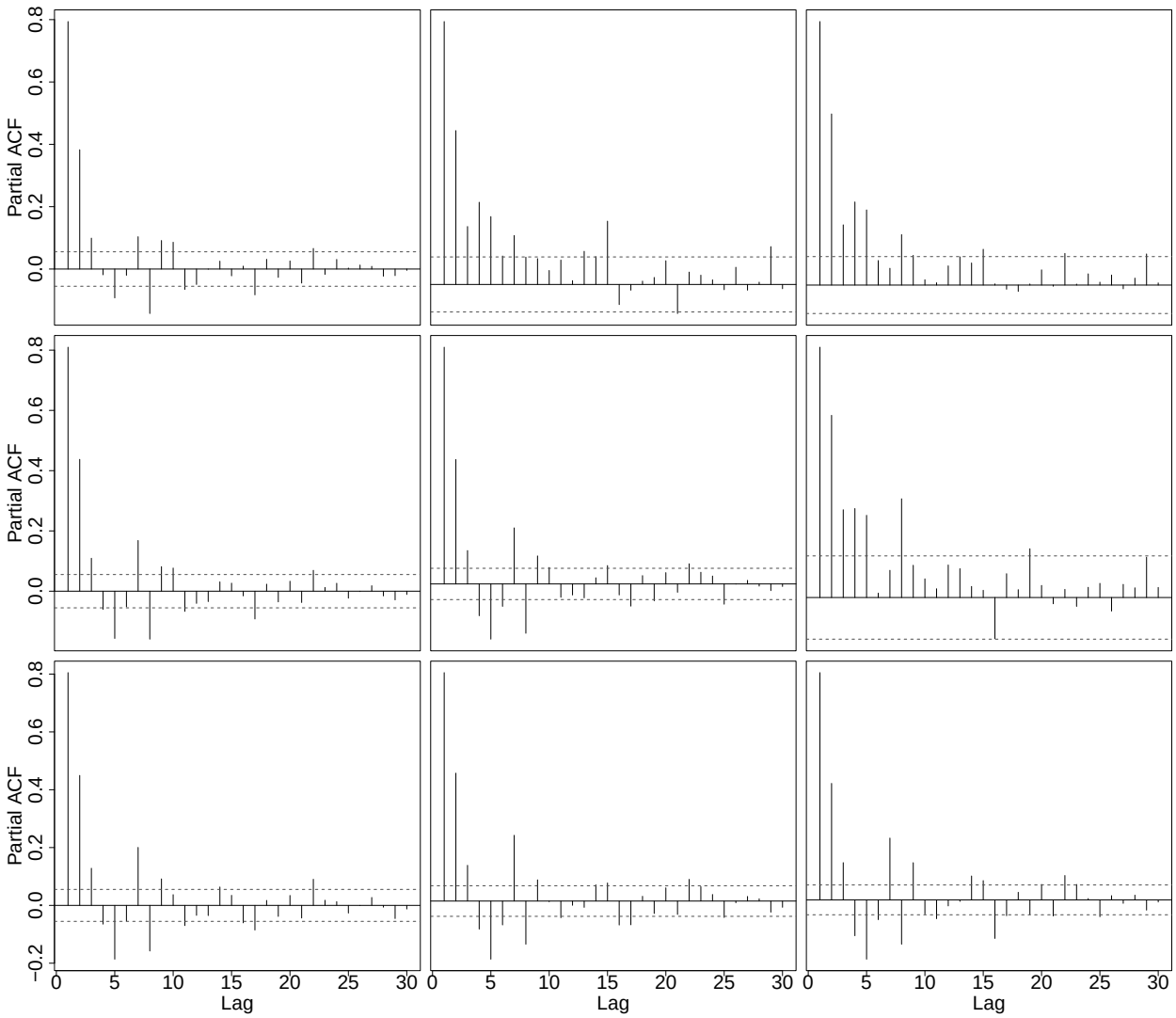


FIGURE 4.6. PACF plots of NASDAQ, S&P 500 and DJIA indices. The diagonals are plots of observed volatilities. The plots in the upper triangle are for observed correlations, while the plots in the lower triangle are for observed covariances.

where HN is the half-normal distribution and $\widehat{\mathbf{\Lambda}}$ is the sample covariance matrix of daily return during the training period.

For model B (correlation model defined in (3.18) to (3.20)), the necessary and sufficient conditions for a stationary and valid volatility process are

$$a_i > 0, \quad b_i > 0, \quad a_i + b_i < 1, \quad i = 1, 2, 3 \tag{4.8}$$

in (3.19). To implement such constraints, we make use of the unit simplex data type. A unit d-simplex is the d-dimention vector $\boldsymbol{x} = (x_1, ..., x_d)^\top \in \mathbb{R}^d$ such that

$$x_i > 0, \quad \text{and} \quad \sum_{i=1}^{d} x_i = 1.$$

For the prior distribution of a simplex, a popular choice is the Dirichlet distribution. The Dirichlet distribution is a family of continuous multivariate probability distributions commonly used as a prior in Bayesian statistics, particularly for problems involving probabilities or proportions. It is parameterised by a vector of positive real numbers $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_d)^\top$, called *concentration* parameters, which determine the distribution's shape. If a random vector follows a Dirichlet distribution, its components are non-negative and sum to one, making it particularly suitable for describing the distribution of a simplex. If $\boldsymbol{x} \sim \text{Dir}(\alpha_1, \ldots, \alpha_d)$, then the marginal distribution of x_i is given by

$$x_i \sim \text{Beta}(\alpha_i, \sum_{j=1}^{d} \alpha_j - \alpha_i). \tag{4.9}$$

It is often referred to as the multivariate generalisation of Beta distribution, as Beta distribution can be viewed as the marginal distribution of Dirichlet distribution. The same formulation also applies to α and β. Furthermore, as we discussed in Section 3.5.4, we expect the sum $\alpha + \beta$ to be away from 1. Thus, we propose a stronger prior for $(\alpha, \beta, 1 - \alpha - \beta)$ by ensuring the concentration parameter of $1 - \alpha - \beta$ to be significantly larger than the concentration parameters of α and β. Hence, the priors of model B, the correlation model, have now differed from (3.34) and (3.35) and are given by

$$(a_i, b_i, 1 - a_i - b_i) \sim \text{Dir}(2, 2, 3),\ i = 1, 2, 3, \quad a_{0i} \sim \text{Gamma(0.5,1)},\ i = 1, 2, 3, \tag{4.10}$$

$$(\alpha, \beta, 1 - \alpha - \beta) \sim \text{Dir}(2, 2, 18), \qquad \nu - 4 \sim \text{Gamma}(5, 1), \tag{4.11}$$

$$\psi_{12} \sim \text{N}(0.6, 0.2), \quad \psi_{13} \sim \text{N}(0.6, 0.2), \quad \psi_{23} \sim \text{N}(0.6, 0.2) \tag{4.12}$$

where $\text{Dir}(\boldsymbol{\alpha})$ is the Dirichlet distribution. We choose the multivariate Student-t distribution for returns to capture the extreme returns and accommodate any possible errors of covariance estimates in the first stage.

4.2.4. In-sample performance of model A and model B using different covariance matrix measures. We apply model A (covariance model) and model B (correlation model) for the covariance mean function of the MCARR model to different covariance matrix measures in (4.4) to (4.7). The results are provided in Table 4.3. We observe that, for each covariance matrix measure, model B consistently exhibits a slightly lower DIC compared to model A in stage one. On the other hand, model A does show a better performance in stage two, which suggests that model B experiences overfitting. However, our experience shows that it is more likely to have a diverged MCMC for model A, which could be a big issue in the one-step-ahead forecast. As a result, the computational time required for model A is significantly longer than that of model B. Considering these advantages of model B, together with its superior interpretability, we conclude that model B is the preferred choice for this application. For the upcoming proposed models, they are defined based on model B.

To choose the covariance matrix measure, we note that we cannot compare the four Stage one models using four covariance matrix measures based on DICs as they represent different data fitted to MCARR models (model A or B). However, since the returns are the same, we can compare the DIC of the second stage of the two return models. Note that while the PK measure is more efficient compared to the sample variance, the stage two DIC of $\boldsymbol{V}_t'^{\cdot}$ is significantly larger (poorer model fit) than the stage two DIC of $\boldsymbol{V}_t''$. Our covariance model specifies a distribution for covariance matrices instead of a marginal distribution for both volatilities and correlation matrices. Thus, the discrepancy in the efficiency of modelling volatility and correlation may result in a consistent model fit. Results show that the covariance matrix measure $\boldsymbol{V}_t$ is preferred as the DIC for the stage

two return models using $\boldsymbol{V}_t$ in stage one is the lowest. In the remaining part of this application, the input for the stage one MCARR model is always $\boldsymbol{V}_t$.

Figure 4.7 plots the in-sample fit of $\{\boldsymbol{V}_t\}$ in terms of variances, covariances and correlations using model A. The credible intervals for model A (so for model B) are computed using the Highest Density Interval (HDI) approach, which defines credible intervals to be the shortest intervals with the highest densities. This is particularly advantageous for skewed distributions, as discussed in Wasserman (1989), and it provides a more accurate and reliable estimation of the credible intervals, yielding reduced errors in comparison to conventional interval estimation techniques.

Measure	Model A		Model B	
	Stage 1	Stage 2	Stage 1	Stage 2
$\boldsymbol{V}_t$	-192986.2	-27315.0	-193039.0	-27265.0
$\boldsymbol{V}_t^{\cdot\prime}$	-193241.5	-27312.6	-193265.9	-27260.7
$\boldsymbol{V}_t^{\prime\cdot}$	-194777.4	-27366.1	-194901.7	-27219.9
$\boldsymbol{V}_t''$	-195199.5	-27306.7	-195220.7	-27252.7

TABLE 4.3. DIC of model A (covariance model) and model B (correlation model) using the four covariance matrix measures in (4.4) to (4.7) and the proposed two-stage MCARR-return models for NASDAQ, S&P 500 and DJIA indices.

4.2.5. In-sample paremeter estimation, out-of-sample volatility and return forecasts and model comparison. Parameter estimates and their corresponding 95% confidence intervals for the order (1,1) two-stage MCARR-return model are reported and compared in Table 4.4 to the results of the order (1,1) DCC-GARCH model defined in (1.66) to (1.70). Parameter estimates are mostly significant except a_{0i} and ϕ_{ij} for the DCC-GARCH model and ϕ_{12}, ϕ_{13} for MCARR-return model model. The long-term variances a_{0i} are consistently estimated to be lower for the MCARR-return model. In addition, the long-term persistence (b_i, β) for variance and correlation are also consistently estimated to be lower than the short-term persistence (a_i, α) for the MCARR-return model. A careful examination of the estimates reveals that the sums of a_i and b_i for the DCC-GARCH model are relatively closer to 1 (0.9905,0.9797,0.9656) than the proposed model (0.9764,0.9749,0.9662), which indicates higher potential for model non-stationarity. Given that the sum of a_i and b_i is approximately the same across assets, the fact that a_{0i} and ϕ_{0i} is highest for NASDAQ again highlights its high returns and volatility for investors.

For returns, the negative ϕ_{ij} estimates show antipersistence, though mostly insignificant. This result agrees with the PACF plot, which suggests that the autoregressive effects in the return series are not statistically significant. Moreover, the degrees of freedom estimates of the Student-t distribution for both models are similar. Additionally, it is noteworthy that ω_i are adjustment factor parameters only for the two-stage MCARR-return models. These parameters satisfy $\omega_i^2 \frac{\upsilon}{\upsilon-2} \approx 1$ for all $i = 1, 2, 3$, which provide strong evidence supporting the efficiency of the PK measure in the MCARR-return model. Moreover, parameters ψ_{ij} are not available in the DCC-GARCH model as they are estimated using the unconditional standardised residuals' covariance matrix in (1.71).

Figure 4.8 presents the observed values, the fitted median, as well as the 75% and 95% confidence intervals for both the in-sample fit and out-of-sample forecast of the covariances (lower triangle), variances (diagonal), and correlations (upper triangle) obtained using model B with a trivariate Wishart distribution. These covariance matrices, estimated as $\nu\boldsymbol{\Lambda}_t$, are subsequently applied to the second stage of the analysis using the trivariate-t return model specified in equations

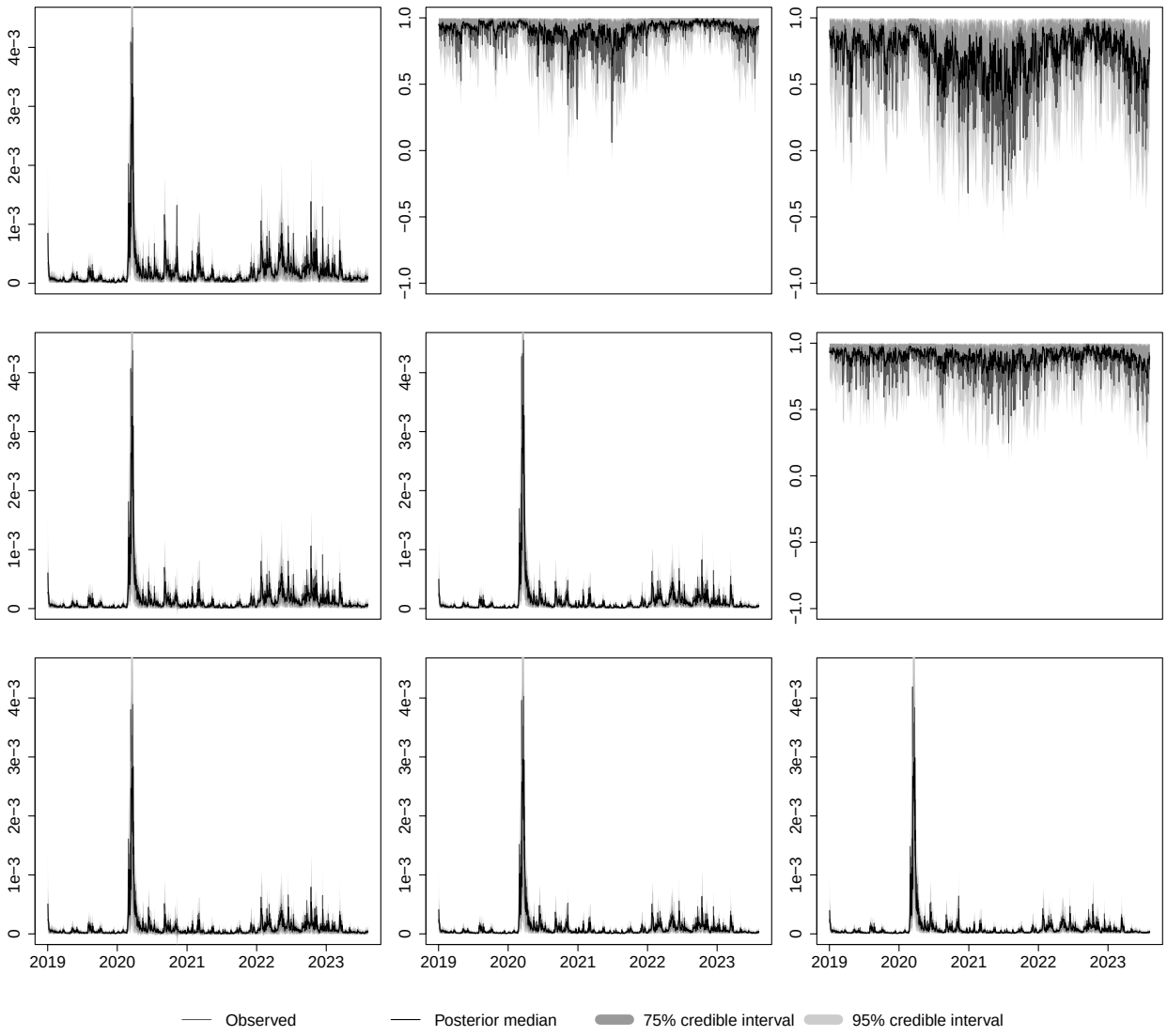


FIGURE 4.7. Observed, predicted median and credible intervals for variances (diagonal), covariances (lower triangle) and correlations (upper triangle) for the in-sample fit using the proposed stage-one MCARR model with model A (on covariance) for NASDAQ, S&P 500 and DJIA indices.

(3.22) and (3.23). Comparing model B (correlation model) with model A (covariance model) in Figure 4.7, it is clear that the median estimates for correlations are more precise and clear for model B than those of model A, confirming our argument that direct modelling of correlation is more efficient than the modelling of covariance. The corresponding in-sample fit and out-of-sample forecast of the return series using the covariance matrix estimated from model B in stage one are illustrated in Figure 4.9, showing near zero mean processes for all three indices and a higher volatility for NASDAQ.

Figure 4.10 presents the fitted variance, covariance and correlation using the DCC-GARCH model and takes the range-based covariance matrix measure $\boldsymbol{V}_t$ with the PK volatility measure and the range-based correlation measure to provide benchmark 'observed' values for comparison. This

Model B	Mean	2.5%	97.5%	DCC-GARCH	Mean	Standard error
a_1	0.4608	0.4132	0.5110	a_1	0.0851	0.0330
b_1	0.5156	0.4666	0.4996	b_1	0.9054	0.0353
${a_{01}}^{\dagger}$	0.8718	0.6264	0.7794	${a_{01}}^{\dagger}$	2.4819	3.1603
a_2	0.4817	0.4015	0.5256	a_2	0.1646	0.0372
b_2	0.4932	0.4515	0.5076	b_2	0.8151	0.0441
${a_{02}}^{\dagger}$	0.5306	0.4040	0.5747	${a_{02}}^{\dagger}$	3.5149	4.5645
a_3	0.4759	0.4269	0.5260	a_3	0.1758	0.0306
b_3	0.4903	0.4406	0.5388	b_3	0.7898	0.1230
${a_{03}}^{\dagger}$	0.4731	0.3445	0.5195	${a_{03}}^{\dagger}$	4.2928	12.4258
α	0.1918	0.1631	0.2220	α	0.0759	0.0109
β	0.7292	0.6846	0.7707	β	0.8905	0.0184
ψ_{12}	0.9525	0.9420	0.9631	-		
ψ_{13}	0.8267	0.7942	0.8696	-		
ψ_{23}	0.9211	0.9055	0.9361	-		
ν	5.8206	5.6685	5.9752	-		
${\phi_{01}}^{\ddagger}$	9.6344	4.4462	14.7650	${\phi_{01}}^{\ddagger}$	6.4744	3.2393
${\phi_{02}}^{\ddagger}$	6.3948	2.6396	10.0929	${\phi_{02}}^{\ddagger}$	5.5574	2.5685
${\phi_{03}}^{\ddagger}$	4.2411	0.5705	7.9052	${\phi_{03}}^{\ddagger}$	5.0268	2.3061
ϕ_{11}	-0.0189	-0.0558	-0.0190	${\phi_{11}}^{\diamond}$	-0.0580	0.0326
${\phi_{12}}^{\diamond}$	-0.0016	-0.0370	0.0336	${\phi_{12}}^{\diamond}$	-0.0537	0.0360
${\phi_{13}}^{\diamond}$	-0.0130	-0.0508	0.0252	${\phi_{13}}^{\diamond}$	-0.0533	0.0351
υ	8.6179	6.9477	10.8952	υ	7.1364	0.6115
ω_1	0.8784	0.8443	0.9131	-		
ω_2	0.8572	0.8265	0.8888	-		
ω_3	0.9017	0.8664	0.9381	-		

$\dagger$: $\times 10^{-6}$; $\ddagger$: $\times 10^{-4}$; $\diamond$: not significant at 0.05.

TABLE 4.4. Parameter estimates, confidence intervals and standard errors for the proposed two-stage MCARR-return model (model B on correlation) and the DCC-MGARCH model for NASDAQ, S&P 500 and DJIA indices.

visualisation also highlights the capacity of these measures to provide reliable and robust estimates of volatility and correlations under various conditions. We observe that the DCC-GARCH model tends to overestimate in-sample volatility. These findings collectively highlight the advantages of model B over the DCC-GARCH model in terms of parameter estimation, computational efficiency, and predictive accuracy. We remark that there is no confidence interval for the volatility and correlation estimates as they are modelled as deterministic in the MGARCH models.

To facilitate a more comprehensive comparison, Figure 4.11 plots the posterior median estimates of model B alongside the corresponding estimates derived from the DCC-GARCH model. A closer inspection of these results reveals that the DCC-GARCH model again consistently produces slightly higher estimates of volatility compared to model B, particularly in periods where volatility remains relatively stable. This observation may indicate a tendency of DCC-GARCH

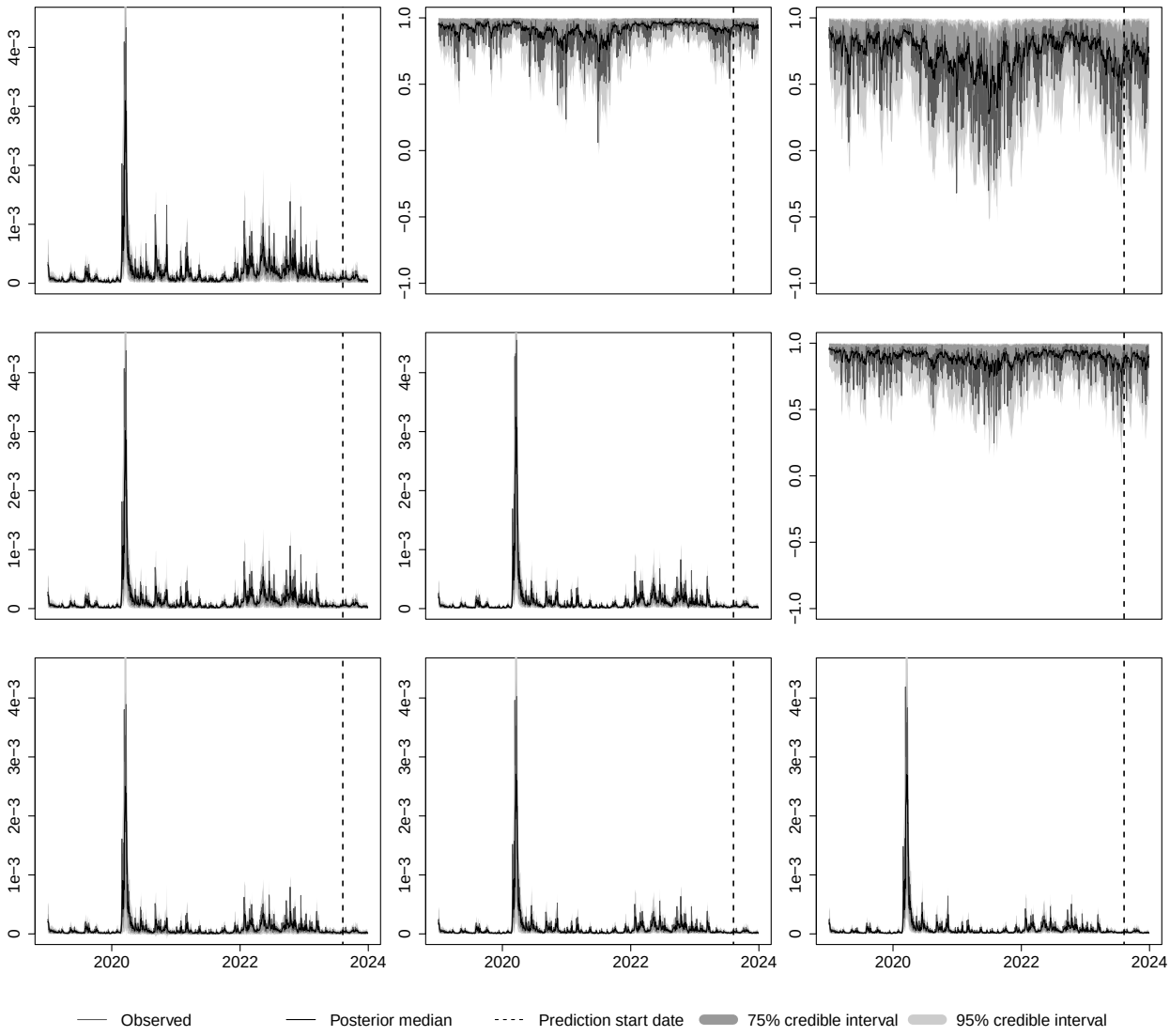


FIGURE 4.8. Observed, predicted median and credible intervals for variances (diagonal), covariances (lower triangle) and correlations (upper triangle) for the in-sample fit and out-of-sample forecast using the proposed stage-one MCARR model (model B on correlation) for NASDAQ, S&P 500 and DJIA indices.

to overestimate volatility in calmer market conditions, potentially due to its reliance on conditional covariance dynamics that assume persistence. In contrast, model B appears to provide more conservative and adaptive volatility estimates in such settings.

Furthermore, Figure 4.12 plots the differences between the posterior median estimates from model B minus DCC-GARCH, showing that the DCC-GARCH model becomes particularly pronounced during periods of heightened market turbulence as there are more negative than positive differences for variances and covariances, indicating possible overestimation. Notably, model B captures the volatility jump observed in early 2020 with greater speed and precision than the DCC-GARCH model, as there are positive differences in the first few days of the jump. This result aligns with the structural differences in the models, particularly the fact that the parameter a_i in model

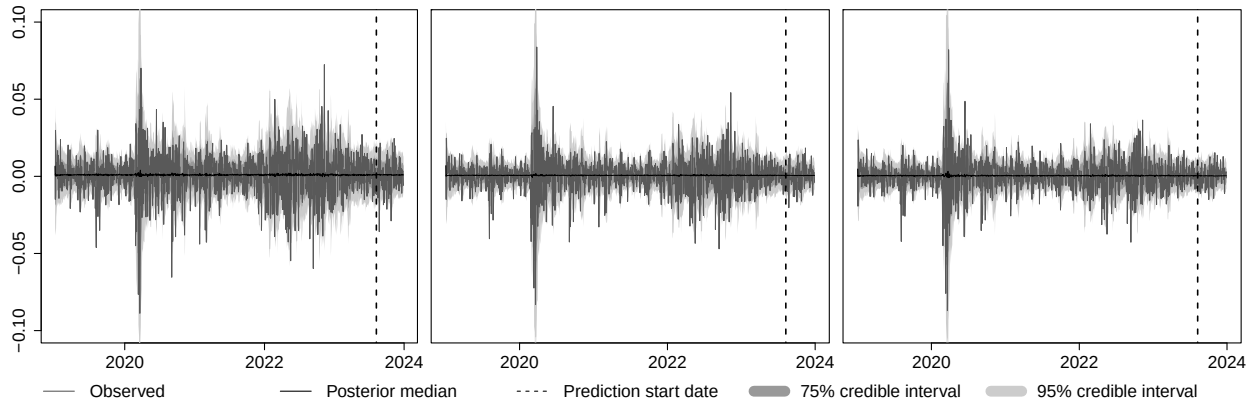


FIGURE 4.9. Posterior median and posterior predictive intervals of returns using the proposed stage two return model with the covariance matrix estimated using model B (correlation model) for NASDAQ, S&P 500 and DJIA indices.

B is larger than its counterpart in the DCC-GARCH model, as we have mentioned. The higher a_i value in model B reflects its increased sensitivity to sudden changes in volatility, allowing it to react more promptly to market shocks. These findings underscore the advantages of model B in effectively capturing abrupt changes in volatility dynamics while maintaining robust performance in stable conditions. The improved responsiveness of model B, the two-stage MCARR-return model with correlation modelling, compared to the DCC-GARCH model highlights its suitability for applications where the accurate and timely identification of volatility jumps is crucial. For the correlation, the (1,2) pair shows overestimation, whereas the (2,3) pair shows estimation for model B relative to the DCC-GARCH model.

We further compare the performance of our correlation model against a traditional moving window sample correlation estimation. A ϖ-window moving average sample correlation is defined as

$$\text{cor}_{ij,t} = \frac{\sum_{\delta=t-(\varpi-1)/2}^{t+(\varpi-1)/2}(r_{i,\delta}-\bar{r}_{i,t})(r_{j,\delta}-\bar{r}_{j,t})}{\sqrt{\sum_{\delta=t-(\varpi-1)/2}^{t+(\varpi-1)/2}(r_{i,\delta}-\bar{r}_{i,t})^2\sum_{\delta=t-(\varpi-1)/2}^{t+(\varpi-1)/2}(r_{j,\delta}-\bar{r}_{j,t})^2}}, \tag{4.13}$$

where $\bar{r}_{i,t} = \frac{1}{\varpi}\sum_{\delta=t-(\varpi-1)/2}^{t+(\varpi-1)/2} r_{i,\delta}$. Figure 4.14 reveals that the correlations estimated by model B align closely with the 21-day moving window sample correlations, thereby providing additional validation of the model's efficiency in capturing the underlying correlation dynamics and providing accurate forecasts. While the moving window estimates are relatively straightforward to implement, their effectiveness is highly sensitive to the choice of window size ϖ. A larger window size typically results in smoother correlation estimates as it averages over a broader range of data points, thereby reducing short-term fluctuations. However, this advantage comes at the cost of introducing greater intrinsic persistence into the correlation estimates due to the extended averaging process, which may not accurately reflect rapid changes in the underlying correlation structure.

Moreover, the use of a larger window size implicitly assumes constant correlation over the entire period covered by the window. This assumption is often impractical, especially in financial markets where volatility and correlations can exhibit significant time variation due to market shocks, structural changes, or economic events. Consequently, a larger window size may fail to capture the temporal heterogeneity of the data, leading to potential biases in the estimated correlations. In contrast, our model overcomes these limitations by directly accounting for time-varying

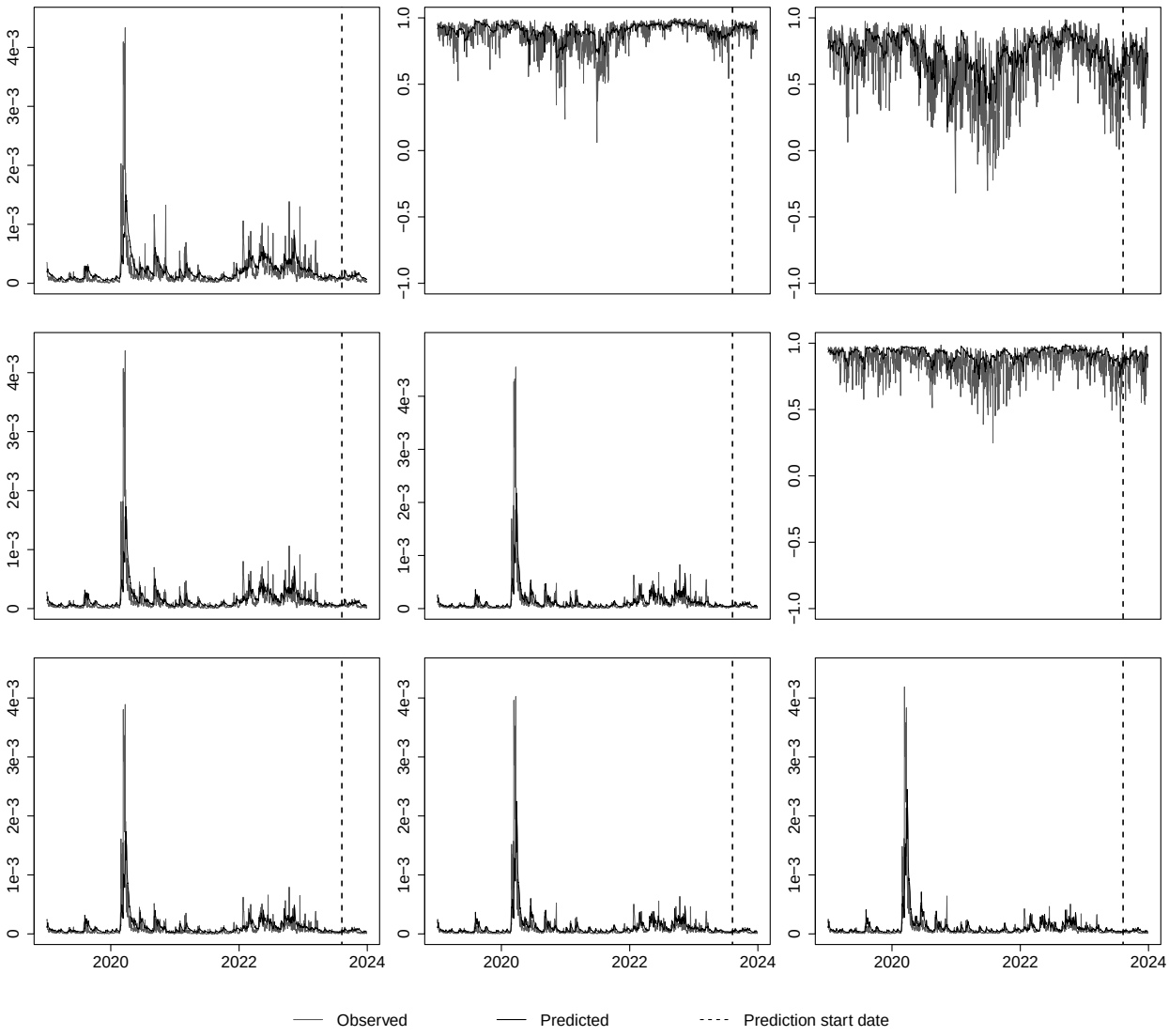


FIGURE 4.10. Observed and predicted using DCC-GARCH model for NASDAQ, S&P 500 and DJIA indices. The diagonals are plots to compare observed (PK) and predicted volatilities. The plots in the lower triangle compare the observed (using PK volatilites and range-based correlation) and predicted covariances, whereas the plots in the upper triangle compare the observed (range-based) and predicted correlations.

volatility and correlation dynamics without relying on a fixed window size. This feature enables the model to adapt to changes in market conditions and provide more precise and responsive estimates. The strong consistency between our model's estimates and the moving window sample correlations further underscores its robustness while highlighting the practical trade-offs involved in choosing window sizes for moving window approaches.

4.2.6. Value-at-Risk forecast. To evaluate the performance of the proposed models, we assess the 1% Value-at-Risk (VaR) during the forecast period. We estimate the one-day-ahead VaR

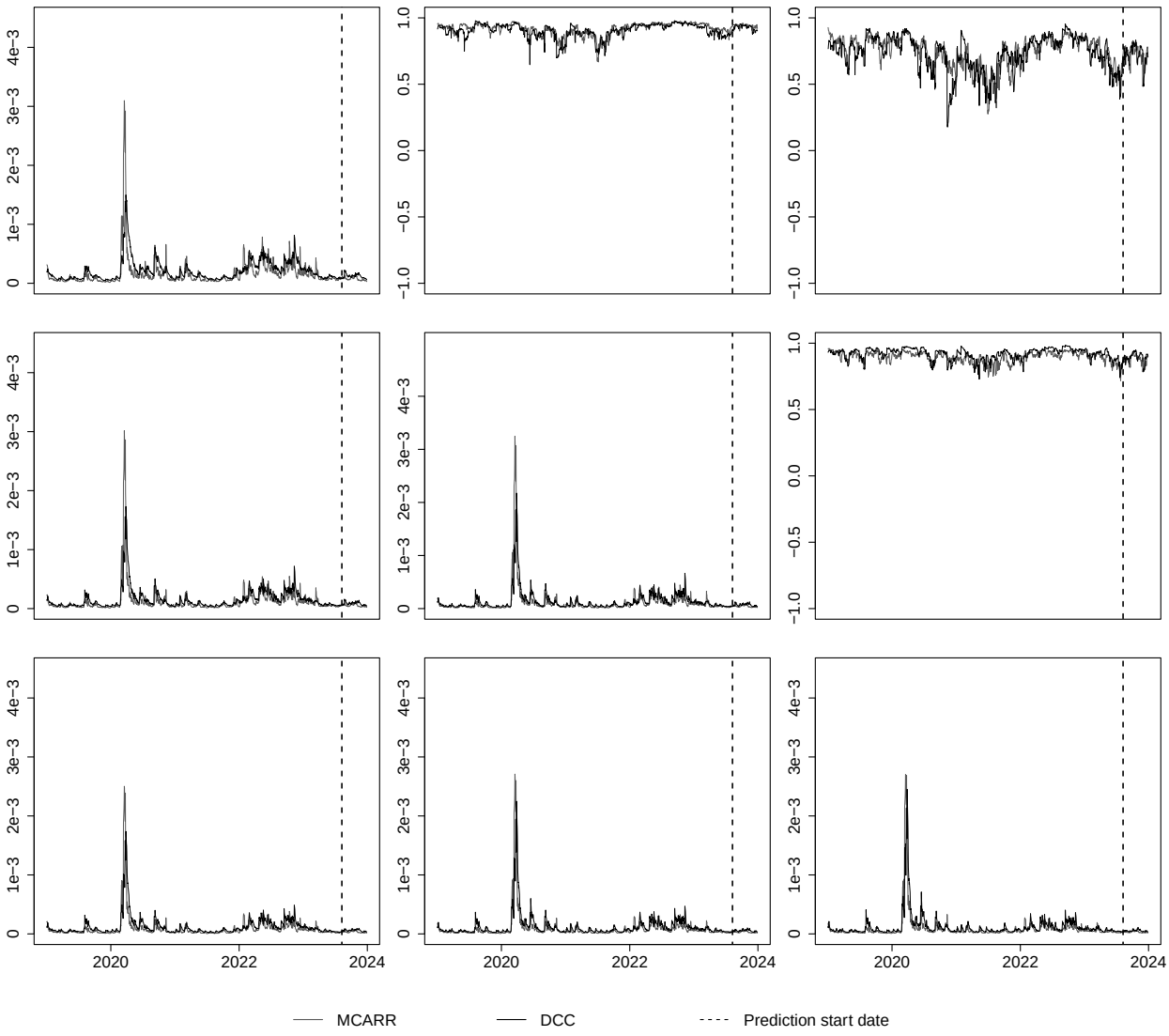


FIGURE 4.11. Comparison between estimates using MCARR (model B) and DCC-GARCH models for NASDAQ, S&P 500 and DJIA indices. The diagonals are plots of volatility. The plots in the upper triangle are for correlation, while the plots in the lower triangle are for covariance.

forecast of return vector $\boldsymbol{r}_{N+1}|\boldsymbol{V}_{N+1}, \mathcal{F}_N$ based on the posterior predictive distributions defined in (3.37). We perform posterior sampling for each asset i using marginal distribution by drawing

$$\hat{r}_{i,N+1}^{(l)} \quad \text{from} \quad f_r\big(r_{i,N+1}|\mu_i(\boldsymbol{\theta}_2^{(l)}), \lambda_{i,N+1}(\boldsymbol{\theta}_1^{(l)}), \boldsymbol{\gamma}_i(\boldsymbol{\theta}_2^{(l)}), \upsilon(\boldsymbol{\theta}_2^{(l)})\big)$$

for the posterior sample l with parameters $\boldsymbol{\theta}^{(l)} = (\boldsymbol{\theta}_1^{(l)}, \boldsymbol{\theta}_2^{(l)}),\ l = 1, \ldots, L$ where $\lambda_{i,N+1}(\cdot)$ is the i-th diagonal entry of the mean covariante matrix $\boldsymbol{\Lambda}_{N+1}(\boldsymbol{\theta}_1^{(l)})$. Then the *point estimate* of the VaR forecasts are the quantiles of the posterior predictive distributions $\hat{r}_{i,N+1}^{(l)}, l = 1, \ldots, L$.

Note that we can also obtain the posterior predictive distribution of VaR, which can be used to obtain the uncertainty of VaR forecast in the form of a credible interval. For instance, if we wish to

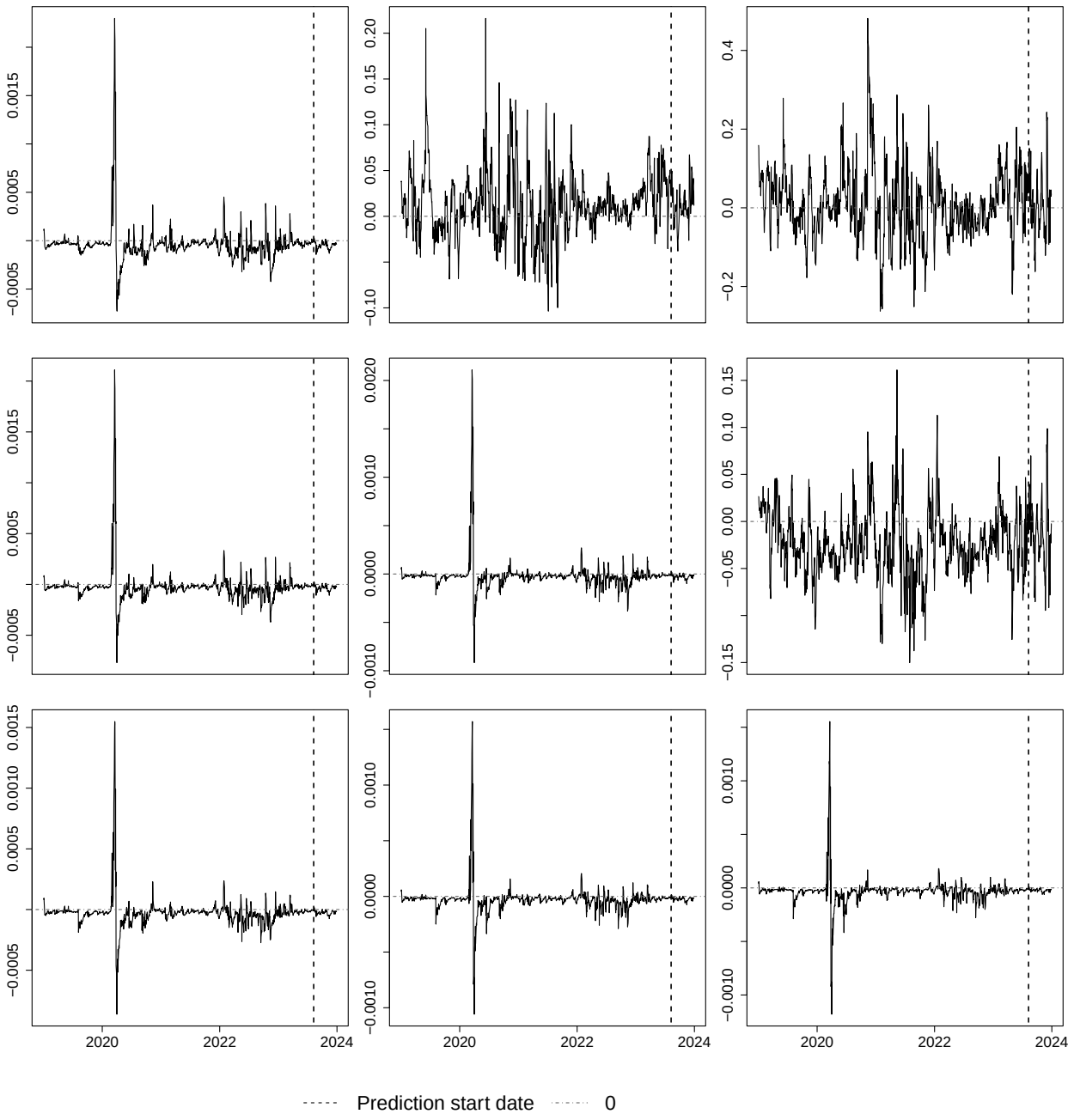


FIGURE 4.12. The difference estimates for those from the MCARR model (model B on correlation) minus those from the DCC-GARCH model. The diagonals are plots for volatilities. Plots in the upper triangle are for correlations, while plots in the lower triangle are for covariances.

evaluate the prediction interval for the 1% VaR of returns, $r_{i,t+1,0.01}$, of which close form quantile function is available, we have

$$r^{(l)}_{i,N+1,0.01} = f_r^{-1}\big(0.01|\mu_i(\boldsymbol{\theta}_2^{(l)}), \lambda_{i,N+1}(\boldsymbol{\theta}_1^{(l)}), \gamma_i(\boldsymbol{\theta}_2^{(l)}), \upsilon(\boldsymbol{\theta}_2^{(l)})\big)$$

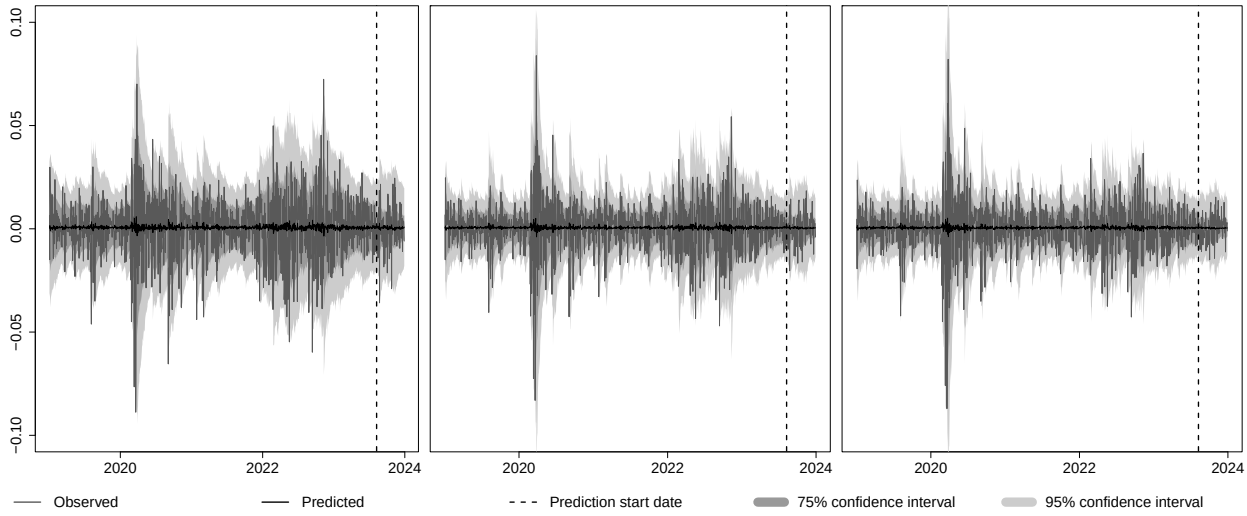


FIGURE 4.13. Observed, predicted median and confidence intervals of returns for NASDAQ, S&P 500 and DJIA indices respectively using DCC-GARCH model.

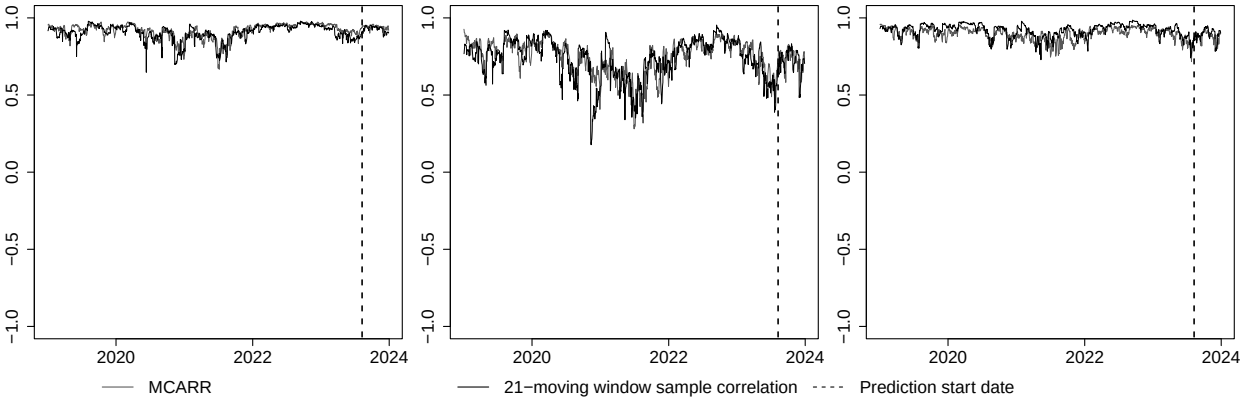


FIGURE 4.14. Correlation estimates using MCARR model (model B on correlation) and 21-moving window sample correlation for NASDAQ, S&P 500 and DJIA indices.

Then the 95% credible interval for the 1% VaR, $r_{i,N+1,0.01}$, can be obtained from the middle 95% values of the posterior sample of $r^{(l)}_{i,N+1,0.01},\ l = 1, \ldots, L$.

If the distribution is complicated with no close form quantile function $f_r^{-1}(\cdot|\cdot)$, we can approximate the 1% quantile, $r_{i,N+1,0.01}$, again based on posterior sampling. Specifically, we can sample L' returns, from the l-th posterior sample with parameters $\boldsymbol{\theta}^{(l)}$ as below:

$$r^{(l,l')}_{i,N+1} \sim f_r\big(r_{i,N+1}|\mu_i(\boldsymbol{\theta}_2^{(l)}), \lambda_{i,N+1}(\boldsymbol{\theta}_1^{(l)}), \gamma_i(\boldsymbol{\theta}_2^{(l)}), \upsilon(\boldsymbol{\theta}_2^{(l)})\big),\ l' = 1, \ldots, L'.$$

For each posterior sample l, $r^{(l)}_{i,N+1,0.01}$ is the 0.01 quantile of the distribution $r^{(l,l')}_{i,N+1},\ l' = 1, \ldots, L'$.

While the VaR of individual returns r_1, r_2, and r_3 provide a measure of volatility for each asset, it does not account for the joint dependence structure or covariance between the assets. To address this, we evaluate the model's performance on two additional types of synthetic assets, which allow for a more thorough assessment of the model's ability to capture both volatility and covariance

dynamics. The first type of asset considered is an equal-weight portfolio constructed pairwise from the three assets. To be more specific, the return of the equal-weight portfolio comprising NASDAQ and S&P 500 is given by $(r_1 + r_2)/2$. Similarly, we construct equal-weight portfolios for the pairs of assets NASDAQ and DJI, and assets S&P 500 and DJI, resulting in three pairwise portfolios. This approach enables us to evaluate the model's ability to estimate portfolio-level volatility, where the returns are influenced by the joint dynamics of the underlying assets. The second type of asset we analyse involves the exchange-rate-like relationships between the three indices. Specifically, we compute the pairwise differences between the returns of the three assets, which are given as $r_1 - r_2$, $r_1 - r_3$ and $r_2 - r_3$. These return differences allow us to assess how well the models capture both marginal volatility and covariance structure. If the model accurately estimates not only the individual volatilities of the three assets but also their covariance, we expect the VaR performance for these synthetic assets to be robust. This is because the volatility of the portfolio and the constructed exchange-rate-like assets can be directly expressed as a function of the volatilities of r_1, r_2, and r_3, along with the pairwise covariances between these assets.

Figures 4.15 and 4.16 illustrate the VaR forecast performance for the NASDAQ, S&P 500, DJIA indices, and the previously defined synthetic assets using model B (correlation model) and DCC-GARCH model, respectively. A careful comparison of the plots reveals that the DCC-GARCH model consistently predicts a lower VaR across all assets. This finding aligns with our earlier observations, where the DCC-GARCH model tends to overestimate volatility, thereby leading to a more conservative (lower) VaR prediction.

In terms of VR performance, we observe that the 95% credible intervals provided by model B closely match the nominated 95% confidence level, suggesting that model B performs well in capturing the overall variability of the return distributions. Table 4.5 summarises the proportion of returns that fall below the VaR for each asset called violation rate (VR) for each return r_i, equal weight portfolio returns $r_i/2 + r_j/2$ and exchange rate returns $r_i - r_j$. These VR can be seen from Figures 4.15 and 4.16 in which returns lower than VaR are indicated with '$\times$'. To formally assess the adequacy of the coverage rates, we conduct the Kupiec test, which evaluates whether the empirical VR is consistent with the nominated coverage probability. Let N be the sample size, p be the observed proportion or empirical VR, and π be the true proportion or the nominated coverage probability. The hypotheses and the Kupiec test statistic are given by

$$\begin{aligned} &\mathrm{H}_0 : \pi = \alpha, \quad \text{vs} \quad \mathrm{H}_1 : \pi \neq \alpha, \\ &\mathcal{K} = 2N \log \frac{p^p (1-p)^{1-p}}{\alpha^p (1-\alpha)^{1-p}} \sim \chi^2_1. \end{aligned} \tag{4.14}$$

The test statistic is exactly twice the likelihood ratio. Hence, asymptotically, the test statistic follows a χ^2 distribution with a degree of freedom 1. We take $\alpha = 0.01$ when applying the test.

The test results indicate that the null hypothesis—stating that the VR equals the expected level—cannot be rejected for any of the assets. This outcome suggests that both models provide statistically acceptable VaR predictions, though it should be noted that the relatively small sample size of 100 observations during the forecast period may limit the power of the test to detect small deviations from the null hypothesis, particularly for a tail probability of $\pi = 0.01$.

To further evaluate the models' VaR predictive performance, we measure the quantile loss for each asset. The quantile loss (QL) function at quantile level α is defined as

$$\mathrm{QL}(x, \hat{x}) = \alpha(x - \hat{x})\mathbf{1}_{x>\hat{x}} + (1-\alpha)(\hat{x} - x)\mathbf{1}_{x<\hat{x}}, \tag{4.15}$$

where $\hat{x}$ is the estimate α-quantile of returns. The QL function assesses the accuracy of the VaR forecasts by penalising deviations between observed returns and predicted quantiles. Our results

indicate that for NASDAQ (r_1) and the exchange-rate-like synthetic assets ($r_i - r_j$), the QL under model B is marginally higher than that of the DCC-GARCH model. This suggests that, for these specific cases, the DCC-GARCH model provides slightly better quantile predictions. However, the observed differences in QL are relatively small and may not be statistically significant. For the remaining assets, model B, the two-stage MCARR-return model with correlation modelling, performs comparably or marginally better than DCC-GARCH, further underscoring its robustness.

TABLE 4.5. Summary of violation rate, p-value of the Kupiec test and quantile loss using the two-stage MCARR-Return (model B on correlation) and DCC-GARCH models for NASDAQ, S&P 500 and DJIA indices during the forecast period.

	2-stage MCARR-Return (model B)			DCC-GARCH		
Order	VR	p-value	QL	VR	p-value	QL
r_1	0.02	0.3763	0.0356$^\diamond$	0	0.1563	0.0325
r_2	0.03	0.1047	0.0224	0	0.1563	0.0240
r_3	0.01	1	0.0155	0	0.1563	0.0210
$r_1/2 + r_2/2$	0.01	1	0.0278	0	0.1563	0.0278
$r_1/2 + r_3/2$	0.01	1	0.0235	0	0.1563	0.0249
$r_2/2 + r_3/2$	0.02	0.3763	0.0187	0	0.1563	0.0219
$r_1 - r_2$	0.03	0.1047	0.0178$^\diamond$	0.01	1	0.0141
$r_1 - r_3$	0.03	0.1047	0.0324$^\diamond$	0	0.1563	0.0228
$r_2 - r_3$	0.04	0.0228	0.0120$^\diamond$	0	0.1563	0.0108

$\diamond$: QL of MCARR model is higher than DCC-GARCH model.

In summary, while DCC-GARCH exhibits a tendency to predict lower VaR due to its overestimation of volatility, both models demonstrate reasonable VR and QL performance. The differences in performance between the models are relatively minor and context-dependent, particularly for the exchange-rate-like assets. These findings suggest that model B, the two-stage MCARR-return model on correlation, offers a competitive alternative to the DCC-GARCH model, with additional benefits such as improved responsiveness to volatility jumps, as noted in earlier analyses.

4.3. Study of COVID-19 impact using different GARCH variate

In the previous section, we investigated the performance of the proposed trivariate MCARR-return model, where all pairwise correlations among the assets are observed to be high. In this chapter, we extend the application to a higher-dimensional dataset with more varied levels of correlation. Specifically, we consider the candlestick data of four assets: Bitcoin, Ethereum, crude oil, and gold. This extension allows us to evaluate the model's ability to fit a more complex and diversified dataset while capturing distinct features and relationships among the assets.

We begin with the expectation of a strong correlation between Bitcoin and Ethereum, two of the most prominent cryptocurrencies. Both assets are characterised as high-risk, high-volatility financial instruments, largely driven by speculative trading, technological developments, and market sentiment. Their shared status as digital assets, coupled with similar demand-supply dynamics, contributes to their typically strong positive correlation. In contrast, we anticipate much weaker correlations between the cryptocurrency pairs (Bitcoin, Ethereum) and the other two assets, crude oil and gold, which are traditionally distinct in both market behaviour and investor perception.

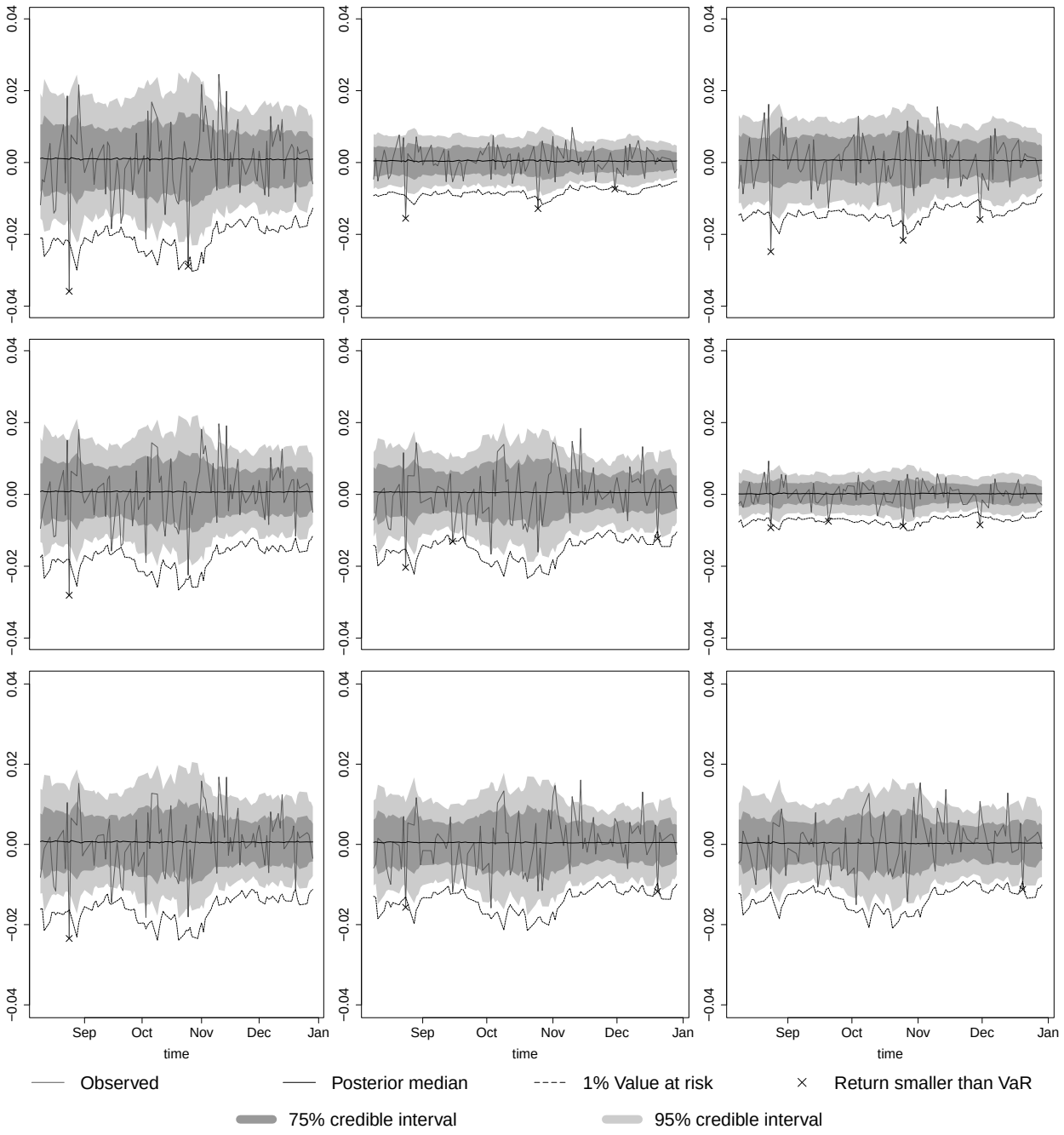


FIGURE 4.15. Observed, predicted, confidence intervals and value at risk estimates of returns for the forecast period in 2023 using model B, the two-stage MCARR-return model with correlation modelling. The diagonals are plots for the returns of the NASDAQ, S&P 500 and DJIA indices respectively. The plots in the upper triangle are returns of exchange rate, while the plots in the lower triangle are returns of equal-weight portfolios.

On the other hand, crude oil behaves as a key indicator of global economic activity, given its role as a critical input for industrial production and transportation. Oil prices are influenced

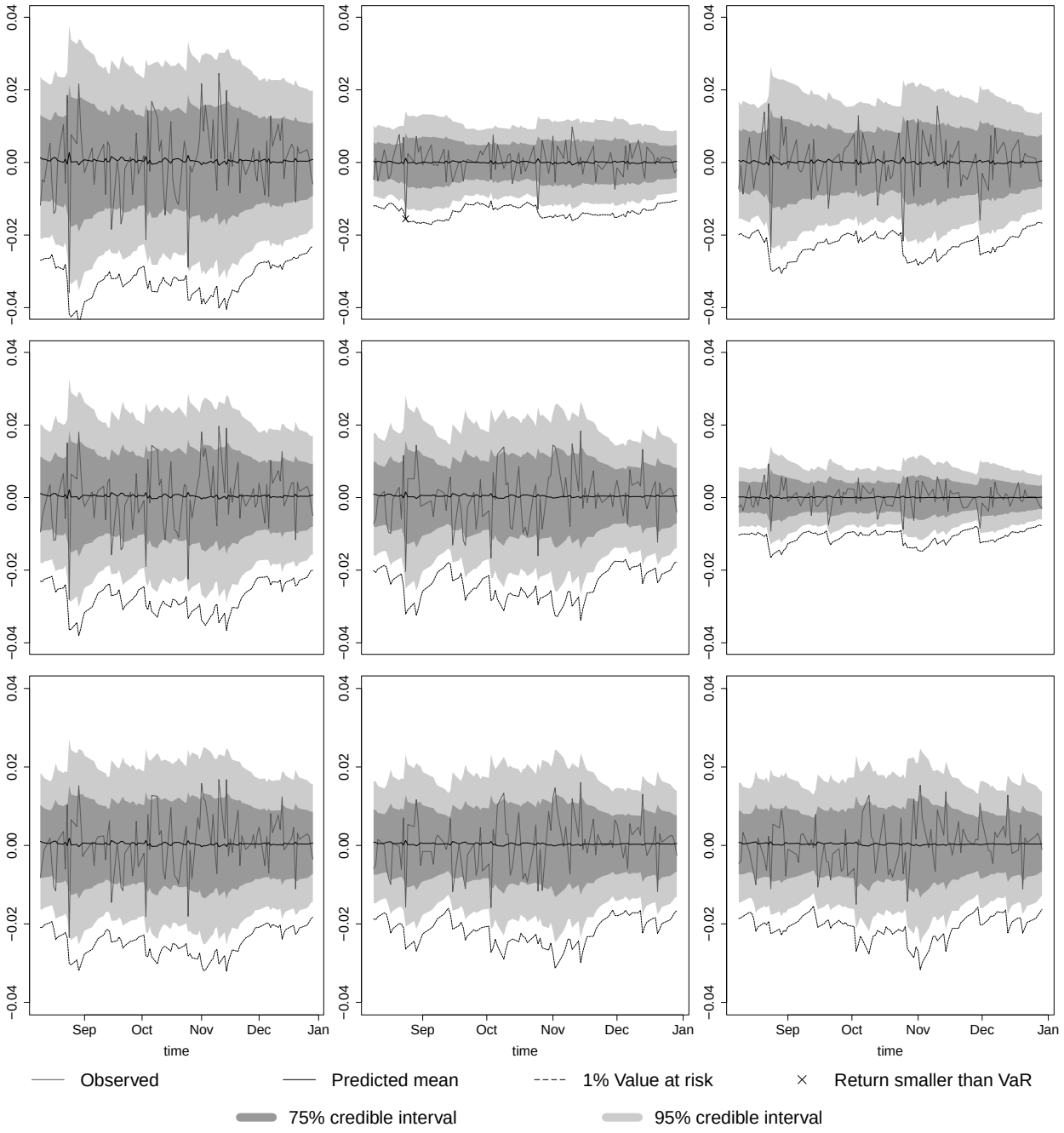


FIGURE 4.16. The confidence intervals and value at risk estimation of returns for prediction period using DCC-GARCH. The diagonals are returns of NASDAQ, S&P 500 and DJIA indices respectively. The upper diagonals are returns of exchange rate, while the lower diagonals are returns of portfolios.

by macroeconomic factors such as supply disruptions, geopolitical events, and changes in global demand. While crude oil prices may not have a direct relationship with cryptocurrency prices, the broader economic conditions that drive oil price changes—such as inflation, liquidity, and economic growth—could indirectly affect investor sentiment and capital flows into digital assets.

Gold, often referred to as a safe-haven asset, is widely regarded as a store of value during periods of economic uncertainty or market turmoil. Its price movement tends to exhibit low volatility relative to cryptocurrencies and is often inversely correlated with riskier assets. Investors seeking stability or hedging against inflation tend to shift capital into gold, which dampens its short-term price fluctuations.

The inclusion of crude oil and gold alongside Bitcoin and Ethereum in our analysis provides an opportunity to assess how our model performs in capturing relationships among assets with highly divergent characteristics. Specifically, Bitcoin and Ethereum exhibit high volatility and are strongly correlated, while gold displays relatively lower volatility and weaker association with cryptocurrencies. Crude oil introduces an additional dimension of economic relevance, as its price dynamics are closely tied to broader macroeconomic indicators.

In this section, we aim to demonstrate that our model can effectively accommodate this higher-dimensional setting while maintaining its ability to capture the diversity in asset features, such as differing levels of volatility and correlation structures. We first want to examine whether the dynamics will change significantly during the COVID-19 period. To achieve this, we adopt a segmented approach to compare the performance of model fitting across three different time periods. By analysing these segments, we can observe how the relationships between assets evolve over time, particularly during periods of calm and economic stress or market shocks during different stages of the COVID-19 pandemic. Additionally, we extend the volatility modelling process by incorporating additional explanatory factors. These factors may enhance the accuracy and flexibility of the model. This extension allows us to better capture complex dynamics in the volatility of each asset and further demonstrates the adaptability of our proposed methodology.

4.3.1. Data. The Bitcoin (BTC, labelled as 1) and Ethereum (ETH, labelled as 2) data are collected from Binance, one of the largest cryptocurrency exchange platforms globally. For crude oil and gold, we choose to use the prices of exchange-traded funds (ETF), which is a collection of securities, such as stocks and bonds, that investors can buy and sell on stock exchanges like regular shares. ETFs of crude oil and gold track these assets, as they generally exhibit higher liquidity compared to direct trading of physical commodities. Specifically, we selected the United States Oil Fund (USO, labelled as 3), which tracks the price of crude oil and the SPDR Gold Shares (GLD; labelled as 4)[1], which closely tracks the price of gold. The ETF data are sourced from the DataBento platform to ensure consistency and reliability in the dataset.

It is important to note that while Bitcoin and Ethereum are traded 24 hours a day, seven days a week on Binance, ETFs such as USO and GLD are only traded during standard market hours. This discrepancy in trading hours necessitated the alignment of the data across all four assets. To address this, we take the intersection of the available time series, ensuring that all observations correspond to periods when trading data are available for both cryptocurrencies and ETFs.

Additionally, on April 20th, 2020, an unprecedented event occurred when the price of crude oil turned negative due to a collapse in oil demand and limited storage capacity during the COVID-19 pandemic. Such an extreme and highly unusual market event made the calculation of log returns infeasible for that specific day. To maintain the integrity of the analysis, we chose to exclude this data point from the dataset.

The period of interest spans from 1st May 2018 to 1st Oct 2023. After preprocessing, the final dataset comprises a total of 1,532 daily observations. The sample correlations of daily returns

[1]It is called "spiders" and is a family of ETFs traded in the US, Europe, Mexico and Asia-Pacific and managed by State Street Global Advisors (SSGA). SPDR is an acronym for the first member of the family, the Standard & Poor's Depositary Receipts, now the SPDR S&P 500 Trust ETF.

suggest a strong positive correlation between BTC and ETH, reflecting the interconnected nature of these two major cryptocurrencies. Their prices often move in tandem due to shared market drivers, such as investor sentiment, technological advancements, and regulatory developments. In contrast, the sample correlations between the remaining asset pairs (BTC-GLD, BTC-USO, ETH-GLD, ETH-USO, GLD-USO) are relatively weak, whereas the sample correlations between BTC-ETH is 0.8582, as shown in Table 4.6. This indicates that traditional assets like gold and crude oil follow distinct price dynamics, primarily influenced by macroeconomic factors, supply-demand imbalances, and global economic conditions, which are less directly tied to cryptocurrency markets.

Based on these observations, we employ a variance-weighted average range correlation estimator ($\mathfrak{m} = 2.4; p = 2$ in (2.50)) for BTC and ETH to account for their strong correlation while properly weighting the variability of returns. For the remaining asset pairs, where correlations are weaker, we choose the simple average correlation estimator ($\mathfrak{m} = 2.2; p = 0$), which provides a robust and straightforward measure of pairwise dependence when significant co-movement is absent. Among all the constructed correlation matrices, each is found to be positive definite, indicating that no NPD correction methods are required. Then $\boldsymbol{V}_t = \widehat{\boldsymbol{\Sigma}}_t$ for fitting the MCARR model is again given by (2.91), (2.53) and (2.57).

Figure 4.17 plots the variance $v_{ii,t}$ and covariance $v_{ij,t}$ in $\boldsymbol{V}_t$ as well as the correlation $\varrho_{ij,t}$ in $\widehat{\boldsymbol{\mathcal{Q}}}_t$. Results highlight the high correlation pair BTC-ETH while all other pairs have low correlation, in agreement with Table 4.6. Moreover, both table and figure show that GLD has much lower volatility than other return series. Then Figure 4.18 plots the four return series, which also reveal the highest volatility of ETH and lowest volatility of GLD. The two extremely low returns for BTC and ETH on March 12, 2020, are due to the announcement of the COVID-19 pandemic by WHO one day earlier. This event also leads to jumps of volatility in Figure 4.17.

	BTC	ETH	USO	GLD
BTC	$7.9923‡$	0.8582	0.0622	0.1851
ETH	$8.1035‡$	$11.1565‡$	0.0739	0.1586
USO	$0.3655‡$	$0.5132‡$	$4.3270‡$	0.0643
GLD	$0.4015‡$	$0.4066‡$	$0.1027‡$	$0.5888‡$

$\dagger: \times 10^{-6}$; $\quad ‡: \times 10^{-4}$.

TABLE 4.6. Sample variance, covariance and correlation of return series of BTC, ETH, USO and GLD. The diagonals are sample variances. The plots in the lower triangle are for covariances, and the plots in the upper triangle are for correlations.

4.3.2. Exploratory data analysis. Table 4.7 presents the summary statistics for $\boldsymbol{R}_t$ (returns) and $\boldsymbol{V}_t$ (covariance matrices). Consistent with findings from previous studies, we observe that the mean of ρ_{ij} aligns well with the sample correlation, reaffirming the accuracy of the range-based correlation measure in capturing the overall correlation structure. While the sample variances of r_1 and r_2 are consistent with the means of v_{11} and v_{22}, respectively, discrepancies are noted in the corresponding statistics for GLD and USO. These discrepancies suggest the presence of a scaling bias in these assets, which warrants further investigation. A detailed discussion on how our model is capable of addressing such scaling biases will be provided in Section 4.3.4, emphasising its robustness in handling these challenges.

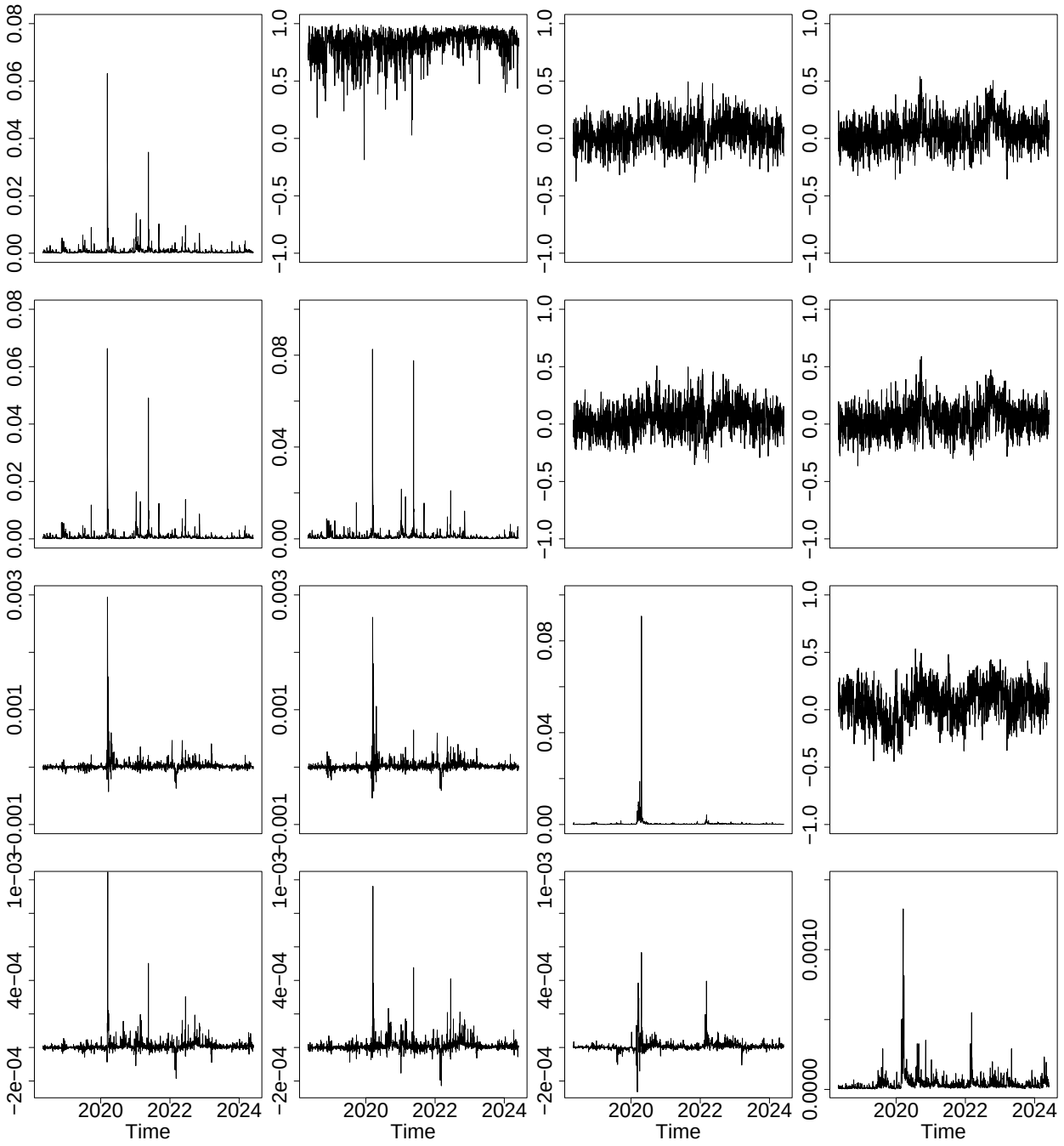


FIGURE 4.17. Plots of volatility, covariance and correlation of BTC, ETH, USO and GLD respectively. The diagonals are plots of observed volatilities. The plots on the upper triangle are for observed correlations, while the plots on the lower triangle are for observed covariances.

Additionally, we observe that the scale of the minimum values of r_1 and r_2 is notably small. This phenomenon can be attributed to the higher likelihood of pin bar formations in cryptocurrency markets. A pin bar is a candlestick pattern widely used in technical analysis to identify potential price reversals, characterised by a small body and a long wick, signalling a strong rejection of a

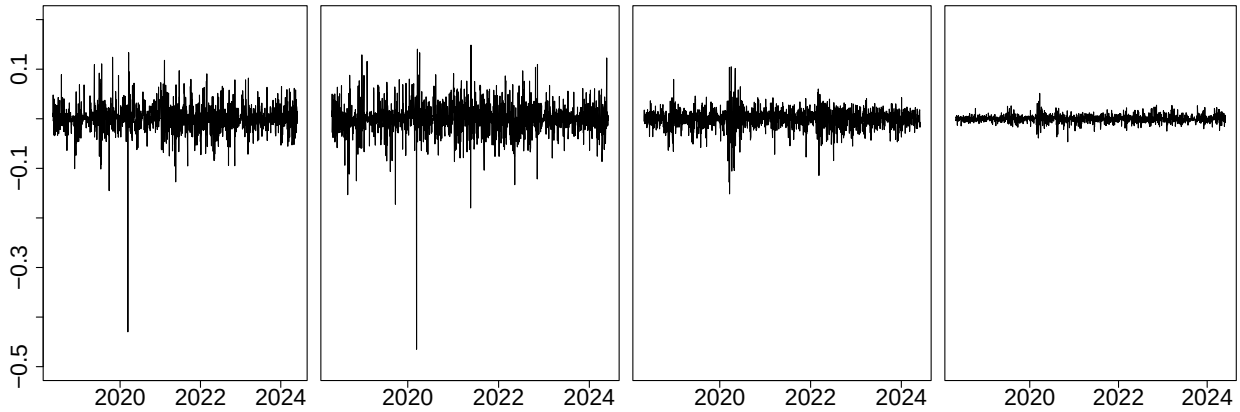


FIGURE 4.18. Plot of return time series for BTC, ETH, USO and GLD.

price level. The cryptocurrency market, with its unique characteristics, is particularly prone to the formation of pin bars. Not only does the high volatility, a defining feature of cryptocurrencies, lead to sharp price movements and rejections and create ideal conditions for pin bar formations, but the market's low depth of cryptocurrency also means that large orders can disproportionately impact prices. Its relatively unregulated nature makes it susceptible to manipulation by large players, often resulting in sharp reversals. Although pin bars have minimal direct impact on spot trading, brokers often exploit the formation of pin bars to trigger forced liquidations in futures contracts. These sharp price rejections can activate margin calls or stop-loss orders, particularly in highly leveraged positions, enabling brokers to capitalise on the volatility inherent in such patterns. These factors collectively contribute to the smaller observed minimum returns in r_1 and r_2.

On the other hand, it is also noteworthy that crude oil prices once turned negative during a significant market anomaly, yet the minimum return of USO is approximately -0.15. This disparity arises because we exclude all negative price points from the dataset, a necessary preprocessing step to ensure the reliability of our analysis. Furthermore, the price of the ETF USO does not perfectly mirror crude oil prices due to the inherent liquidity differences between ETFs and their underlying assets. ETFs, being financial instruments traded on exchanges, exhibit their own supply-demand dynamics, which may lead to deviations from the actual commodity prices they track. This divergence underscores the complexity of modelling such assets and the importance of accounting for market-specific characteristics in our analysis.

Similar to the US stock market returns, the returns for BTC, ETH, USO, and GLD exhibit minimal evidence of autoregressive or moving average effects, as suggested by the ACF and PACF plots presented in Figure 4.19(a) and (b), respectively. Both plots reveal that the autocorrelations for returns at various lags are small and largely insignificant. This is a common characteristic of financial return series, particularly in high-frequency and highly volatile markets.

To further validate this finding, we perform the Ljung-Box test, ADF test and KPSS test. The summary of the test in Table 4.8. The results of the test are not significant for all return series, confirming the absence of significant serial correlation. Despite these findings, we choose to model the returns using an AR(1) process. We conclude that there is no unit root present in the return, volatility, and covariance time series other than v_{24} and v_{34} series. This decision is motivated by the fact that while the return series do not display strong autoregressive properties, the inclusion of an AR(1) term provides a simple yet flexible structure to account for any minor linear dependencies that may not be visually discernible in the ACF or PACF plots. An AR(1) process also ensures

	Mean	Variance	Min	Max
r_1	$7.0497\ddagger$	$7.9923\ddagger$	-0.4298	0.1334
r_2	$3.8972\ddagger$	$11.1565\ddagger$	-0.4649	0.1486
r_3	$-6.5168\ddagger$	$4.3270\ddagger$	-0.1515	0.1050
r_4	$0.7953\ddagger$	$0.5888\ddagger$	-0.0465	0.0513
v_{11}	$7.2999\ddagger$	$5.1755\dagger$	$0.1325\ddagger$	0.0626
v_{22}	$11.0722\ddagger$	$0.1147\ddagger$	$0.2813\ddagger$	0.0826
v_{33}	$3.7626\ddagger$	$6.1956\dagger$	$4.8793\dagger$	0.0908
v_{44}	$0.4157\ddagger$	$0.0047\dagger$	$2.0813\dagger$	0.0013
v_{12}	$7.6023\ddagger$	$6.3457\dagger$	$-1.4961\ddagger$	0.0663
v_{13}	$0.2396\ddagger$	$0.0150\dagger$	$-4.2764\ddagger$	0.0030
v_{14}	$9.8951\dagger$	$0.0017\dagger$	$-1.8455\ddagger$	$10.4620\ddagger$
v_{23}	$0.2767\ddagger$	$0.0167\dagger$	$-5.4043\ddagger$	$26.1158\dagger$
v_{24}	$0.1116\ddagger$	$0.0020\dagger$	$-2.2810\ddagger$	$9.6149\ddagger$
v_{34}	$8.1352\dagger$	$0.0011\dagger$	$2.6476\ddagger$	$5.6627\ddagger$
ϱ_{12}	0.8290	0.0145	-0.1860	0.9963
ϱ_{13}	0.0462	0.0162	-0.3809	0.4942
ϱ_{14}	0.0536	0.0174	-0.3572	0.5400
ϱ_{23}	0.0465	0.0166	-0.3530	0.5082
ϱ_{24}	0.0493	0.0175	-0.3646	0.5889
ϱ_{34}	0.0577	0.0246	-0.4511	0.5310

$\dagger: \times 10^{-6}$; $\ddagger: \times 10^{-4}$.

TABLE 4.7. Summary statistics of the observed returns r_i in $\boldsymbol{R}_t$; variance v_{ii} and covariance v_{ij} in $\boldsymbol{V}_t$; and correlation ϱ_{ij} in $\widehat{\boldsymbol{\mathcal{Q}}}$ for BTC, ETH, USO and GLD.

consistency with standard financial time series models, which often incorporate such terms as a baseline for capturing short-term dynamics. The ADF test results indicate that all-time series are stationary, as evidenced by the rejection of the null hypothesis of a unit root.

For the analyses of volatility and covariance, the behaviour of BTC, ETH, and USO differs significantly from the stock market indices examined in the previous application. Specifically, the returns of BTC, ETH, and USO do not exhibit the same degree of volatility persistence that is typically observed in stock market indices. In the ACF plots for volatility, as shown in Figure 4.20, we observe that the autocorrelations for BTC, ETH, and USO decay more rapidly compared to the stock indices we explored in the previous application. This suggests that volatility shocks in these assets are less enduring and dissipate more quickly over time. Furthermore, ACF plots also demonstrate the presence of high persistence in all volatility, covariance and correlation time series.

However, a notable exception arises in the case of USO. The PACF plot for USO volatility in Figure 4.21 reveals that the AR effect does not decay quickly, indicating a more persistent autocorrelation structure. This is consistent with the unique characteristics of crude oil markets, where volatility can be influenced by macroeconomic shocks, geopolitical events, and supply-demand imbalances that often create prolonged periods of elevated volatility. For instance, events

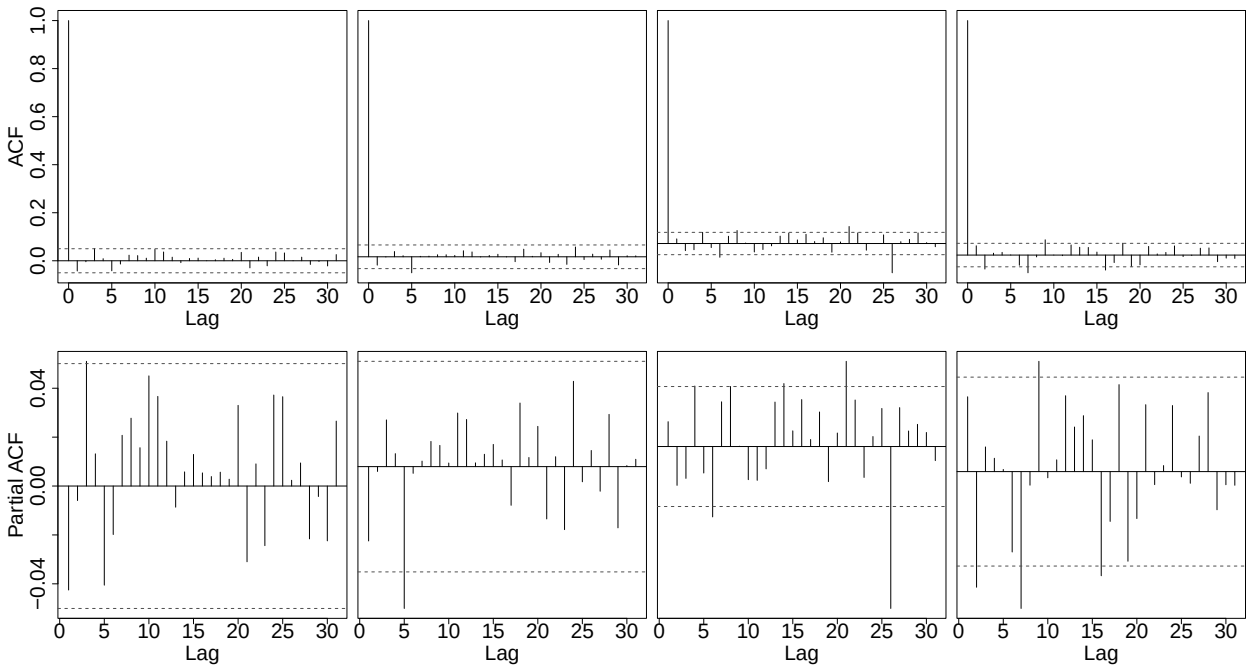


FIGURE 4.19. (a) Row one: ACF plots; (b) Row two: PACF plots for the whole period of BTC, ETH, USO and GLD return series respectively.

such as supply disruptions or changes in OPEC production quotas[2] can have lasting impacts on oil prices, contributing to stronger persistence in volatility.

These observations highlight the distinct features of the volatility and covariance dynamics for BTC, ETH, and USO compared to traditional stock indices. The relatively low volatility persistence for BTC and ETH aligns with the speculative and highly reactive nature of cryptocurrency markets, where price movements are often driven by investor sentiment, market news, and technological developments. In contrast, the higher persistence in crude oil volatility underscores the importance of structural and macroeconomic factors in commodity markets.

4.3.3. Leverage effect by asymmetric AR effect and order selection. Similar to the well-known GJR-GARCH model, the leverage effect can be captured by introducing asymmetry into the volatility process in (1.42). Specifically, this asymmetry is incorporated by allowing the autoregressive term in the volatility process to become larger when returns are negative, reflecting the empirical observation that markets tend to react more strongly to adverse shocks or negative

[2]These quotas are upper limits on the amount of oil each member of the Organization of the Petroleum Exporting Countries (OPEC) is allowed to produce.

TABLE 4.8. Test results of Ljung-Box test and Augmented Dickey-Fuller test of BTC, ETH, USO and GLD time series for the whole period.

	Ljung–Box test		ADF test		KPSS test	
Order	Statistic	p-value	Statistic	p-value	Statistic	p-value
r_1	2.7707	0.0960	-9.9568	< 0.01	0.2002	> 0.1
r_2	1.9157	0.1663	-10.7718	< 0.01	0.1980	> 0.1
r_3	0.6583	0.4171	-11.7184	< 0.01	0.1189	> 0.1
r_4	2.4101	0.1205	-10.9654	< 0.01	0.0477	> 0.1
v_1	283.7152	0.0000	-9.6696	< 0.01	0.2894	> 0.1
v_2	152.2738	0.0000	-9.5042	< 0.01	0.4046	0.0752
v_3	128.5931	0.0000	-6.7240	< 0.01	0.3268	> 0.1
v_4	652.3231	0.0000	-6.7250	< 0.01	0.4180	0.0694
v_{12}	207.1362	0.0000	-9.7440	< 0.01	0.2835	> 0.1
v_{13}	260.1859	0.0000	-8.1641	< 0.01	0.2132	> 0.1
v_{14}	164.3688	0.0000	-9.1958	< 0.01	0.4555	0.0532
v_{23}	173.5078	0.0000	-8.4830	< 0.01	0.3214	> 0.1
v_{24}	125.7279	0.0000	-8.7923	< 0.01	0.5582	0.0286
v_{34}	424.0538	0.0000	-10.0474	< 0.01	0.5208	0.0370
ϱ_{12}	125.7316	0.0000	-8.4059	< 0.01	6.4301	< 0.01
ϱ_{13}	33.9251	5.7274×10^{9}	-7.3500	< 0.01	2.1607	< 0.01
ϱ_{14}	64.4345	9.9920×10^{-16}	-7.4764	< 0.01	3.0301	< 0.01
ϱ_{23}	42.5496	6.8912×10^{-11}	-7.3596	< 0.01	2.3684	< 0.01
ϱ_{24}	69.4450	1.1102×10^{-16}	-7.5117	< 0.01	2.8872	< 0.01
ϱ_{34}	41.9691	9.2728×10^{-11}	-8.4257	< 0.01	2.8599	< 0.01

returns. We denote this as an asymmetric volatility model (model C) :

$$\text{1-stage: } \boldsymbol{V}_t \sim \text{W}_d(\boldsymbol{\Lambda}_t, \nu), \tag{4.16}$$

$$\boldsymbol{\Lambda}_t = \text{diag}(\sigma_{1,t}, \ldots, \sigma_{d,t})\boldsymbol{Q}_t\text{diag}(\sigma_{1,t}, \ldots, \sigma_{d,t}),$$

$$\sigma^2_{i,t} = a_{i0} + \sum_{k=1}^{p}(a_{ik} + a'_{ik}\mathbf{1}_{r_{i,t-k}<0})v_{ii,t-k} + b_i\sigma^2_{i,t-1}, \tag{4.17}$$

$$\boldsymbol{Q}_t = \boldsymbol{\Psi} + \alpha(\widehat{\boldsymbol{\mathcal{Q}}}_{t-1} - \boldsymbol{\Psi}) + \beta(\boldsymbol{Q}_{t-1} - \boldsymbol{\Psi}), \tag{4.18}$$

$$\text{2-stage: } \boldsymbol{R}_t = \boldsymbol{\mu}_t + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \text{ST}_d(\mathbf{0}, \boldsymbol{\Omega}\nu\boldsymbol{\Lambda}_t\boldsymbol{\Omega}, \upsilon), \tag{4.19}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{R}_{t-1} \tag{4.20}$$

where a'_{ik} are additional parameters for the negative lag-1 returns on volatilities. To evaluate the performance of model C, we compare its DIC with that of model B (correlation model) across different combinations of (p, q) orders, where p and q represent the order of the autoregressive and moving average components, respectively, in the volatility process. A summary of the results is presented in Table 4.9. The DIC serves as a model selection criterion, balancing model fit and complexity, where a lower DIC indicates a better trade-off between accuracy and parsimony.

Upon examining the results, we first observe that for any combination of (p, q) orders, model C consistently achieves a lower DIC compared to model B across both Stage 1 and Stage 2 of modelling covariances and returns respectively. This finding underscores the importance of including

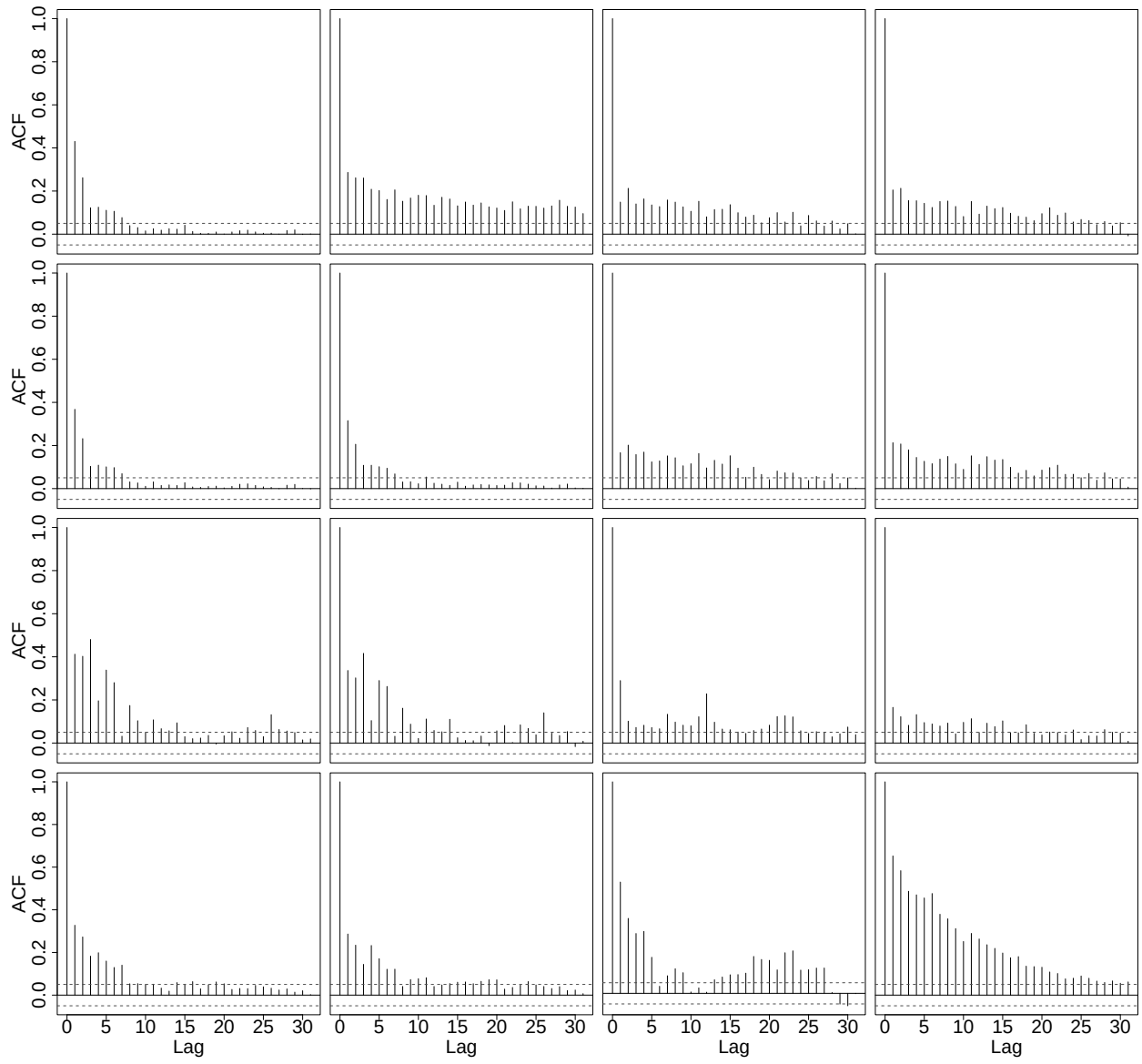


FIGURE 4.20. ACF plots of BTC, ETH, GLD and USO return series respectively for the whole period. The diagonal the plot of observed volatilities. The plots in the upper triangle are for observed correlations, while the plots in the lower triangle are for observed covariances.

the leverage effect to capture the volatility dynamics, as it significantly improves the model performance. Further inspection of the results reveals that among models with orders up to $(3, 3)$, the configuration $(2, 3)$ yields the best performance, as indicated by the lowest DIC values. This suggests that an autoregressive order $p = 2$ is sufficient to capture the short-term dynamics in the volatility process.

While increasing the persistence order q beyond 3 does yield further reductions in DIC for both stages, the improvements become marginal at higher values of q. Specifically, the reduction in DIC from order $(2, 4)$ to $(2, 5)$ and subsequently to $(2, 6)$ is relatively small, particularly in Stage 2. This diminishing improvement in performance suggests that the additional complexity introduced

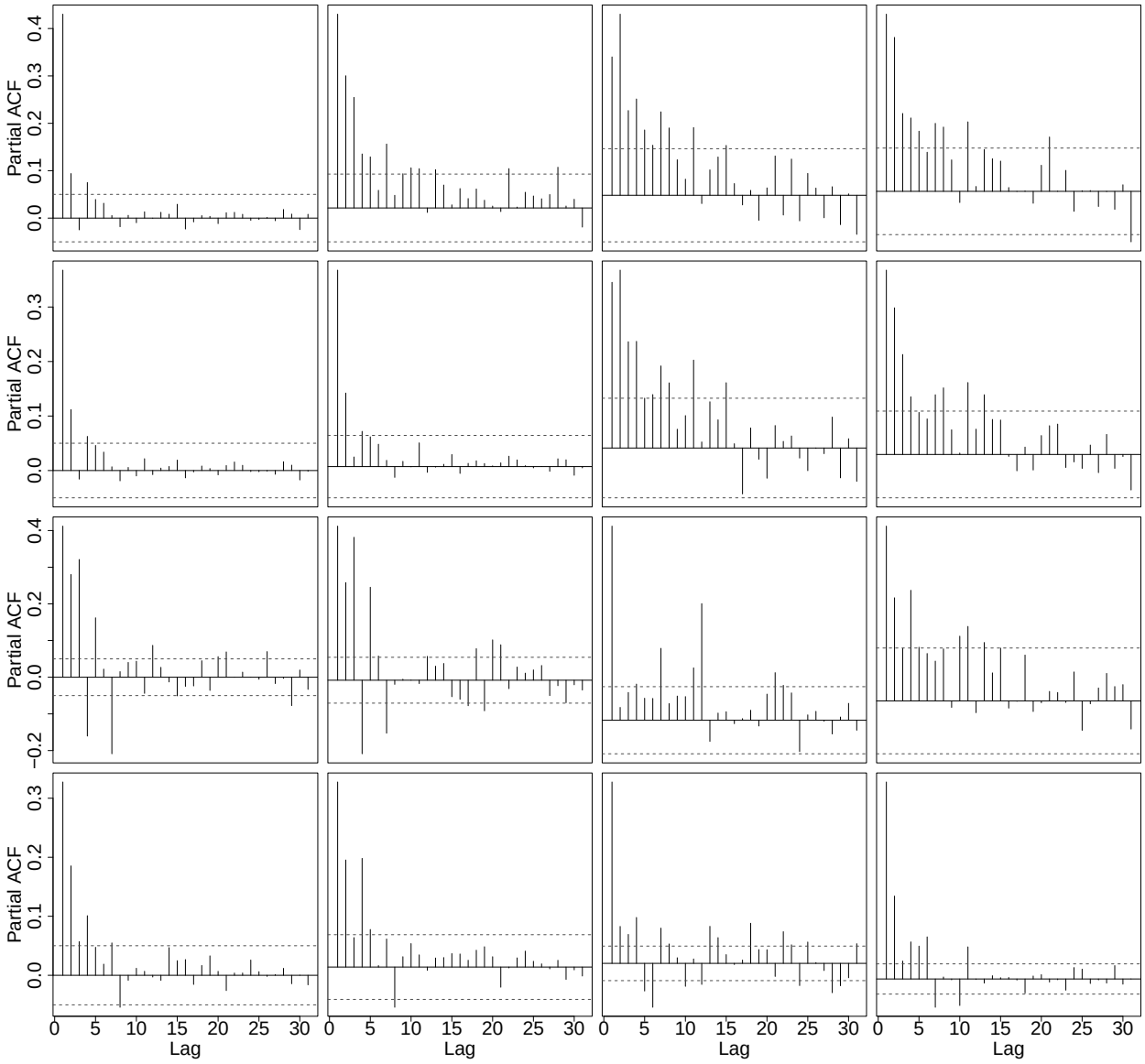


FIGURE 4.21. PACF plots of BTC, ETH, USO and GLD time series respectively for the whole period. The diagonals are plots for observed volatilities. The plots in the upper triangle are for observed correlations, while the plots in the lower triangle are for observed covariances.

by higher q values offers limited practical benefit. To balance model performance with simplicity and computational efficiency, we choose the $(2, 4)$ order for model C. This choice captures the key dynamics of the leverage effect and volatility clustering without introducing unnecessary complexity. The selection of $p = 2$ and $q = 4$ ensures that the model remains parsimonious while effectively describing the asymmetric behaviour and persistence observed in the volatility process.

We further examine the in-sample fit of both volatility and returns. We also consider the two types of constructed assets we discussed in Section 4.2.6. Given that the results exhibit similar patterns across different pairs and sub-periods, we provide the plots for in-sample fit of variances/covariances/correlations as well as returns/VaR of indices/equal weight portfolio/exchange

TABLE 4.9. DIC and p_D (in parentheses) for the two-stage MCARR-Return model with the covariance matrix estimated by model B (correlation model) and C (asymmetric volatility model) applied to the whole period of BTC, ETH, USO and GLD time series.

	Model B		Model C	
Order	1-stage	2-stage	1-stage	2-stage
(1,1)	−226177.19 (20.01)	−31004.29 (12.99)	−226184.68 (21.00)	−31009.69 (12.89)
(1,2)	−226226.68 (22.02)	−31010.63 (12.92)	−226233.13 (23.29)	−31015.69 (12.77)
(1,3)	−226250.24 (23.60)	−31009.75 (12.84)	−226261.33 (24.60)	−31015.33 (12.84)
(2,1)	−226252.63 (22.09)	−31016.80 (13.01)	−226273.06 (23.77)	−31024.41 (12.87)
(2,2)	−226307.57 (22.65)	−31023.76 (12.92)	−226331.55 (25.23)	−31030.45 (12.69)
(2,3)	−226346.83 (25.56)	−31025.98 (12.86)	−226381.21 (27.89)	−31033.14 (12.80)
(3,1)	−226256.33 (23.29)	−31016.24 (12.91)	−226268.19 (26.43)	−31022.16 (12.78)
(3,2)	−226296.96 (23.09)	−31020.71 (12.75)	−226311.53 (25.99)	−31026.88 (12.82)
(3,3)	−226338.08 (25.68)	−31023.97 (13.01)	−226364.96 (28.44)	−31031.95 (12.91)
(2,4)	**−226355.92 (26.93)**	**−31029.22 (13.00)**	**−226388.31 (29.23)**	**−31036.35 (12.94)**
(2,5)	−226405.20 (27.62)	−31030.17 (12.92)	−226434.72 (29.94)	−31036.87 (12.83)
(2,6)	−226440.42 (28.50)	−31031.41 (12.88)	−226468.10 (30.96)	−31037.69 (12.79)
(2,7)	−226454.54 (29.58)	−31039.59 (13.04)	−226483.42 (31.68)	−31046.14 (12.91)
(2,8)	−226468.52 (29.95)	−31037.64 (12.91)	−226497.78 (32.25)	−31044.61 (12.82)

rate in Figures 4.22 and 4.23 for model B, and Figures 4.24 and 4.25 for model C, focusing specifically on the model with order $(2, 4)$, which has been identified as the preferred specification based on its performance. The summary of in-sample VR and QL of VaR is provided in Table 4.10. Again, the VRs can be checked from the 1% VaR in Figures 4.23 and 4.25. We observe that the QL of model C is lower than that of model B for the majority of asset pairs, indicating that model C provides superior VaR estimation. There are some significant Kupiec test results showing inconsistency between VR and $\pi = 0.01$ in H_0 in (4.14), but they are consistent between models B and C. These findings align with the model selection results based on the Deviance Information Criterion (DIC), further supporting the robustness of model C.

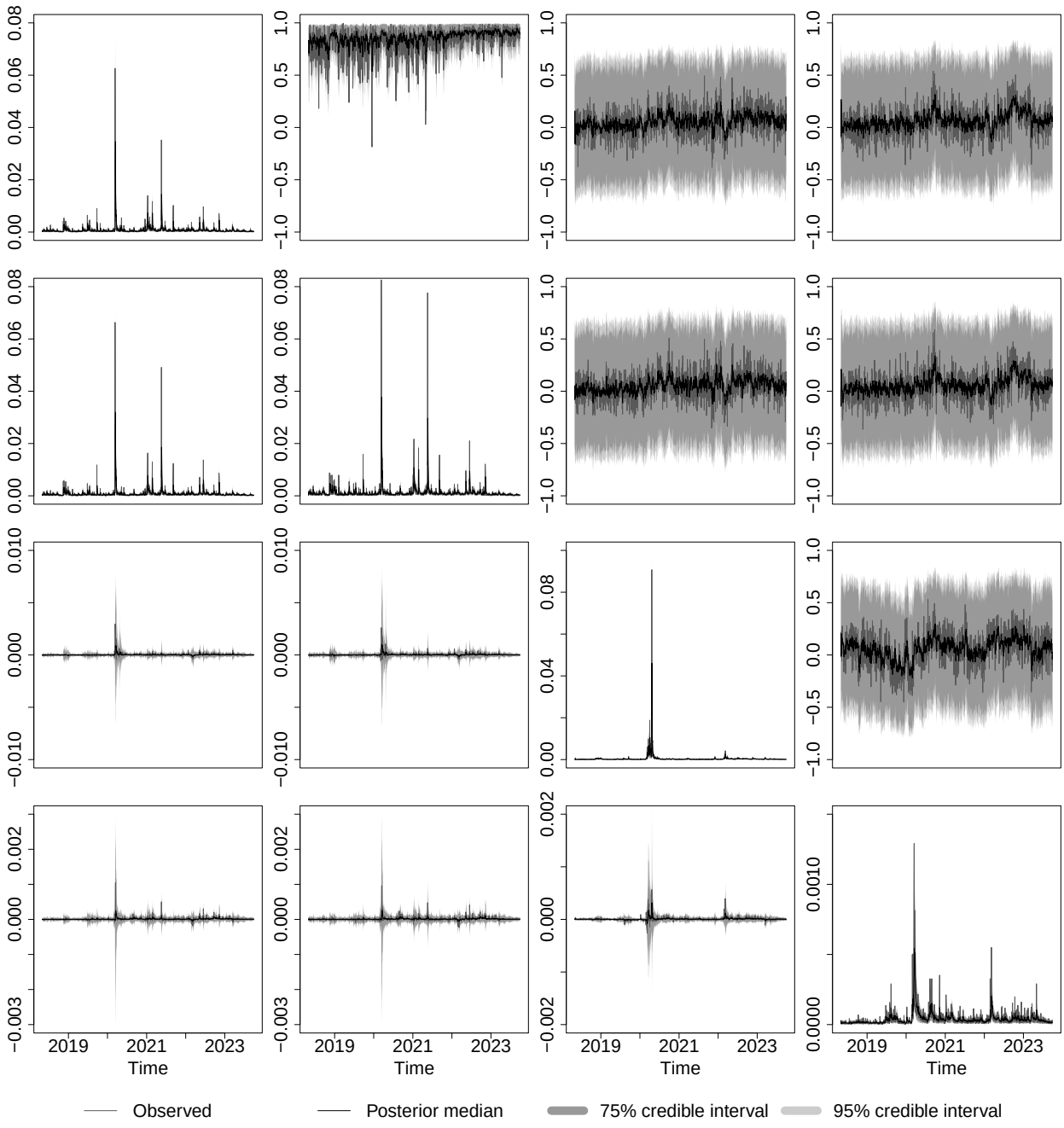


FIGURE 4.22. Observed, posterior median, 1% VaR and credible intervals for BTC, ETH, USO and GLD during the entire period using the two-stage MCARR-Return model with the covariance matrix estimated by model B (correlation model). The diagonals represent volatility. The plots in the lower triangle are for covariances. The plots in the upper triangle are for correlations.

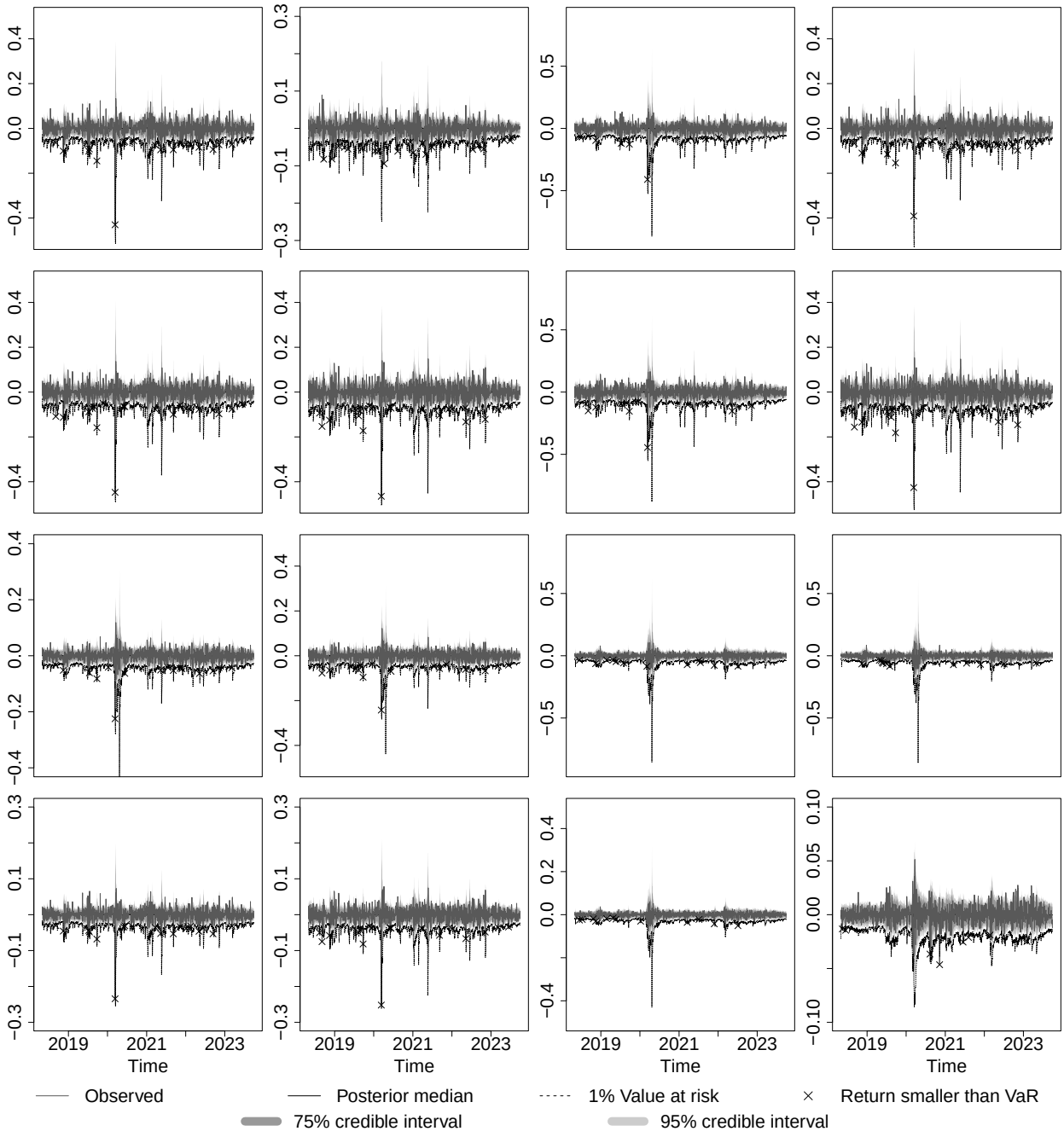


FIGURE 4.23. Observed, posterior median, 1% VaR and credible intervals for BTC, ETH, USO and GLD time series during the entire period using model B. The diagonals are for each return series. The plots in the upper triangle are for exchange rates, and the plots in the lower triangle are for equal-weight portfolios.

From these plots, we observe that for asset pairs with low correlation, the credible intervals for the estimated volatility and covariance matrices are notably wide. This finding indicates substantial uncertainty in the estimated correlations for pairs that exhibit weak dependence. Specifically,

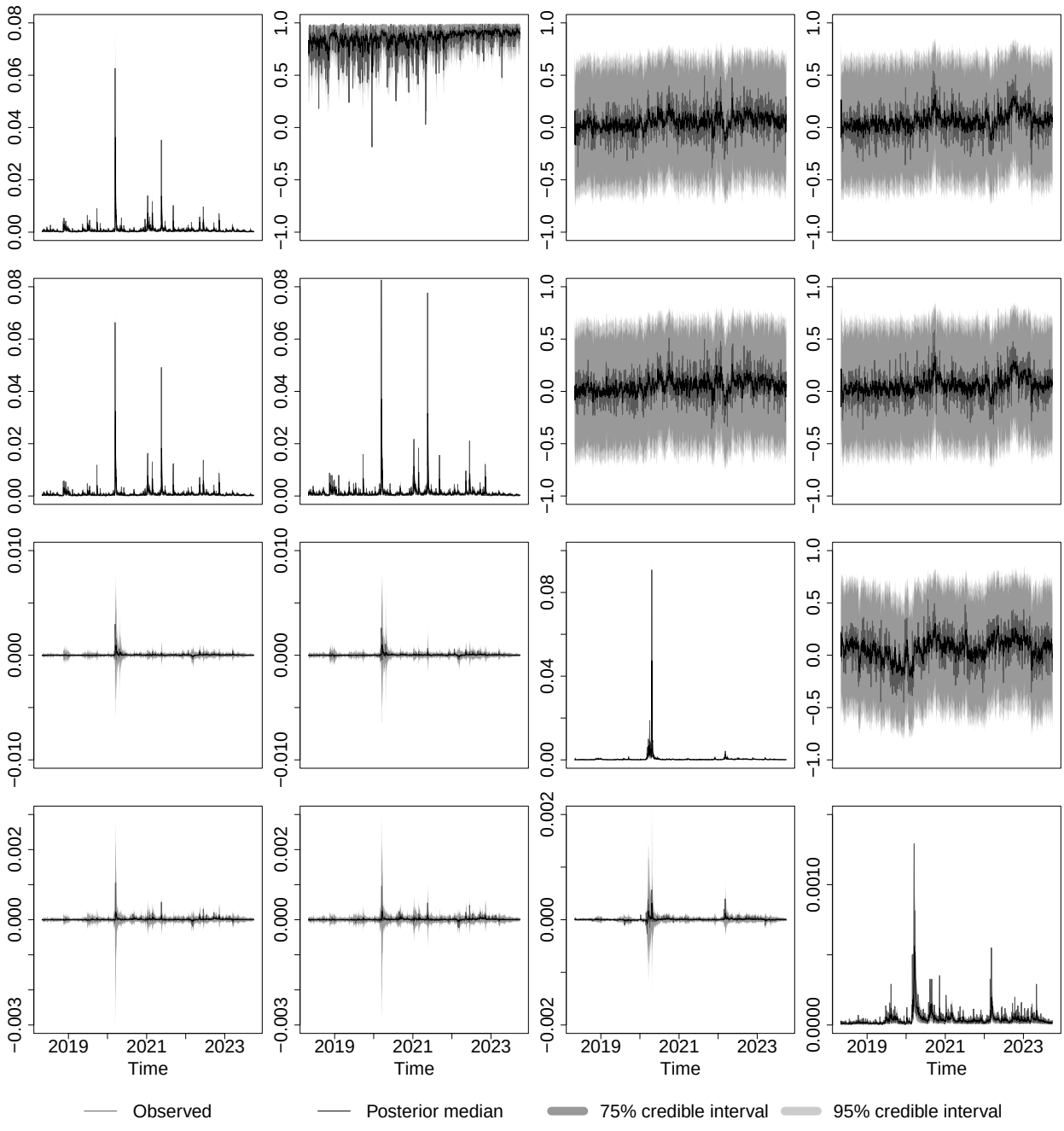


FIGURE 4.24. Observed, posterior median, 1% VaR and credible intervals for BTC, ETH, USO and GLD during the entire period using model C (asymmetric volatility model). The plots in the diagonals represent volatilities. The plots in the lower triangle are for covariances. The plots in the upper triangle are for correlations.

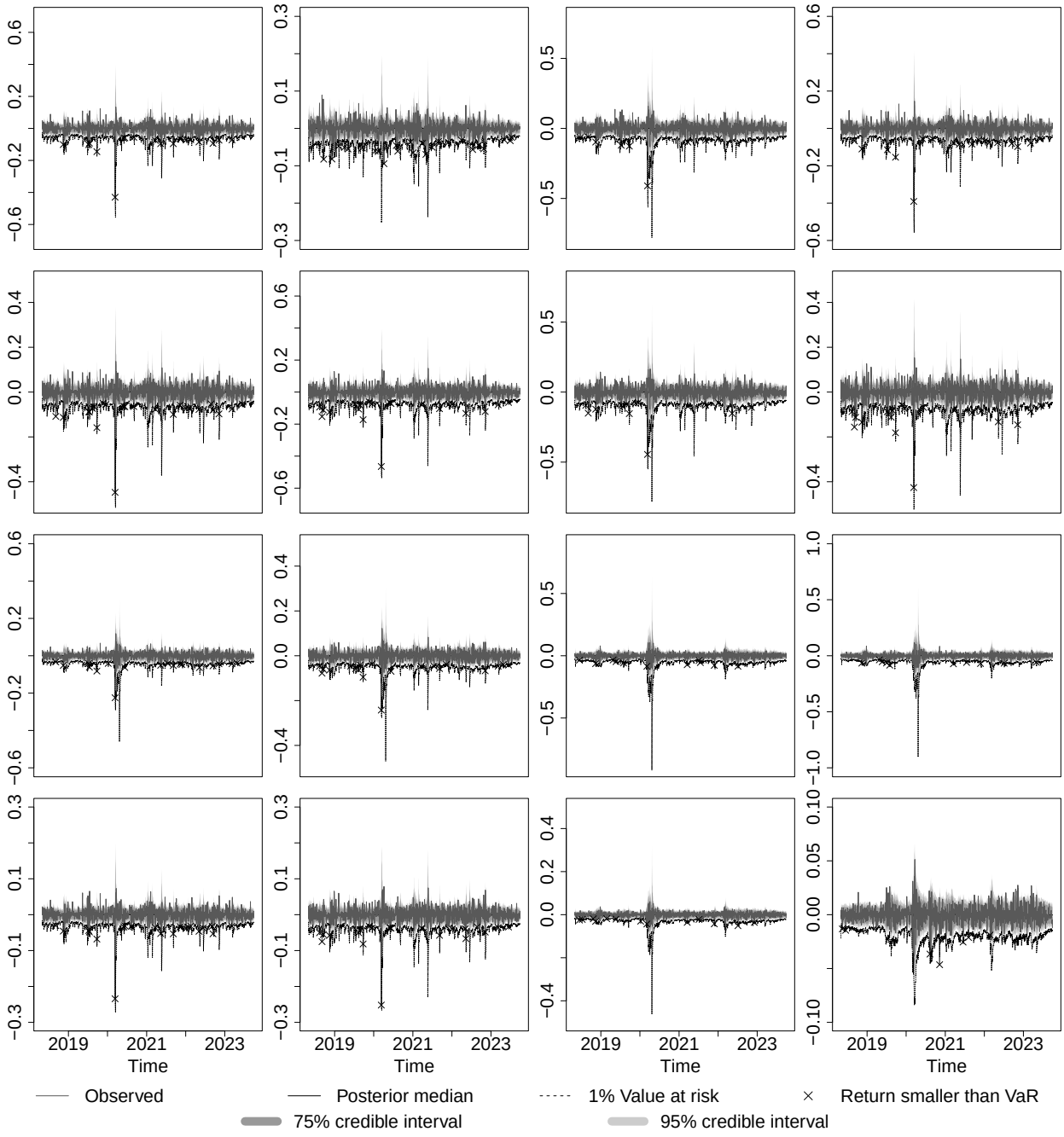


FIGURE 4.25. Observed, posterior median, 1% VaR and credible intervals for BTC, ETH, USO and GLD time series during the entire period using model C (asymmetric volatility model). The diagonals are for each return series. The plots in the upper triangle are for exchange rates, and the plots in the lower triangle are for equal-weight portfolios.

TABLE 4.10. Summary of violation rate, p-value of the Kupiec test and quantile loss using the two-stage MCARR(2,4)-return model using model B (correlation model) and model C (asymmetric volatility model) for BTC, ETH, USO and GLD indices during the forecast period.

	Model B			Model C		
Return	VR	p-value	QL	VR	p-value	QL
r_1	0.0147	0.1031	1.7394	0.0139	0.1718	1.7026
r_2	0.0147	0.1031	2.0597	0.0154	0.0634	2.0555
r_3	0.0081	0.4660	0.9512	0.0081	0.4660	0.9317
r_4	0.0051	0.0447†	0.3270	0.0051	0.0447†	0.3265
$r_1/2 + r_2/2$	0.0183	0.0058†	1.8528	0.0176	0.0109†	1.8144
$r_1/2 + r_3/2$	0.0110	0.7150	0.9417	0.0132	0.2577	1.8144
$r_1/2 + r_4/2$	0.0139	0.1718	0.8808	0.0147	0.1031	0.8651
$r_2/2 + r_3/2$	0.0132	0.2577	1.1120	0.0147	0.1031	1.0917
$r_2/2 + r_4/2$	0.0154	0.0634	1.0244	0.0154	0.0634	1.0229
$r_3/2 + r_4/2$	0.0081	0.4660	0.5021	0.0095	0.8516	0.5051$^{\diamond}$
$r_1 - r_2$	0.0176	0.0109†	0.9453	0.0191	0.0027†	0.9540$^{\diamond}$
$r_1 - r_3$	0.0081	0.4660	1.6401	0.0066	0.1787	1.6221
$r_1 - r_4$	0.0139	0.1718	1.6888	0.0132	0.2577	1.6586
$r_2 - r_3$	0.0103	0.9118	1.9835	0.0103	0.9118	1.9854$^{\diamond}$
$r_2 - r_4$	0.0139	0.1718	2.0440	0.0139	0.1718	2.0399
$r_3 - r_4$	0.0103	0.9118	0.9845	0.0073	0.2926	0.9548

$\diamond$: QL of model C is higher than model B; †: VR is significantly different from 0.01.

the induced correlation matrix derived from the Wishart distribution appears to be overly dispersive around zero correlation. As a result, the variability of the covariance estimates tends to be overstated, leading to excessively large credible intervals.

The observed dispersion is a well-known property of the Wishart distribution when applied to low-correlation settings. In cases where the true correlations are close to zero, the inherent variability of the Wishart process can result in inflated uncertainty estimates. This occurs because the Wishart distribution tends to spread probability mass more broadly around zero, which makes the credible intervals disproportionately wide compared to what would be expected under a more concentrated distribution. Consequently, the variability in the covariance estimates may appear exaggerated, even if the underlying data exhibit minimal correlation. From a practical standpoint, the large credible intervals observed for weakly correlated pairs suggest that the covariance estimates may be less reliable in such cases. This result underscores the importance of considering alternative prior formulations or modifications to the covariance structure that can provide tighter estimates of uncertainty for weakly dependent assets. The in-sample fit analysis reveals that while the proposed model performs well overall, the induced correlation matrix of the Wishart distribution tends to overestimate the variability of covariance for asset pairs with low correlation.

4.3.4. Three modeling periods, MCARR model order and additivity of DIC. To better analyse the temporal dynamics of volatility and correlation, we divide the overall period of interest into three distinct sub-periods:

- Pre COVID-19 period 1: 1st May 2018 to 1st March 2020,

- COVID-19 period 2: 1st March 2020 to 1st Jan 2022,
- Post COVID-19 period 3: 1st Jan 2022 to 1st Oct 2023.

This segmentation allows us to identify changes in model performance and parameter behaviour across time, particularly during periods of significant market stress or structural shifts.

It is important to note that the deviance used in model selection represents the sum of the log-likelihood over the whole period. As such, when the data is split into multiple sub-periods, the sum of the DIC values across the three sub-periods should approximate the DIC in (3.41) of a single model applied to the full dataset but with potentially time-varying coefficients. This is sensible as the flexibility of a model which combines multiple models is also penalised by the sum of the effective number of parameters defined in (3.42) of each model. We call this model taking a segmented approach as a 3-period model. This approach enables a more granular analysis, as splitting the dataset into sub-periods allows the model to accommodate temporal variations in parameters that may arise due to changing market conditions.

The summary of DIC and p_D values for the three-period MCARR$(1,1)$-return and MCARR$(2,4)$-return models are presented in Table 4.11. Comparing order (1,1) and (2,4) MCARR models, it is clear that order (2,4) models are consistently better but the increase in p_D is less than 16, the actual increase of the number of parameters. Such model improvement for MCARR also brings consistent improvement to return models. Comparing with the whole-period models in boldface in Table 4.9, we observe that applying a single model to the whole period results in higher DIC values compared to the models fitted separately to each of the three sub-periods. This suggests that the model's performance improves when the temporal heterogeneity of the data is accounted for, reinforcing the need for a segmented approach to better capture shifts in volatility and correlation dynamics. We also observe that the p_D of the 1-stage MCARR(2,4) model is significantly smaller than that of the other models in period 3. This phenomenon may be attributed to the higher degree of cross-dependency within the parameter space during Period 3, which could have influenced the model's dynamics and estimation performance.

TABLE 4.11. DIC and effective number of parameters p_D (in parentheses) for the 3-period MCARR(1,1)-return and MCARR(2,4)-return models using model B (correlation model) for BTC, ETH, USO and GLD indices.

	MCARR(1,1)-return		MCARR(2,4)-return	
Period	1-stage	2-stage	1-stage	2-stage
Period 1	−109904.13 (20.49)	−10871.93 (12.78)	−109975.40 (22.81)	−10889.25 (12.80)
Period 2	−106745.04 (19.94)	−9989.28 (12.49	−106821.57 (23.48)	−10005.67 (12.66))
Period 3	−105157.21 (24.21)	−10176.08 (12.61)	−105260.08 (14.17)	−10182.47 (12.68)
Total	−321806.38 (64.63)	−31037.30 (37.88)	−322057.06 (60.47)	−31077.39 (38.14)

To further investigate the source of these improvements, Table 4.12 provides a detailed comparison of the estimated model parameters across the three sub-periods. All parameters are significant as they are greater than two times the standard error. We first observe that the short-term effect (relative to persistence) in the volatility process, represented by the order-1 parameters a_{i1} (b_{i1}),

is highest (lowest) for BTC, ETH, and USO during Period 2 compared to Period 1 and Period 3. This finding suggests that the volatility dynamics for these assets are significantly more pronounced during the COVID-19 period. The heightened a_{i1} (lowered b_{i1}) values indicate increased sensitivity to immediate past shocks, which aligns with the general market turbulence and heightened uncertainty experienced during the pandemic. Financial markets, particularly those involving high-risk assets like cryptocurrencies (BTC, ETH) and commodities such as crude oil (USO), are known to exhibit amplified volatility in response to systemic crises. The dramatic shifts in investor sentiment, coupled with rapid news dissemination and policy changes, likely contributed to these elevated short-term effects. In contrast, GLD (gold) displays a markedly different pattern. The short-term effect for gold during Period 2 does not exhibit the same level of increase observed in BTC, ETH, and USO. This divergence may be attributed to the unique role of gold as a safe-haven asset. During periods of market distress, investors often turn to gold as a store of value and a hedge against financial instability. As a result, gold's price dynamics tend to be less reactive to short-term shocks compared to more speculative or economically sensitive assets. Instead, gold's volatility may remain relatively stable or even decrease during crises, reflecting its role as a stabilising force in investor portfolios. The contrast between the behaviour of GLD and the other assets underscores the heterogeneous impact of the COVID-19 pandemic on different segments of the financial market. While cryptocurrencies and oil are highly reactive to short-term market turbulence, gold's resilience highlights its enduring appeal as a hedge during periods of uncertainty. These findings emphasise the importance of tailoring volatility models to account for asset-specific characteristics and their varying responses to systemic events. Other characteristics, such as the kurtoses (ν, υ; degrees of extremeness) for the distributions of covariances and returns, remain relatively stable across sub-periods.

A further examination of the results reveals the parameters α and a, which govern the autoregressive structure and persistence of volatility, change significantly across the three periods. These findings highlight that although the overall volatility process remains relatively stable, its persistence and responsiveness to shocks vary considerably over time, reflecting evolving market conditions. Furthermore, note that $\omega_i^2 \frac{\upsilon}{\upsilon-2} \approx 1$ for $i = 1, 2$ and $\omega_i^2 \frac{\upsilon}{\upsilon-2} \approx 2$ for $i = 3, 4$. This observation suggests that USO and GLD do not exhibit sufficient trading activity within the 15-minute intervals. As a result, the recorded high and low prices within these intervals may fail to converge to the true high and low values of the underlying latent price process. The result justifies the role of the diagonal matrix $\mathbf{\Omega}$ of adjustment factors ω_i to adjust the possible scale bias for $\nu\mathbf{\Lambda}_t$ in the return model (4.19).

To quantify the changes in the correlation process, we employ the loss of ergodicity (LE) metric defined in (3.61). This metric allows us to assess the variability in correlations across the three subperiods. The results indicate that Period 2 (LE $= 3.7491$) is characterised by a remarkably more volatile correlation process compared to Period 1 (LE $= 1.2316$) and Period 3 (LE $= 0.8796027$). This observation coincides with the onset of the COVID-19 pandemic, a period marked by extreme market uncertainty, significant volatility spikes, and increased interdependence across asset classes. During this time, financial markets experienced unprecedented levels of stress, leading to heightened and unstable correlations, particularly among risky assets such as cryptocurrencies and commodities.

The segmentation of the data into sub-periods thus provides valuable insights into the evolving dynamics of volatility and correlation over time. By comparing model performance and parameter estimates across periods, we can identify structural changes and periods of market turbulence. In

particular, the significant variation in the α and β coefficients, combined with the increased volatility in the correlation process during Period 2, underscores the impact of the COVID-19 pandemic on financial markets. This segmented approach not only enhances model accuracy but also offers a deeper understanding of how market conditions influence the behaviour of key parameters in the volatility and correlation processes.

4.3.5. Study of covariate effects. In this section, we explore the possibility of adding additional covariates in our stage-1 MCARR model. It is natural to believe that factors influencing the return will also have an impact on volatilities. Hence, we first consider the most fundamental factor model in finance, the Capital Asset Pricing Model (CAPM) proposed by Sharpe (1964). CAPM explains the relationship between the expected return of an asset and its risk by assuming a simple linear relationship between the expected return of an asset based on its sensitivity to market risk and the expected return of the overall market. CAPM assumes that investors are rational and seek to maximise returns while minimising risk so that they can diversify away from all risks except market risk. The model suggests that an asset's expected return is determined solely by its risk exposure coefficient beta[3], with higher-beta assets requiring higher expected returns to compensate for their higher risk.

While CAPM is influential in asset pricing and portfolio theory, it is criticised for its simplifying assumptions. The Fama-French Three-Factor Model, a prominent asset pricing model proposed by Fama and French (1993), expands on the CAPM by incorporating two additional factors alongside the traditional market risk factor. The model is given by

$$r_{it} = \text{RF}_t + \beta_{i1}\text{Mkt-RF}_t + \beta_{i2}\text{SMB}_t + \beta_{i3}\text{HLM}_t + \epsilon_{it}. \tag{4.21}$$

The three factors are:

1. Mkt-RF: The excess return of the market over the risk-free rate.
2. SMB: The difference in returns between small-cap and large-cap companies.
3. HML: The difference in returns between companies with high book-to-market ratios (value stocks) and low book-to-market ratios (growth stocks).

The model suggests that smaller companies and companies with higher book-to-market ratios tend to outperform the market on a risk-adjusted basis. By accounting for these factors, the Fama-French model provides a more comprehensive explanation of stock returns than the CAPM, which only considers market risk. The model is now widely adopted in financial research and portfolio management as it offers a more nuanced understanding of the drivers of stock returns, emphasising the importance of firm size and value characteristics in addition to market risk. The daily values of three factors are collected from *Fama-French Data Library*, an open-source research database maintained by Kenneth R. French. We further rescale the data by dividing it by 1000 to ensure it aligns with the scale of $\boldsymbol{V}_t$.

The plot of how 3 factors change across time is provided in Figure 4.26, from which we observe that all three factors exhibited significantly higher variability during period 2 compared to other timeframes. This heightened variability reflects the extraordinary market conditions and elevated uncertainty that characterised the pandemic's impact on financial markets. The increased variability in Mkt-RF can be attributed to the heightened swings in overall market returns during this period, driven by rapidly changing investor sentiment, unpredictable policy responses, and unprecedented economic disruptions.

[3]A beta coefficient is a value that compares the overall volatility of an asset to the overall market. A high beta coefficient indicates that a security is more volatile than the market as a whole.

	Period 1	Period 2	Period 3	Whole period
a_{10}	5.8711E-7	8.0954E-7	2.3077E-7	4.9879E-7
a_{11}	0.3626	0.4697	0.3203	0.4108
a_{12}	0.1885	0.2520	0.2521	0.3006
b_{11}	0.1081	0.0810	0.0542	0.0524
b_{12}	0.0574	0.0468	0.0699	0.0833
b_{13}	0.0288	0.0217	0.1444	0.0253
b_{14}	0.0256	0.0143	0.0214	0.0132
a_{20}	1.3338E-6	1.0959E-6	2.1005E-7	7.2448E-7
a_{21}	0.3468	0.4387	0.2841	0.3888
a_{22}	0.1468	0.2581	0.2914	0.2999
b_{21}	0.0998	0.0513	0.0405	0.0596
b_{22}	0.0451	0.0336	0.0351	0.0441
b_{23}	0.0198	0.0653	0.1740	0.0539
b_{24}	0.0191	0.0245	0.0883	0.0322
a_{30}	6.1894E-8	6.5775E-8	4.9559E-8	4.9630E-8
a_{31}	0.4113	0.5754	0.3543	0.5073
a_{32}	0.0581	0.0335	0.1571	0.0451
b_{31}	0.0933	0.0474	0.1236	0.0787
b_{32}	0.1000	0.0798	0.1088	0.1105
b_{33}	0.1016	0.1196	0.1147	0.0917
b_{34}	0.1432	0.1091	0.0820	0.1266
a_{40}	5.1449E-9	1.6943E-8	9.0768E-9	4.9349E-9
a_{41}	0.3267	0.2218	0.1972	0.2601
a_{42}	0.0527	0.1575	0.0778	0.1110
b_{41}	0.0590	0.0872	0.1273	0.0770
b_{42}	0.0453	0.0782	0.1625	0.0826
b_{43}	0.3008	0.2874	0.2430	0.3208
b_{44}	0.1688	0.0674	0.1147	0.1167
r_{12}	0.8110	0.8641	0.8875	0.8843
r_{13}	0.0054	0.0576	0.0752	0.0458
r_{14}	0.0021	0.0913	0.1015	0.0778
r_{23}	0.0032	0.0558	0.0739	0.0470
r_{24}	0.0045	0.0873	0.0973	0.0733
r_{34}	0.0136	0.0702	0.1187	0.0613
α	0.2811	0.3184	0.1211	0.2669
β	0.0662	0.4359	0.4378	0.6441
ν	8.1778	8.3700	10.2650	8.6291
ϕ_{01}	1.2827E-4	0.0024	0.0018	−5.6189E-5
ϕ_{02}	0.0014	0.0028	0.0019	−3.7805E-4
ϕ_{03}	−7.4691E-4	0.0014	16.6617	1.8394E-4
ϕ_{04}	−9.8695E-5	3.2383E-4	−3.0218E-4	−6.3512E-5
ϕ_{11}	0.0414	0.0308	−0.0451	0.0387
ϕ_{12}	−0.0069	0.0161	0.0437	0.0256
ϕ_{13}	0.0024	0.0094	0.0865	0.0282
ϕ_{14}	0.0140	0.0747	0.0271	0.0424
υ	4.5140	4.5210	4.5291	4.4247
ω_1	0.7464	0.7475	0.7470	0.7378
ω_2	0.7380	0.7363	0.7393	0.7297
ω_3	1.0478	1.0397	1.0467	1.0375
ω_4	1.0643	0.9632	1.0637	1.0375

TABLE 4.12. Parameter estimates of the two-stage MCARR$(2,4)$-return models using model B (correlation model) in the 3-period and whole-period for BTC, ETH, USO and GLD indices. All parameters are all significant (p-value< 0.05).

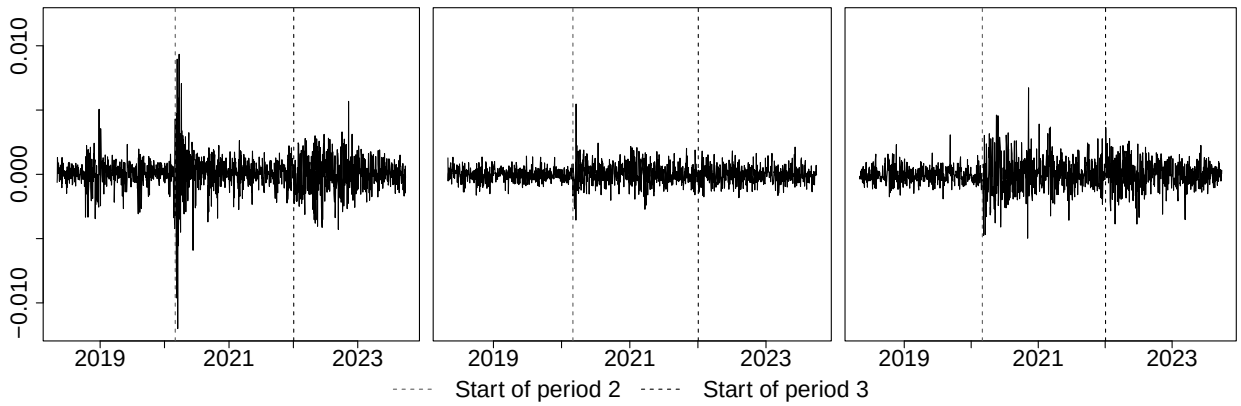


FIGURE 4.26. Plots of Mkt-RF, SMB and HML across three periods.

As the COVID-19 pandemic unfolded, global markets experienced sharp declines followed by rapid recoveries, reflecting the significant uncertainty about the pandemic's trajectory and its economic consequences. Similarly, the variability in SMB, which measures the performance differential between small-cap and large-cap stocks, highlights the differing sensitivities of firms to the pandemic's economic impact. Small-cap companies, often with less diversified revenue streams and limited financial resilience, were particularly vulnerable to the adverse effects of lockdowns and supply chain disruptions. Conversely, large-cap firms, many of which are better capitalised and globally diversified, demonstrated relative stability, contributing to the heightened dispersion in the SMB factor.

The HML factor, which captures the performance difference between value stocks and growth stocks, also experienced elevated variability. This may reflect the contrasting performance dynamics between value stocks, which are often tied to traditional industries disproportionately affected by the pandemic (e.g., energy, manufacturing), and growth stocks, particularly those in technology and healthcare, which benefited from the accelerated shift to digital services and the global focus on healthcare solutions. The higher variability observed in all three factors underscores the profound impact of COVID-19 on different segments of the market and the dynamics of asset volatility. We are interested in determining whether the impact of COVID-19 can be effectively captured using these factors.

4.3.5.1. *Additional covariate effects.* Adding these factors to volatility and ensuring that it is always non-negative is not straightforward. Let $\boldsymbol{x}_t = (x_{1,t}, ..., x_{d,t})^\top$ be the vector of d covariates we measured at day t, where d is the number of covariates to be included in our model. We explore two different approaches. The first formulation adds the quadratic effect of the factor. It is the additive covariate model (model D) given by

$$\begin{aligned}
\text{1-stage: } \boldsymbol{V}_t &\sim \mathrm{W}_d(\boldsymbol{\Lambda}_t, \nu), && (4.22)\\
\boldsymbol{\Lambda}_t &= \mathrm{diag}(\sigma_{1,t}, \ldots, \sigma_{d,t})\boldsymbol{Q}_t\mathrm{diag}(\sigma_{1,t}, \ldots, \sigma_{d,t}),\\
\sigma^2_{i,t} &= a_{i0} + \sum_{k=1}^{p} a_{ip} v_{ii,t-p} + \sum_{k=1}^{q} b_{ik}\sigma^2_{i,t-k} + \boldsymbol{c}_i^\top(\boldsymbol{x}_{t-1} \odot \boldsymbol{x}_{t-1}), && (4.23)\\
\boldsymbol{Q}_t &= \boldsymbol{\Psi} + \alpha(\widehat{\boldsymbol{\mathcal{Q}}}_{t-1} - \boldsymbol{\Psi}) + \beta(\boldsymbol{Q}_{t-1} - \boldsymbol{\Psi}), && (4.24)\\
\text{2-stage: } \boldsymbol{R}_t &= \boldsymbol{\mu}_t + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathrm{ST}_d(\boldsymbol{0}, \boldsymbol{\Omega}\nu\boldsymbol{\Lambda}_t\boldsymbol{\Omega}, \upsilon), && (4.25)\\
\boldsymbol{\mu}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{y}_{t-1}. && (4.26)
\end{aligned}$$

where $\odot$ refers to the entry-wise product. Compared to model B, the additional parameters are $\boldsymbol{c}_i = c_{1:\mathfrak{l}}, i = 1, \ldots, d$ where $\mathfrak{l} = 1, 2, 3$ is the total number of factors in model D. Under this formulation, the square of return factors can be treated as an estimation of the volatility of the factors, which contribute to the total volatility of the assets. The second formation consists of an exponential scaling effect in $\sigma_{i,t}^2$. The exponential scaling covariate model (model E) is given by

$$
\begin{aligned}
\text{1-stage: } \boldsymbol{V}_t &\sim \mathrm{W}_d(\boldsymbol{\Lambda}_t, \nu), && (4.27)\\
\boldsymbol{\Lambda}_t &= \mathrm{diag}(\sigma_{1,t}, \ldots, \sigma_{d,t})\boldsymbol{Q}_t \mathrm{diag}(\sigma_{1,t}, \ldots, \sigma_{d,t}),\\
\sigma_{i,t}^2 &= \left(a_{i0} + \sum_{k=1}^{p} a_{ip} v_{ii,t-p} + \sum_{k=1}^{q} b_{ik}\sigma_{i,t-k}^2\right) \exp(\boldsymbol{c}_i^\top \boldsymbol{x}_{t-1}), && (4.28)\\
\boldsymbol{Q}_t &= \boldsymbol{\Psi} + \alpha(\widehat{\boldsymbol{\mathcal{Q}}}_{t-1} - \boldsymbol{\Psi}) + \beta(\boldsymbol{Q}_{t-1} - \boldsymbol{\Psi}), && (4.29)\\
\text{2-stage: } \boldsymbol{R}_t &= \boldsymbol{\mu}_t + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathrm{ST}_d(\boldsymbol{0}, \boldsymbol{\Omega}\nu\boldsymbol{\Lambda}_t\boldsymbol{\Omega}, \upsilon), && (4.30)\\
\boldsymbol{\mu}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1}. && (4.31)
\end{aligned}
$$

Both formulations guarantee that volatility is positive.

In our study, the maximum number of covariates is 3. Thus, we fit all possible models with different combinations of factors and choose the optimal models. The summary of model fit for the two formulations with different covariates is provided in Table 4.13. The results obtained from our analysis provide some notable insights. First, model E (exponential scaling covariate model) exhibits minimal improvement in stage-2 returns compared to model D (additive covariate model) but is slightly worse in estimating stage-1 covariances. A further examination of the fitted model reveals that the coefficients of all covariates are not statistically significant. This lack of improvement indicates that the inclusion of the additional factors using exponentially scaling structure of model E is not suitable for our dataset and so, does not meaningfully contribute to explaining the observed volatility or covariance dynamics.

Furthermore, compared with model B (basic correlation model), while the inclusion of the Mkt-RF factor in the volatility process appears to reduce the DIC in stage 1, it simultaneously leads to a deterioration of DIC in stage 2. This feature is the same for other models with different combinations of factors. These results suggest that adding, for instance, the Mkt-RF factor, particularly at lag 1, introduces an overfitting problem in the first stage. The model begins to fit the noise in the data rather than capturing the underlying structure, resulting in poorer performance when extended to the second stage of the analysis. Hence, we consider an alternative approach.

4.3.5.2. *Lagged covariates replacing persistence.* We hypothesise that the $\boldsymbol{\Lambda}_{t-1}$ already accounts for the influence of the external factors. Specifically, if we assume that the three factors are independent of each other, then by (4.21) we have

$$
\mathrm{Var}(r_{i,t}) = \beta_{i1}^2 \mathrm{Var}(\text{Mkt-RF}_t) + \beta_{i2}^2 \mathrm{Var}(\text{SMB}_t) + \beta_{i3}^2 \mathrm{Var}(\text{HML}_t) + \mathrm{Var}(\epsilon_{i,t}). \tag{4.32}
$$

Thus $\boldsymbol{\Lambda}_{t-1}$ inherently captures the square of these factors. This interpretation aligns with the idea that persistence terms in volatility processes naturally reflect the cumulative impact of past shocks, including those originating from external factors.

TABLE 4.13. DIC and effective number of parameters p_D (in parentheses) for the whole-period MCARR(2,4)-return D (additive covariate) and E (exponential scaling covariate) models for BTC, ETH, USO and GLD indices.

Models	Model D		Model E	
	1-stage	2-stage	1-stage	2-stage
Benchmark model B	−226355.92	−31029.22	-	-
	(26.93)	(13.00)	-	-
Mkt-RF	−226524.36	−31023.46	−226348.38	−31029.52
	(30.63)	(12.87)	(26.96)	(12.80)
SMB	−226353.35	−31028.99	−226348.20	−31030.19
	(28.07)	(12.94)	(27.52)	(12.58)
HML	−226379.93	−31027.06	−226349.85	−31028.97
	(28.66)	(12.90)	(27.79)	(12.98)
Mkt-RF + SMB	−226522.02	−31022.47	−226341.21	−31029.06
	(31.83)	(12.87)	(27.16)	(13.00)
Mkt-RF + HML	−226529.52	−31021.41	−226342.39	−31028.97
	(32.84)	(12.83)	(27.64)	(12.90)
SMB + HML	−226377.35	−31027.42	−226342.42	−31029.35
	(29.78)	(12.87)	(28.16)	(12.80)
Mkt-RF + SMB + HML	−226527.26	−31022.11	−226334.25	−31029.56
	(33.63)	(12.75)	(28.45)	(12.61)

To address this issue and correctly model the impact of the factors, we propose a refined approach in the factor volatility model (model F) given by:

$$\begin{aligned}
\text{1-stage: } \boldsymbol{V}_t &\sim \mathrm{W}_d(\boldsymbol{\Lambda}_t, \nu), && (4.33)\\
\boldsymbol{\Lambda}_t &= \mathrm{diag}(\sigma_{1,t}, \ldots, \sigma_{d,t})\boldsymbol{Q}_t\mathrm{diag}(\sigma_{1,t}, \ldots, \sigma_{d,t}),\\
\sigma^2_{i,t} &= a_{i0} + \sum_{k=1}^{p} a_{ip} v_{ii,t-p} + \sum_{k=1}^{q} \boldsymbol{c}_{\boldsymbol{ik}}^{\top} \cdot (\boldsymbol{x}_{t-k} \odot \boldsymbol{x}_{t-k}), && (4.34)\\
\boldsymbol{Q}_t &= \boldsymbol{\Psi} + \alpha(\widehat{\boldsymbol{\mathcal{Q}}}_{t-1} - \boldsymbol{\Psi}) + \beta(\boldsymbol{Q}_{t-1} - \boldsymbol{\Psi}), && (4.35)\\
\text{2-stage: } \boldsymbol{R}_t &= \boldsymbol{\mu}_t + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathrm{ST}_d(\boldsymbol{0}, \boldsymbol{\Omega}\nu\boldsymbol{\Lambda}_t\boldsymbol{\Omega}, \upsilon), && (4.36)\\
\boldsymbol{\mu}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1}. && (4.37)
\end{aligned}$$

Relative to model B (basic correlation model), the additional parameters are $\boldsymbol{c}_{ik} = c_{ik,1:\mathfrak{l}}$ where $\mathfrak{l} = 1, 2, 3$ is the number of factors in model F which differs from model D (additive covariate model) and E (exponential scaling covariate model) because order q is also considered for each factor $\boldsymbol{x}_{\mathfrak{l},1:N}, l = 1, 2, 3$ similar to persistence. In model F, we replace the $\boldsymbol{\Lambda_{t-k}}$ with the lagged values of the factors. By explicitly introducing the lagged factors into the model, we aim to separate the direct influence of these factors from the persistence structure. This adjustment ensures that the model avoids overfitting the first stage while appropriately capturing the dynamic contributions of the external factors to volatility and covariance.

The proposed modification in model F allows for a more structured and interpretable representation of the factor effects. By disentangling the factor contributions from the persistence term, we can better assess the true influence of the external factors on the volatility process. This approach

also enhances the flexibility of the model, as it avoids redundancy and ensures that the persistence term does not inadvertently absorb the impact of the factors.

Table 4.14 presents a summary of model F performance in terms of DIC incorporating different combinations of the three factors. A careful examination of the results reveals that none of the models (1-stage and 2-stage) outperform model B with the selected order $(2, 4)$, which serves as our benchmark model. This finding suggests that the inclusion of these factors does not universally enhance the overall model fit for both returns and covariances when compared to the baseline persistence-based model.

However, it is worth noting that the model incorporating both Mkt-RF and SMB factors performs better than model B with a simpler order of $(2, 1)$ when persistence terms are dropped in model F. This result indicates that model F with Mkt-RF and SMB (in boldface) has measurable impacts on volatility. These factors appear to add incremental explanatory power, particularly when the autoregressive and moving average orders of the persistence term are lower, highlighting their importance in environments where simpler volatility structures are considered.

Another significant advantage of model F relative to model B lies in its ability to incorporate investor views about future market behaviour in a structured and transparent manner for the forecasts based on the forecasted Mkt-RF and SMB. Unlike traditional volatility models, which rely solely on historical data to estimate volatility persistence, model F allows for the inclusion of forward-looking factors, such as the forecasted market excess return. For instance, if investors hold the belief that the market excess return will increase in the near future, reflecting an optimistic outlook on overall market performance, model F can incorporate this expectation directly into the volatility process. By including lagged factors explicitly, the model provides a framework for generating volatility and covariance predictions that align with such views.

This feature makes model F particularly appealing in scenarios where subjective investor expectations, economic forecasts, or macroeconomic indicators are considered reliable inputs for modelling volatility. For example, during periods of economic recovery, rising market excess returns and a positive size factor (SMB) may signal higher growth potential, which could directly impact volatility dynamics. Model F accommodates such scenarios by integrating these factors into its predictive framework, offering enhanced flexibility over traditional models that rely purely on persistence terms derived from past data.

To further explore the impact of COVID-19, we divide the entire sample period into three subperiods and reestimate Model F for each segment. This approach allows us to identify temporal changes in the model's parameters that may be associated with the pandemic. Upon analysis, we observed consistent patterns in the parameters α and β, as well as in the short-term effect coefficient a_{ik}. These findings suggest that while the general structure of volatility dynamics remained stable, specific changes in the parameters could reflect underlying shifts induced by the pandemic. As the trends of these parameters across periods mostly agree with the results in Table 4.12 for model B, we omit reporting these parameters.

On the other hand, Table 4.15 presents only the parameters $\boldsymbol{c}_{ik}$ for the factors Mkt-RF and SMB in MCARR(2,4) model F. Notably, we observed some changes in the significance of parameters across the three sub-periods. More factor parameters $\boldsymbol{c}_{ik}$ are significant for crude oil before COVID-19, but they become gradually insignificant across periods. Reverse trends can be observed for ETH, but no other obvious trends can be observed. This variation indicates that the pandemic introduced shifts in market conditions and asset behaviours and, hence, highlights the role of COVID-19 in influencing the dynamics of volatility indirectly by affecting the magnitude

and behaviour of these key parameters of $\boldsymbol{c}_{ik}$ within the model. However, these results also underscore the complexity of capturing the systemic drivers for the effects of the COVID-19 pandemic on market volatility. The influence of factors Mkt-RF and SMB on the COVID-19 pandemic may manifest indirectly through other existing market factors, making it challenging to isolate their impact by simply comparing parameter estimates. Future studies may benefit from incorporating additional macroeconomic or policy-related factors alongside dynamic parameter estimation to provide a more nuanced understanding of how global crises affect market volatility.

Lastly, we also attempt to incorporate COVID-19-specific indices, such as daily new case counts, into our selected model F to examine their potential impact on the dynamics of volatility. However, the parameters are mostly not statistically significant, suggesting that the influence of COVID-19 may already be indirectly captured through existing factors, such as Mkt-RF and SMB, or through the parameters governing the dynamics of volatility. The priors for model B-F are given in Appendix 4.A.

TABLE 4.14. DIC and effective number of parameters p_D (in parentheses) for the whole-period two-stage MCARR(2,4)-return model using model F (factor volatility model) for BTC, ETH, USO and GLD indices. Stage-2 model F with Mkt-RF + SMB has lower DIC than MCARR(2,1)-return model using model B (basic correlation model) and is highlighted in boldface.

Models	Model F	
	1-stage	2-stage
Benchmark model B(2,4)	−226355.92 (26.93)	−31029.22 (13.00)
Benchmark model B(2,1)	−226252.63 (22.09)	−31016.80 (13.01)
Mkt-RF	−226238.08 (34.25)	−31014.02 (12.69)
SMB	−226059.36 (28.89)	−31019.48 (12.80)
HML	−226032.48 (30.04)	−30999.36 (12.69)
Mkt-RF + SMB	−226288.97 (41.10)	**−31023.65** (12.90)
Mkt-RF + HML	−226258.31 (41.52)	−31012.12 (12.87)
SMB + HML	−226093.89 (37.00)	−31013.50 (12.80)
Mkt-RF + SMB + HML	−226297.98 (47.19)	−31021.24 (12.99)

4.4. Conclusion

In this chapter, we present two primary applications of our proposed model. The first application focuses on the U.S. stock market indices, comparing the in-sample performance of model A (covariance model) and model B (correlation model). The efficiency of the range-based measures

	Period 1	Period 2	Period 3
c_{111}	0.0624	0.2932	1.1749†
c_{112}	0.2058	0.4654	0.4020
c_{121}	1.3104†	0.4199	0.8092†
c_{122}	0.1530	5.4744†	1.3513†
c_{131}	0.1045	3.4958†	0.1101
c_{132}	0.3264	0.1608	0.3307
c_{141}	0.1021	0.0555	3.4846†
c_{142}	0.2211	0.1694	0.2660
c_{211}	0.4439	0.1041	1.8593†
c_{212}	0.2392	0.8524	0.8914†
c_{221}	2.4657†	0.0867	2.5238†
c_{222}	0.2299	0.6165	0.1911
c_{231}	0.2944	4.7430†	0.2861†
c_{232}	0.3282	2.3903†	1.9166†
c_{241}	0.2437	0.0996	4.9554†
c_{242}	0.3435	0.2225	4.7252†
c_{311}	1.2219†	0.1458	0.6173†
c_{312}	0.6793†	0.1563	0.2149
c_{321}	0.2416	0.1931	0.1097
c_{322}	1.0513†	0.1345	0.3997
c_{331}	0.3741†	1.3743†	0.2044†
c_{332}	1.2621†	0.7820†	0.5571
c_{341}	0.5579†	0.5624†	0.6734†
c_{342}	1.8464†	0.0807	0.3339
c_{411}	0.0323†	0.0295	0.0638†
c_{412}	0.3419†	0.0288	0.0312
c_{421}	0.0136	0.2371†	0.0104
c_{422}	0.0641	0.0423	0.0275
c_{431}	0.0126	0.1615†	0.0209†
c_{432}	0.0642	0.1550†	0.0235
c_{441}	0.0078	0.0222	0.2972†
c_{442}	0.2375†	0.0638	0.0198

Parameters with † are significant.

TABLE 4.15. The change of parameters c_{ikl} across the 3 periods for MCARR(2,4) model F for index i (BTC, ETH, USO and GLD) of order $k = 1, \ldots, q$ $(q = 4)$ and covariates $l = 1, \ldots, \mathfrak{l}$ $(\mathfrak{l} = 2)$ (Mkt-RF + SMB).

is evaluated by contrasting the model fit using the range correlation and volatility measures with that of the sample covariance measures. The results indicate that the proposed range correlation and volatility measures provide a more efficient framework for modelling the volatility of returns. Additionally, the 1-step-ahead VaR forecasts for the stock market indices, as well as the equal weight portfolio and exchange rate, demonstrate the strong predictive capability of our two-stage MCARR-return model (model B).

The second application aims to analyse the impact of the COVID-19 pandemic on financial markets. We show that the proposed model effectively captures the dynamics of the covariance matrix for assets with diverse characteristics, highlighting its versatility. To investigate temporal changes, we divide the period of interest into three sub-periods, pre, during and post pandemic, and demonstrate that the model parameters vary significantly across these periods, reflecting distinct market behaviours. Furthermore, using the LE measure proposed in (3.61), we conclude that the COVID-19 period exhibits lower ergodicity of correlations, indicating heightened instability in the correlation structure during this time of market stress. Lastly, we explore the potential for incorporating additional features into the model, including model C (asymmetric volatility model) with leverage effect, models D (additive covariate model) and E (exponential scaling covariate model) with covariate effects, and model F (factor volatility model) with lagged covariate effects replacing the lagged persistent effects. Our analysis suggests that including external factors alongside persistence terms can lead to overfitting, particularly in scenarios where the added complexity does not provide substantial improvement in model performance. These findings underscore the importance of balancing model flexibility with parsimony to ensure robustness and interpretable results.

Appendices

4.A. Parametrisation and prior of parameter

The prior specification for the order (p, q) covariance models B, C, D, E and F are as below:

Model B: $(a_{i1}, ..., a_{ip}, b_{i1}, ..., b_{iq}, 1 - \sum_{k=1}^{p} a_{ik} - \sum_{k=1}^{q} b_{ik}) \sim \text{Dir}(2, ..., 2, 3), \quad i = 1, 2, 3,$

$a_{0i} \sim \text{Gamma}(0.5, 1), \quad i = 1, \ldots, 4,$

$(\alpha, \beta, 1 - \alpha - \beta) \sim \text{Dir}(2, 2, 18),$

$\nu - 5 \sim \text{Gamma}(5, 1),$

$\psi_{12}, \psi_{13}, \psi_{14}, \psi_{23}, \psi_{24}, \psi_{34} \sim \text{N}(0.6, 0.2).$

Model C: $(a_{i1}, ..., a_{ip}, a'_{i1}, ..., a'_{ip}, b_{i1}, ..., b_{iq}, 1 - \sum_{k=1}^{p} a_{ik} - \sum_{k=1}^{p} a'_{ik} - \sum_{k=1}^{q} b_{ik}) \sim \text{Dir}(2, ..., 2, 3), \ i = 1, 2, 3,$

$a_{0i} \sim \text{Gamma(0.5,1)}, \ i = 1, 2, 3,$

$(\alpha, \beta, 1 - \alpha - \beta) \sim \text{Dir}(2, 2, 18),$

$\nu - 4 \sim \text{Gamma}(5, 1),$

$\psi_{12}, \psi_{13}, \psi_{14}, \psi_{23}, \psi_{24}, \psi_{34} \sim \text{N}(0.6, 0.2).$

Model D: $(a_{i1}, ..., a_{ip}, b_{i1}, ..., b_{iq}, 1 - \sum_{k=1}^{p} a_{ik} - \sum_{k=1}^{q} b_{ik}) \sim \text{Dir}(2, ..., 2, 3), \ i = 1, 2, 3,$

$a_{0i} \sim \text{Gamma(0.5,1)}, \ i = 1, \ldots, 4$

$c_{ij} \sim \text{Gamma}(0.5, 1),$

$(\alpha, \beta, 1 - \alpha - \beta) \sim \text{Dir}(2, 2, 18),$

$\nu - 4 \sim \text{Gamma}(5, 1),$

$\psi_{12}, \psi_{13}, \psi_{14}, \psi_{23}, \psi_{24}, \psi_{34} \sim \text{N}(0.6, 0.2).$

Model E: $(a_{i1}, ..., a_{ip}, b_{i1}, ..., b_{iq}, 1 - \sum_{k=1}^{p} a_{ik} - \sum_{k=1}^{q} b_{ik}) \sim \text{Dir}(2, ..., 2, 3), \ i = 1, 2, 3,$

$a_{0i} \sim \text{Gamma(0.5,1)}, \ i = 1, \ldots, 4$

$c_{ij} \sim \text{N}(0, 0.1),$

$(\alpha, \beta, 1 - \alpha - \beta) \sim \text{Dir}(2, 2, 18),$

$\nu - 4 \sim \text{Gamma}(5, 1),$

$\psi_{12}, \psi_{13}, \psi_{14}, \psi_{23}, \psi_{24}, \psi_{34} \sim \text{N}(0.6, 0.2).$

$$\text{Model F:}(a_{i1}, ..., a_{ip}, 1 - \sum_{k=1}^{p} a_{ik}) \sim \text{Dir}(2, ..., 2, 3),\ i = 1, 2, 3,$$
$$a_{0i} \sim \text{Gamma(0.5,1)},\ i = 1, \ldots, 4$$
$$c_{ij} \sim \text{Gamma}(0.5, 1),$$
$$(\alpha, \beta, 1 - \alpha - \beta) \sim \text{Dir}(2, 2, 18),$$
$$\nu - 4 \sim \text{Gamma}(5, 1),$$
$$\psi_{12}, \psi_{13}, \psi_{14}, \psi_{23}, \psi_{24}, \psi_{34} \sim \text{N}(0.6, 0.2).$$

CHAPTER 5

Conclusion and further research

5.1. Conclusion

This dissertation presents a series of advancements in modelling and estimating correlation and covariance dynamics, with applications to financial data and broader statistical frameworks. By addressing limitations in existing methodologies, the research offers new tools and insights for capturing the complex relationships among multi-dimensional time series, particularly in volatile and heterogeneous environments.

In Chapter 1, the literature review highlights key challenges in covariance modelling, including issues of heteroskedasticity, non-positive definiteness, and the lack of robust estimators for dynamic correlation structures. The identified research gaps provide the motivation for the proposed methodologies and applications detailed in subsequent chapters.

In Chapter 2, we extended the PK measure from variance estimation to multi-dimensional covariance matrices, proposing novel range-based correlation estimators tailored for high-dimensional data. The key contributions include the development of efficient estimators based on simple and weighted averages of range measures, offering flexibility and precision in capturing the dynamic relationships between assets. Through extensive simulation studies, we demonstrate the superiority of these estimators over traditional sample covariance methods under varying correlation structures and levels of heteroskedasticity.

The findings reveal that the simple average mode estimator is optimal for moderate correlations, while the MSE-weighted mode estimator performs better when correlations are high. These results are further strengthened by incorporating non-positive definite (NPD) transformations, which improve estimator performance, especially in higher-dimensional settings. By addressing the limitations of existing approaches, such as inflated variability under heteroskedasticity, we provide a more robust framework for estimating covariance matrices in practical scenarios.

Additionally, we apply both two-sample and one-sample tests to evaluate the statistical properties of the proposed estimators. The results confirm that range-based covariance matrices have higher power and lower type I error rates compared to sample-based matrices, making them more reliable for detecting inconsistencies in covariance structures across different scenarios. The proposed methodologies allow for the construction of dynamic time series of covariance matrices, which are essential for understanding correlation patterns over time and optimising portfolios to hedge risk effectively. Ultimately, the methods developed in this chapter set the stage for dynamic covariance modelling, addressing the random noise inherent in time series data and providing a foundation for more sophisticated analyses in upcoming chapters.

Chapter 3 focuses on extending the two-stage MCARR-return models to directly model the covariance matrix series using the Wishart distribution. We introduce two approaches—modelling the covariance matrix directly (Model A) and modelling variances and correlation matrices separately (Model B)—and evaluate their performance through simulation studies. The results consistently demonstrate that Model B outperformed Model A, providing more accurate estimates of variances and correlations, particularly under dynamic market conditions. We explore the impact

of the shape parameter ν in the Wishart distribution on variance and correlation estimates. Higher values of ν are associated with fewer spikes in variance and greater stability in correlations, emphasising its role in controlling the balance between variability and precision. Additionally, studies on the stability of correlation estimates (ρ_t) revealed that stability is related to α, $\alpha + \beta$ as well as ν. This relationship leads to the development of the LE metric, a novel quantitative measure for assessing ergodicity and stability in correlation dynamics.

Despite the good performance of MCARR models, they rely on the variability of the marginal Wishart distribution for correlation estimates. We demonstrate the potential of Wishart-based models for covariance matrix analysis while providing valuable insights into parameter selection, stability, and performance evaluation. These advancements pave the way for more robust and interpretable modelling frameworks in financial and other multi-dimensional applications.

Chapter 4 applies the methodologies developed in earlier chapters to real-world financial datasets, specifically focusing on U.S. stock market indices and the impact of the COVID-19 pandemic. These applications validate the robustness and adaptability of the proposed models, highlighting their effectiveness in capturing dynamic volatility and correlation structures in complex market environments. For U.S. stock market indices, the range-based estimators demonstrate superior in-sample fit and predictive accuracy, particularly in forecasting 1-step-ahead VaR. By comparing Model A and Model B, the findings reaffirm the advantages of modelling variances and correlations separately in real applications, while Model B only slightly underperformed Model A in the second stage. The efficiency of the range-based correlation and volatility measures is particularly evident in their ability to handle heteroskedasticity and variability, providing a robust framework for risk management and portfolio optimization.

In analysing the impact of the COVID-19 pandemic, we segment the data into three distinct periods to capture temporal changes in market behaviour. The results reveal significant variations in model parameters across these periods, reflecting the heightened volatility and correlation instability during the pandemic. Using the proposed LE metric, we quantify the loss of ergodicity of correlations in this period, highlighting the challenges posed by extreme market conditions. These findings underscore the importance of dynamic modelling approaches in adapting to rapidly changing financial landscapes. Additionally, we explore the integration of external features into the model, such as macroeconomic factors and investor sentiment. While these features can enhance interpretability, their inclusion alongside persistence terms led to overfitting in certain scenarios. This observation highlights the need for careful feature selection to balance model complexity and performance. We demonstrate the practical applicability of the proposed methodologies in addressing real-world challenges, offering insights into market dynamics and improving predictive accuracy. The ability to adapt to diverse scenarios, from stable markets to periods of crisis, highlights the versatility of the models developed in this dissertation. These applications provide a strong foundation for future research and practical implementations in financial risk management and beyond.

5.2. Future research

Although our research presents some promising results, we also realise some limitations of our studies. Firstly, our proposed range-based correlation matrix measure is constructed under the assumptions of zero drift and homoscedasticity. While we have examined the performance of our proposed correlation estimator in heteroscedastic scenarios in Section 2.5.1, our exploration of non-zero drift scenarios remains limited. Our preliminary simulation study suggests that the range-based correlation estimator may exhibit invariance to drift. However, the evidence is insufficient

to draw a definitive conclusion due to the narrow range of drift values tested in our preliminary simulation studies. Although the observed differences in mean squared error (MSE) between cases with varying drift levels are marginal, it is not sufficient to conclude that the estimator is invariant to drift, given the inherent randomness of our simulation experiments. For instance, real-world applications often involve scenarios with more pronounced drift, such as trending financial markets or datasets influenced by long-term structural changes.

Similarly, interactions between drift and volatility, as seen in periods of market shocks or high autocorrelation, might affect the estimator's properties in scenarios not included by our limited testing conditions. To robustly evaluate the proposed estimator's performance, further studies should investigate its properties under a broader set of conditions, including non-zero drift combined with varying degrees of heteroscedasticity, autocorrelation, and structural breaks. Such investigations will provide a more comprehensive understanding of the estimator's behaviour and its applicability in diverse real-world contexts.

Throughout the course of deriving the range-based correlation estimators, our approach remains primarily empirical. This reliance on simulation and numerical methods limits the ability to generalise or formalise the theoretical underpinnings of the proposed estimators. A more theoretical framework will significantly enhance the robustness and applicability of the range-based estimator. For the proposed range correlation measure, we demonstrate the properties of an efficient range-based correlation measure and construct numerical estimators to implement it. One promising direction for future research involves exploring the joint distribution of $(\boldsymbol{h}, \boldsymbol{l}, \boldsymbol{c})$ using possibly properties of order statistics to derive more efficient estimators, as functions of $(\boldsymbol{h}, \boldsymbol{l}, \boldsymbol{c})$. Apart from deriving the functional form of f_μ and f_m in our approach, we can obtain a maximum likelihood estimator based on the distribution of $(\boldsymbol{h}, \boldsymbol{l}, \boldsymbol{c})$, which contain some nice properties such as asymptotically consistency and efficiency under standard regularity conditions. This means that the estimator will not only converge to the true parameter value as the sample size increases but also achieve the lowest possible variance among unbiased estimators, maximising the precision of the range-based correlation measure. Such a theoretical formulation can provide deeper insights into the estimators' behaviours under various data-generating processes and market conditions.

Apart from incorporating these theoretical advancements, the integration of range-based measures into a broader array of statistical and econometric models, will advance the modelling capability and enhance the utility of range-based measures for real-world applications such as financial risk management, portfolio optimisation, and volatility forecasting. However, a key limitation in our two-stage MCARR-return models arises from the use of the Wishart distribution as the foundational matrix distribution. Specifically, the marginal distribution of the diagonal elements, representing variances, follows a scaled χ^2 distribution with a common degree of freedom, ν. This uniform degree of freedom across all marginal volatility distributions fails to reflect the inherent variability of distinct volatility processes for different assets. For instance, assets with higher volatility persistence or differing responses to market shocks are inadequately represented under this framework. To address this issue, a promising direction for future research will involve generalising the Wishart distribution to accommodate variable degrees of freedom in the marginal distributions. Such a generalised framework will provide greater flexibility, enabling more accurate modelling of diverse volatility behaviours while maintaining computational tractability. However, the matrix F distribution (Olkin and Rubin, 1964; Dawid, 1981) with extra degrees of freedom, though present flexibility, will not be adequate to achieve the goal of individual degrees of freedom. Instead, the idea of Choy et al. (2014) may be used to derive such a distribution, borrowing the idea for Student-t distributions.

Additionally, future research could focus on enhancing the computational efficiency of the proposed estimators, as the computational cost associated with MCMC within a Bayesian framework is substantial, especially for high-dimensional models. The slow convergence of MCMC presents a practical barrier to scaling the model for larger datasets or higher dimensions. To mitigate this, one promising approach is to project the high-dimensional parameter space onto a lower-dimensional subspace, followed by reconstructing the higher-dimensional structure from the reduced representation. Such a dimensionality reduction technique could dramatically improve computational efficiency without sacrificing model accuracy. Factor analysis represents another viable strategy for addressing the dimensionality issue. By identifying latent factors that explain most of the variability in the covariance or correlation matrix, we can reduce the number of parameters to be estimated, thereby simplifying the model and accelerating convergence.

However, factor analysis introduces additional complexity in selecting the optimal number of factors and ensuring the interpretability of the results. Another approach is to replace the traditional MCMC with Variational Bayesian methods, which are a family of techniques for approximating intractable integrals arising in Bayesian inference and machine learning. These computational issues originated from the curse of dimensionality, which has posted severe limitations to statistical models. On the other hand, advancements in neural networks, particularly deep learning and density networks, offer an exciting avenue for handling our high-dimensional MCARR models and improving computational efficiency. The tuning of hyperparameters and ensemble methods in neural networks resemble the running of MCMC to approximate the posterior distributions. Besides, convolution neural networks (CNN) (Venkatesan and Li, 2017) also offer a potential venue for modelling high-dimensional covariance matrices (Fang et al., 2021). By leveraging their capacity to handle high-dimensional data and learn complex dependencies, neural networks can provide faster, scalable alternatives for dynamic covariance modelling.

The implementation of neural networks could be challenging due to the positive definiteness of covariance and correlation matrices, which made the parameter space highly constrained. Lam et al. (2024) proposed modelling the Cholesky decomposition matrix to ensure positive definiteness. Geng and Zhang (2016) combined the adaptive neuro-fuzzy inference system (ANFIS) with the CARRX model and proposed a nonparametric CARRX model under the ANFIS framework.

Apart from handling the computational issues, there are potential extensions for the proposed two-stage MCARR-returns by proposing a dedicated distribution for the correlation matrices. Currently, the correlation matrices are modelled indirectly through the modelling of the covariance matrices, which limits the ability to model correlations dynamically and independently of variances. By introducing a distribution tailored for correlation matrices, the model can directly capture the dynamics of correlations, allowing for more precise and interpretable estimation and incorporating distinct sources of information beyond returns and variances. By refining this approach, the model can achieve greater precision in estimating confidence intervals and capturing dynamic behaviours. This is particularly valuable in contexts where correlations exhibit significant temporal variability or asymmetry, such as during periods of market stress or asset-specific shocks. Developing such a distribution, however, requires careful consideration of the constraints on correlation matrices, including their positive semi-definiteness and the bounded nature of correlation values between -1 and 1. Such developments will mark a significant step forward in the field, addressing the current limitations of our study and expanding its potential impact.

A related challenge in MCARR models lies in the scalar coefficients (α and β) used in the ARMA process for modelling correlation dynamics. While these coefficients provide a simple

mechanism for modelling correlation persistence, they are insufficient for capturing the heterogeneity of marginal dynamics within the correlation matrix. In an attempt to address this limitation, we explore a simple generalisation of α and β by having different coefficients for each pairwise correlation. However, this effort reveals the complexity of such an extension due to the strict constraints imposed on correlation matrices. A feasible generalisation must respect these constraints while allowing for greater flexibility in modelling diverse marginal behaviours. Future work could investigate advanced mathematical frameworks, such as manifold-based approaches or constrained optimisation techniques, to achieve this goal.

While our analyses demonstrate significant changes in the coefficients governing volatility dynamics during the pandemic period, we do not quantify the extent of these impacts. To address this, we attempt to incorporate daily new COVID-19 cases as a covariate in the model, hypothesising that this variable could capture the direct influence of the pandemic on market dynamics. Specifically, we explore the possibility of modelling the coefficients α and β as logistic functions of the daily new cases, aiming to reflect the potential nonlinear relationship between pandemic severity and volatility persistence. However, the parameters are not statistically significant, indicating that daily new cases alone may not sufficiently explain the changes in dynamics. Future research can explore other pandemic-related variables or a combination of factors, such as vaccination rates, mobility indices, or government policy interventions, to identify the key drivers of these changes.

Furthermore, the research period is manually divided into three distinct sub-periods based on our expert judgment and prior knowledge of the market dynamics during the timeframe under consideration. This approach, while pragmatic, relies heavily on subjective interpretation and may not fully capture the optimal segmentation points that reflect the underlying structural changes in the data. An alternative and more rigorous method for determining these splits could involve systematically comparing the likelihood of the entire model fitted to the data as a single period against the likelihood of models fitted to various candidate segmentations. By identifying the segmentation that maximises the overall likelihood, one could achieve a more precise and data-driven partitioning of the research period. Such a methodology will account for changes in the parameters of the model, such as shifts in volatility dynamics, correlations, or other structural properties, and ensure that the segmentation reflects statistically significant changes rather than relying solely on subjective criteria. For example, likelihood-based change-point models can identify breakpoints corresponding to major market events, such as the onset of the COVID-19 pandemic, the introduction of fiscal stimulus measures, or periods of market recovery. This type of analysis will not only enhance the robustness of the segmentation process but also provide deeper insights into the temporal dynamics of the data. It will enable a more precise investigation of how different periods exhibit varying market behaviours and how these changes impact model parameters. Future research can further incorporate methodologies such as structural break tests or Bayesian change-point detection to refine the process of periodisation and improve the overall explanatory power of the model.

In our proposed model D, we observe that adding covariate effects directly to the volatility process often leads to overfitting. This overfitting arises because the effects of covariates may already be partially or entirely captured by the lagged volatility terms, which reflect the persistence inherent in the data. As a result, the inclusion of additional covariates can redundantly amplify the model's complexity without contributing meaningful new information. On the other hand, Model F, which is specifically designed to incorporate covariate effects by replacing the persistence effects, demonstrate an ability to account for these impacts more explicitly. However, the performance of model F is found to be highly sensitive to the selection of efficient covariates.

If the chosen covariates fail to fully explain the variability in the data, the model's accuracy and stability can deteriorate significantly. To strike a balance between the competing formulations of models D and F, a promising approach involves leveraging covariance regression.

Specifically, we first fit a covariance regression model defined in (1.74) to get coefficient matrices $\boldsymbol{B}$ and $\boldsymbol{C}$. Now define

$$\boldsymbol{O}_t = \nu\boldsymbol{\Lambda}_t - (\boldsymbol{C} + \boldsymbol{B}\boldsymbol{x}_t\boldsymbol{x}_t^\top\boldsymbol{B}^\top) \tag{5.1}$$

where $\boldsymbol{O}_t$ can be treated as the residuals of regression fitting. Marginally, we have

$$\sigma_{i,t}^2 = o_{ii,t}/\nu + c_{ii} + \sum_{l=1}^{\mathfrak{l}}\sum_{m=1}^{\mathfrak{l}} b'_{ilm}x_{l,t}x_{m,t}/\nu, \tag{5.2}$$

where $o_{ii,t}$ is the i-th diagonal of $\boldsymbol{O}_t$, b'_{ilm} is the coefficient of $x_{l,t}x_{m,t}$ term in i-th diagonal of matrix $\boldsymbol{B}\boldsymbol{x}_t\boldsymbol{x}_t^\top\boldsymbol{B}^\top$. Plugging (5.2) into (3.19), we have

$$\sigma_{i,t}^2 = a_{0i} + a_i v_{ii,t-1} + b_i(o_{ii,t}/\nu + c_{ii} + \sum_{l=1}^{\mathfrak{l}}\sum_{m=1}^{\mathfrak{l}} b'_{ilm}x_{l,t}x_{m,t}/\nu). \tag{5.3}$$

By comparing Equation 5.3 to the volatility equation in model F, it becomes evident that $o_{ii,t}/\nu$ captures the component of volatility that cannot be explained by $\boldsymbol{x}_t$. Building upon this innovation, we propose the following model:

$$\begin{aligned}
\text{1-stage: } \boldsymbol{V}_t &\sim \mathrm{W}_d(\boldsymbol{\Lambda}_t, \nu), && (5.4)\\
\boldsymbol{\Lambda}_t &= \mathrm{diag}(\sigma_{1,t},\ldots,\sigma_{d,t})\boldsymbol{Q}_t\mathrm{diag}(\sigma_{1,t},\ldots,\sigma_{d,t}),\\
\boldsymbol{O}_t &= \nu\boldsymbol{\Lambda}_t - \boldsymbol{C} + \boldsymbol{B}\boldsymbol{x}_t\boldsymbol{x}_t^\top\boldsymbol{B}^\top && (5.5)\\
\sigma_{i,t}^2 &= a_{i0} + \sum_{k=1}^{p} a_{ip}v_{ii,t-p} + \sum_{k=1}^{q}\boldsymbol{c}_{\boldsymbol{k}}^\top\cdot(\boldsymbol{x}_{t-k}\odot\boldsymbol{x}_{t-k}) + \sum_{k=1}^{q} d_k|o_{ii,t-k}|, && (5.6)\\
\boldsymbol{Q}_t &= \boldsymbol{\Psi} + \alpha(\widehat{\boldsymbol{\mathcal{Q}}}_{t-1} - \boldsymbol{\Psi}) + \beta(\boldsymbol{Q}_{t-1} - \boldsymbol{\Psi}), && (5.7)\\
\text{2-stage: } \boldsymbol{R}_t &= \boldsymbol{\mu}_t + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathrm{ST}_{\boldsymbol{R},d}(\boldsymbol{0}, \boldsymbol{\Omega}\nu\boldsymbol{\Lambda}_t\boldsymbol{\Omega}, \upsilon), && (5.8)\\
\boldsymbol{\mu}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{y}_{t-1}. && (5.9)
\end{aligned}$$

Comparing (5.6) with (5.3), we drop interacting effect term $x_{l,t}x_{m,t}, l \neq m$. Furthermore, we establish the coefficients as independent of $b_{il}b_{im}$ to introduce greater flexibility in modelling the persistence effects. This formulation allows each covariate to exhibit its own distinct persistent influence on the volatility process. By incorporating a regression framework within the covariance structure, we can separate the effects of covariates from the volatility dynamics. Specifically, the absolute residual term from a covariance regression can be introduced into the model to capture the unexplained volatility after accounting for the covariate effects, while maintaining $\sigma_{i,t}^2$ positive. This method not only allows for a more parsimonious representation of volatility dynamics but also ensures that the covariate effects are distinctly modelled, reducing the risk of overfitting. Moreover, the covariance regression approach provides a clearer interpretation of the contributions of covariates and residual volatility, offering insights into the underlying mechanisms driving market behaviour during periods of heightened uncertainty, such as the pandemic.

Bibliography

Ahmad, M. R. and von Rosen, D. (2015). Tests for high-dimensional covariance matrices using the theory of U-statistics. *Journal of Statistical Computation and Simulation*, 85(13):2619–2631.

Alomari, M. W. (2019). Numerous approximations of riemann-stieltjes double integrals.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2010). Parametric and nonparametric volatility measurement. In *Handbook of financial econometrics: Tools and techniques*, pages 67–137. Elsevier.

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Bakdash, J. Z. and Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in psychology*, 8:252904.

Baur, P. W. (2020). An examination of the impact of covid-19 on the financial markets and how this directs investment into the market for fine art. *Journal of Economic and Financial Sciences*, 13(1):1–22.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.

Bergomi, L. (2015). *Stochastic volatility modeling*. CRC press.

Betancourt, M. J. and Girolami, M. (2013). Hamiltonian monte carlo for hierarchical models.

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

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, 72(3):498–505.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.

Boyle, J. P. and Dykstra, R. L. (1986). A method for finding projections onto the intersection of convex sets in Hilbert spaces. In *Advances in Order Restricted Statistical Inference: Proceedings of the Symposium on Order Restricted Statistical Inference held in Iowa City, Iowa, September 11–13, 1985*, pages 28–47. Springer.

Brandt, M. W. and Diebold, F. X. (2006). A no-arbitrage approach to range-based estimation of return covariances and correlations. *The Journal of Business*, 79(1):61–74.

Brandt, M. W. and Jones, C. S. (2006). Volatility forecasting with range-based EGARCH models. *Journal of Business & Economic Statistics*, 24(4):470–486.

Cai, T., Liu, W., and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108(501):265–277.

Campbell, N., Lopuhaä, H. P., and Rousseeuw, P. J. (1998). On the calculation of a robust S-estimator of a covariance matrix. *Statistics in Medicine*, 17(23):2685–2695.

Cepoi, C.-O. (2020). Asymmetric dependence between stock market returns and news during covid-19 financial turmoil. *Finance research letters*, 36:101658.

Chan, J. S. K., Ng, K. H., and Ragell, R. (2019). Bayesian return forecasts using realised range and asymmetric CARR model with various distribution assumptions. *International Review of Economics & Finance*, 61:188–212.

Chen, S. X., Zhang, L.-X., and Zhong, P.-S. (2010a). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

Chen, Y., Wiesel, A., Eldar, Y. C., and Hero, A. O. (2010b). Shrinkage algorithms for mmse covariance estimation. *IEEE transactions on signal processing*, 58(10):5016–5029.

Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking*, 37(3):561–582.

Chou, R. Y. and Liu, N. (2008). The economic value of volatility timing using a range-based volatility model. *Available at SSRN 1330908*.

Chou, R. Y., Wu, C.-C., and Liu, N. (2009). Forecasting time-varying covariance with a range-based dynamic conditional correlation model. *Review of Quantitative Finance and Accounting*, 33:327–345.

Choy, S. B., Chen, C. W., and Lin, E. M. (2014). Bivariate asymmetric GARCH models with heavy tails and dynamic conditional correlations. *Quantitative Finance*, 14(7):1297–1313.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association*, 82(400):1079–1091.

Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika*, 68(1):265–274.

Drouet Mari, D. and Kotz, S. (2001). *Correlation and dependence*. World Scientific.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics letters B*, 195(2):216–222.

Elyasiani, E. and Mansur, I. (1998). Sensitivity of the bank stock returns distribution to changes in the level and volatility of interest rate: A GARCH-M model. *Journal of banking & finance*, 22(5):535–563.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business Economic Statistics*, 20(3):339–350.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, 5(1):1–50.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric theory*, 11(1):122–150.

Engle, R. F., Lilien, D. M., and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica: journal of the Econometric Society*, pages 391–407.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(4):603–680.

Fang, Y., Yu, P. L., and Tang, Y. (2021). CNN-based realized covariance matrix forecasting. *arXiv preprint arXiv:2107.10602*.

Feller, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *The Annals of Mathematical Statistics*, 22(3):427–432.

Fernandes, M., de Sá Mota, B., and Rocha, G. (2005). A multivariate conditional autoregressive range model. *Economics Letters*, 86(3):435–440.

Fisher, T. J. (2012). On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size. *Journal of statistical planning and inference*, 142(1):312–326.

Francq, C. and Zakoian, J.-M. (2019). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.

Freedman, D. (2012). *Brownian motion and diffusion*. Springer Science & Business Media.

Garman, M. B. and Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 53(1):67–78.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741.

Geng, L. and Zhang, Z. (2016). Application of anfis-based carrx model to stock volatility forecasting. *International Journal of Simulation Systems, Science & Technology*, 17(22):7.

Gerlach, R. and Wang, C. (2016). Forecasting risk via realized GARCH, incorporating the realized range. *Quantitative Finance*, 16(4):501–511.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801.

Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). *The Annals of Mathematical Statistics*, 34(1):152–177.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of Applied Econometrics*, 20(7):873–889.

Hansen, P. R., Lunde, A., and Voev, V. (2014). Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. *Journal of Applied Econometrics*, 29(5):774–799.

Higham, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343.

Hoeffding, W. (1994). A non-parametric test of independence. *The Collected Works of Wassily Hoeffding*, pages 214–226.

Hoff, P. D. and Niu, X. (2012). A covariance regression model. *Statistica Sinica*, pages 729–753.

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623.

Hotelling, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society. Series B (Methodological)*, 15(2):193–232.

Hsieh, D. A. (1991). Chaos and nonlinear dynamics: application to financial markets. *The journal of finance*, 46(5):1839–1877.

Jorion, P. (1996). Risk^2: Measuring the risk in value at risk. *Financial Analysts Journal*, 52(6):47–56.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series

have a unit root? *Journal of econometrics*, 54(1-3):159–178.

Lam, F., Chan, J., and Choy, B. (2024). Temporal multivariate density networks for portfolio optimization. *Temporal Multivariate Density Networks for Portfolio Optimization (October 17, 2024)*.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30(4):1081–1102.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411.

Li, H. and Hong, Y. (2011). Financial volatility forecasting with range-based autoregressive volatility model. *Finance Research Letters*, 8(2):69–76.

Lowther, G. (2023). The minimum and maximum of brownian motion.

Madan, D. B. and Seneta, E. (1990). The variance gamma (V.G.) model for share market returns. *The Journal of Business*, 63(4):511–524.

Markowitz, H. (1952). Modern portfolio theory. *Journal of Finance*, 7(11):77–91.

Meilijson, I. (2011). The Garman-Klass volatility estimator revisited. *Revstat*, 9(3):199–212.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics*, 8(4):323–361.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.

Mikosch, T. and Stărică, C. (2000). Limit theory for the sample autocorrelations and extremes of a garch (1, 1) process. *The Annals of Statistics*, 28(5):1427–1451.

Muirhead, R. J. and Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika*, 67(1):31–43.

Nagao, H. (1973). On some test criteria for covariance matrix. *The Annals of Statistics*, 1(4):700–709.

Neal, R. M. (1994). An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics*, 111(1):194–203.

Neal, R. M. (2011). Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2:113–162.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

Nitithumbundit, T. and Chan, J. S. (2022). Covid-19 impact on cryptocurrencies market using multivariate time series models. *The Quarterly Review of Economics and Finance*, 86:365–375.

Olkin, I. and Rubin, H. (1964). Multivariate beta distributions and independence properties of the wishart distribution. *The Annals of Mathematical Statistics*, pages 261–269.

Page Jr, T. J. (1984). Multivariate statistics: A vector space approach. *Journal of Marketing Research*, 21(2):236–236.

Paparoditis, E. and Politis, D. N. (2018). The asymptotic size and power of the augmented dickey–fuller test for a unit root. *Econometric Reviews*, 37(9):955–973.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 53(1):61–65.

Rogers, L. C. G. and Satchell, S. E. (1991). Estimating variance from high, low and closing prices. *Annals of Applied Probability*, 1(4):504–512.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639.

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society*, 35(2):251–272.

Srivastava, M. S., Kollo, T., and von Rosen, D. (2011). Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis*, 102(6):1090–1103.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794.

Tan, S. K., Ng, K. H., and Chan, J. S.-K. (2022). Predicting returns, volatilities and correlations of stock indices using multivariate conditional autoregressive range and return models. *Mathematics*, 11(1):13.

Taylor, S. J. (1986). *Modelling Financial Time Series*. John Wiley and Sons, Chichester.

Team, S. D. et al. (2016). Rstan: the R interface to Stan. *R package version*, 2(1):522.

Van Jaarsveldt, C., Peters, G. W., Ames, M., and Chantler, M. (2024). Long/short equity risk premia parity portfolios via implicit factors in regularized covariance regression. *IEEE Access*, 12:119405–119432.

Venkatesan, R. and Li, B. (2017). *Convolutional neural networks in visual computing: a concise guide*. CRC Press.

Wasserman, L. A. (1989). A robust Bayesian interpretation of likelihood regions. *The Annals of Statistics*, pages 1387–1393.

Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1/2):32–52.

Yang, D. and Zhang, Q. (2000). Drift independent volatility estimation based on high, low, open, and close prices. *Journal of Business*, 73(3):477–492.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.

Zhang, D., Hu, M., and Ji, Q. (2020). Financial markets under the global pandemic of covid-19. *Finance research letters*, 36:101528.

Zhang, X., Shi, H., Zhou, N., Tan, F., and Guo, X. (2024). Quantile generalized measures of correlation. *Statistics and Computing*, 34(2):92.

Zheng, S., Shi, N.-Z., and Zhang, Z. (2012). Generalized measures of correlation for asymmetry, nonlinearity, and beyond. *Journal of the American Statistical Association*, 107(499):1239–1252.