universität wien

DISSERTATION | DOCTORAL THESIS

Titel | Title

Three Essays in Financial Econometrics and Portfolio Optimization

verfasst von | submitted by

Yuan Chen MSc UZH ETH

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 794 370 145
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Finance
Betreut von | Supervisor:	Univ.-Prof. Dr. Nikolaus Hautsch

Acknowledgements

First and foremost, I would like to thank Nikolaus Hautsch for his guidance, support, and encouragement throughout my PhD. Without his supervision and collaboration, this thesis would not have been possible.

I would also like to thank Melanie Schienle and Jérémy Leymarie for their support and insightful discussions, which led to our first paper. I learned a great deal from this project, and it significantly shaped my research interests. A special thanks goes to Immanuel Bomze for the many opportunities to present at conferences and for the numerous discussions that guided me into the field of optimization. Without his support, as well as that of Bo Peng, the second chapter of this thesis would not have existed.

Moreover, I would like to express my appreciation to Christian T. Brownlees and Prof. Richard Luger for reviewing my thesis.

Many thanks to everyone on the 5th floor at Kolingasse 14-16 and to my colleagues at VGSF for being such great company, whether through coffee chats, academic discussions, or countless dinners. In particular, I would like to thank Janka Möller, Luca Gonzato, Francesca Primavera, Dominik von Gehlen, Aron Bodisz, and Elizabeth Rinde.

I am also grateful to my friends with whom I had many online conversations throughout my PhD journey, making this experience much more enjoyable: Yanchen He, Zhenwei Xu, Kiwan Wong, and Hanlin Yang.

A special thanks to those who supported me during my job market journey, including Michael Wolf, Tobin Hanspal, Wei Li, and Ilya Archakov, as well as many others I had the pleasure of meeting over the past year.

I am deeply thankful to my parents for their unwavering support throughout my studies, especially my mother, Yongzhong Deng. Without her endless encouragement and unconditional support, I would not have achieved what I have today.

Finally, my deepest gratitude goes to my fiancé, Jona Marie Hassenbach, for her unconditional love and support. I could not have undertaken this journey without you.

Abstract

This thesis consists of three essays in financial econometrics and portfolio optimization. The first paper provides statistical inference for marginal expected shortfall (MES) and delta conditional value-at-risk (ΔCoVaR) measures, which are semiparametrically estimated by a two-step procedure in a multivariate GARCH-type framework. We establish the asymptotic properties of the corresponding estimators and illustrate how the estimation uncertainty can be decomposed into dynamic univariate marginal and time-varying dependence components. Moreover, we propose tests for differences in systemic risk in order to construct confidence sets for companies' ranks in systemic risk rankings. In an empirical application based on 50 large US financial institutions, the framework provides novel evidence on the informativeness of such rankings and highlights the importance of accounting for time-varying return dependence. The second paper proposes a portfolio optimization model that reconciles Keynes's advocacy for concentrated investments with Markowitz's emphasis on diversification. By incorporating cardinality constraints into the mean-variance framework, investors can focus on a small set of assets while still using the sample covariance matrix in high-dimensional settings with limited data, balancing diversification needs and mitigating estimation errors. The third paper investigates the role of correlations in portfolio selection. It proposes a transformation that isolates the directional component of correlations and shows that both fully ignoring and fully relying on correlation are suboptimal. Empirically, the directional component captures the most relevant information for diversification and improves portfolio performance, providing a framework for constructing more robust portfolios.

Kurzfassung

Diese Dissertation besteht aus drei Essays in Financial Econometrics und Portfoliooptimierung. Der erste Beitrag liefert statistische Inferenz für Marginal Expected Shortfall (MES) und Delta Conditional Value-at-Risk (ΔCoVaR), die mittels eines zweistufigen semiparametrischen Verfahrens in einem multivariaten GARCH-Typ-Rahmen geschätzt werden. Wir etablieren die asymptotischen Eigenschaften der entsprechenden Schätzer und zeigen, wie die Schätzunsicherheit in dynamische univariate Randkomponenten und zeitvariierende Abhängigkeitskomponenten zerlegt werden kann. Darüber hinaus schlagen wir Tests für Unterschiede im systemischen Risiko vor, um Konfidenzmengen für Ranglisten von Unternehmen im systemischen Risiko zu konstruieren. In einer empirischen Anwendung auf 50 große US-Finanzinstitute liefert der Ansatz neue Evidenz zur Aussagekraft solcher Ranglisten und unterstreicht die Bedeutung der Berücksichtigung zeitvariierender Renditeabhängigkeiten. Der zweite Beitrag schlägt ein Portfoliooptimierungsmodell vor, das Keynes' Befürwortung konzentrierter Investitionen mit Markowitz' Betonung der Diversifikation in Einklang bringt. Durch die Einbeziehung von Kardinalitätsrestriktionen in den Mean-Variance-Rahmen können sich Investoren auf eine kleine Anzahl von Assets konzentrieren und dennoch die Stichprobenkovarianzmatrix in hochdimensionalen Umgebungen mit begrenzten Daten verwenden, wodurch Diversifikationsbedürfnisse ausbalanciert und Schätzfehler reduziert werden. Der dritte Beitrag untersucht erneut die Rolle von Korrelationen in der Portfolioselektion. Er schlägt eine Transformation vor, die die Richtungskomponente von Korrelationen isoliert, und zeigt, dass sowohl das vollständige Ignorieren als auch das vollständige Berücksichtigen von Korrelation suboptimal sind. Empirisch erfasst die Richtungskomponente die relevanteste Information für Diversifikation und verbessert die Portfolioleistung, wodurch ein Rahmen zur Konstruktion robusterer Portfolios bereitgestellt wird.

Contents

Introduction

This thesis consists of three essays in financial econometrics and portfolio optimization. Each essay addresses a distinct research question in finance, ranging from systemic risk inference at the market level to portfolio construction at the investor level. Despite their different focuses, the three essays share a common objective: using rigorous mathematical and statistical methods to improve financial decision-making. In what follows, I provide a brief overview of each essay and the motivation behind it.

The first chapter focuses on systemic risk. The recent financial crisis has fostered extensive research in this area, and understanding how financial institutions are interconnected has become critically important. A large body of literature has focused on the measurement of systemic risk, and in this chapter we study two of the most widely used measures, the Marginal Expected Shortfall (MES) (Acharya et al., 2012, 2017) and the Delta Conditional Value-at-Risk (ΔCoVaR) (Adrian and Brunnermeier, 2011, 2016). These measures have become important tools for regulators and policymakers seeking to monitor systemic vulnerabilities and design policies aimed at preventing future financial crises.

Despite their widespread use in academic research and policy discussions, an important gap remains in the literature. In practice, systemic risk measures are often used to rank financial institutions according to their systemic importance, yet relatively little is known about the asymptotic properties of the corresponding estimators and the associated statistical inference for these rankings. From a financial econometrics perspective, this raises the need for rigorous inference procedures that allow researchers and policymakers to assess whether differences in systemic risk across institutions are statistically meaningful.

To address this issue, the first chapter develops a semi-parametric framework for estimating MES and ΔCoVaR in dynamic financial systems with time-varying volatility and dependence structures. The chapter derives asymptotic theory for these measures under multivariate GARCH specifications and proposes statistical inference procedures for systemic risk rankings, including tests for equality in systemic risk across institutions and confidence sets for rankings. An empirical application to large US banks highlights the importance of modeling dynamic correlations and demonstrates that the informativeness of systemic risk rankings varies substantially over time.

The second chapter shifts the focus from systemic risk measurement to portfolio optimization for individual investors. One of the foundations of modern finance is the seminal work of Markowitz (1952a), which introduced the *mean-variance* framework for portfolio selection. Since then, diversification has become a central principle of asset allocation and portfolio management. Under the classical view, investors should hold a large set of assets in order to diversify idiosyncratic risk.

However, an important practical question remains: how many assets are actually

needed to achieve sufficient diversification? While larger portfolios may provide additional diversification benefits, they also increase estimation risk and the complexity of portfolio management. In particular, estimating large covariance matrices becomes increasingly difficult in high-dimensional settings. Since the seminal contribution of Statman (1987), the question of how many stocks are sufficient for a well-diversified portfolio has remained an important topic in the portfolio optimization literature.

This chapter studies the trade-off between diversification benefits and estimation uncertainty through a cardinality-constrained portfolio optimization framework. The resulting optimization problem is NP-hard and therefore computationally challenging in large dimensions. To address this issue, the chapter develops tractable relaxations together with an efficient optimization algorithm. Combined with formal statistical testing procedures, the proposed framework provides investors with a practical tool for constructing customized and well-diversified portfolios using only a relatively small subset of assets.

Importantly, the proposed cardinality constraints do not distort the structural composition of portfolios across industries. Instead, the optimization procedure primarily acts as a dimensionality reduction mechanism, preserving the broader economic structure of the investment universe while reducing portfolio complexity. This result is particularly relevant for individual investors, who often prefer manageable portfolios without sacrificing the benefits of diversification.

The third chapter returns to portfolio management and studies the challenge of covariance matrix estimation from a behavioral finance perspective rather than a purely statistical one. Accurately estimating the covariance matrix of asset returns is one of the central problems in portfolio optimization, as it determines the dependence structure and diversification benefits across assets. This challenge becomes particularly important in high-dimensional settings, where the number of assets is large relative to the available sample size. A substantial literature, including the pioneering work of Ledoit and Wolf (2004), addresses this issue through shrinkage methods, while other studies (Rothman et al., 2009; Bickel and Levina, 2008; Karoui, 2008) focus on regularization techniques. At the same time, empirical studies often advocate simplifying portfolio decisions through heuristic approaches, such as the equal-weighted ($1/N$) portfolio (DeMiguel et al., 2009b) or strategies that partially or fully ignore correlation information (Kirby and Ostdiek, 2012).

In contrast, this chapter is motivated by the behavioral finance literature on "correlation neglect," which documents that investors often ignore correlation or pay limited attention to correlation information. More recently, Ungeheuer and Weber (2021) show that some investors instead rely primarily on simpler directional patterns of comovement. We establish a formal connection between partial correlation neglect and robust rank-based dependence measures, which, to the best of our knowledge, has not previously been explored in the literature. Exploiting the relationship between Pearson correlation and Kendall's τ, the chapter decomposes the sample correlation matrix into a directional component and a magnitude component. The analysis shows that the directional component contains stable and economically meaningful information about dependence structures, while the

magnitude component is substantially more sensitive to estimation error.

Based on this insight, the chapter develops a portfolio construction approach that relies primarily on directional dependence information, providing a new perspective on covariance matrix estimation, shrinkage, and robust portfolio optimization. More broadly, the chapter connects shrinkage methods with investor behavior by showing that correlation neglect can be interpreted as a practical response to estimation uncertainty and model complexity. In this sense, simplifying heuristics may sometimes improve portfolio decisions by reducing sensitivity to noisy estimates and unstable correlation structures. The chapter therefore offers both a behavioral and statistical interpretation of robust portfolio construction.

Together, these three essays contribute to the broader literature by providing rigorous statistical and optimization tools for risk measurement and portfolio construction, with implications for both researchers and practitioners.

Bibliographic information on the papers contained in this thesis

This thesis is a collection of the following papers. We made minor modifications to the respective paper in order for them to better fit the format of this thesis.

Chapter 1:
Multivariate Inference for Dynamic Systemic Risk Measures
Yuan Chen, Nikolaus Hautsch, Jérémy Leymarie, Melanie Schienle
Not published elsewhere yet.
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.
The paper has been revised and resubmitted to the Journal of Econometrics.

Chapter 2:
Cardinality-Constrained Optimization for Large-Scale Portfolio
Immanuel M. Bomze, Yuan Chen, Bo Peng, Nikolaus Hautsch
Not published elsewhere yet.
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.

Chapter 3:
Is Correlation Neglect Bad for Portfolio Diversification?
Yuan Chen
Not published elsewhere yet.
The paper was my job market paper.

1 Multivariate Inference for Dynamic Systemic Risk Measures

Abstract. This paper provides statistical inference for marginal expected shortfall (MES) and delta conditional value-at-risk (ΔCoVaR) measures, which are semi-parametrically estimated by a two-step procedure in a multivariate GARCH-type framework. We establish the asymptotic properties of corresponding estimators and illustrate how the estimation uncertainty can be decomposed into dynamic univariate marginal and time-varying dependence components. Our methodology reveals good finite sample performance for estimation and prediction of risk. Moreover, we propose tests for differences in systemic risk in order to construct confidence sets for companies' ranks in systemic risk rankings. In an empirical application based on 50 large US financial institutions, our framework provides novel evidence on the informativeness of such rankings. Moreover, our findings highlight the importance of accounting for time-varying return dependence in systemic risk estimators.

Keywords: Multivariate dynamic systemic risk; multivariate GARCH dynamics; two-step multivariate statistical inference; credible systemic risk rankings.

1.1 Introduction

The recent financial crisis has fostered extensive research on systemic risk, encompassing its definition, measurement, and regulation. A key area of focus has been the identification of financial institutions that contribute the most to overall financial system risk - referred to as Systemically Important Financial Institutions (SIFIs). The Financial Stability Board (2011) defines SIFIs as "*financial institutions whose distress or disorderly failure, because of their size, complexity and systemic interconnectedness, would cause significant disruption to the wider financial system and economic activity*". Given the potential threat SIFIs pose to the system, regulators and policymakers worldwide have called for stricter supervision, additional capital requirements, and liquidity buffers. These mechanisms are underpinned by systemic risk measures (SRMs), which are critical tools for identifying vulnerabilities within the financial system (see De Bandt and Hartmann, 2002; Benoit et al., 2017, for a survey).

This paper provides estimation theory and multivariate inference of SRMs in a dynamic financial system of stock returns[1], where not only each univariate component is dynamic and conditionally heteroscedastic but also the dependence between constituents might be conditionally time-varying. We analytically derive the effect of such time variation on the

[1]The provided theory is agnostic of the specific asset class and also holds for a system of option prices, bond prices or CDS spreads.

uncertainty of SRM estimators in their asymptotic expansion and highlight its relevance in simulations and empirical applications. Building on these results, we develop statistical inference for tests of equal systemic risk across banks. This yields a comprehensive framework to statistically assess the informativeness of systemic risk rankings commonly used in financial practice.

We focus on the two most prominent SRMs, the Marginal Expected Shortfall (MES) (Acharya et al., 2012, 2017) and the Delta Conditional Value-at-Risk (ΔCoVaR) (Adrian and Brunnermeier, 2011, 2016). They also serve as key elements and building blocks for further SRMs like SRISK (Acharya et al., 2012; Brownlees and Engle, 2017) and the Systemic Expected Shortfall (SES) (Acharya et al., 2017). Since SRISK and SES are functions of MES and lagged observable firm-specific characteristics, our inference developed for the MES estimator directly extend to them. For details we refer to Appendix A.

Let $\boldsymbol{y}_t = (y_{m,t}, y_{i,t})'$ be the vector of two time series at time t, where $y_{m,t}$ is the market return and $y_{i,t}$ the stock return of the i-th financial institution. Let $\mathcal{F}_{t-1}$ be the information set available at time $t-1$, with $(\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, ...) \subseteq \mathcal{F}_{t-1}$. Following Acharya et al. (2012), we define the MES of the institution as its short-run expected equity loss conditional on the market taking a loss more severe than its Value-at-Risk (VaR). The α-level MES for institution i at time t given $\mathcal{F}_{t-1}$ is defined as

$$MES_{i,t}(\alpha) = \mathbb{E}(y_{i,t} | y_{m,t} \leq VaR_{m,t}(\alpha); \mathcal{F}_{t-1}), \tag{1.1}$$

where $VaR_{m,t}(\alpha)$ is the α-level VaR of $y_{m,t}$ with $VaR_{m,t}(\alpha) = F^{-1}_{y_m}(\alpha; \mathcal{F}_{t-1})$, F_{y_m} denoting the cumulative distribution function (cdf) of y_m, and $\alpha \in [0,1]$. Thus for α small such as 10% or 5%, MES measures how the financial institution contributes to the overall risk of the financial system.

The Delta Conditional Value-at-Risk (ΔCoVaR) of Adrian and Brunnermeier (2016) is based on the so-called CoVaR, defined as the VaR of the firm return $y_{i,t}$ conditionally on the financial system being under stress. The latter event is associated with the market return $y_{m,t}$ that we quantify as $y_{m,t} = VaR_{m,t}(\alpha)$ with α small such as 10% or 5%.[2] The β-level CoVaR for firm i at time t is the quantity $CoVaR_{i,t}(\beta, \alpha)$ and is defined as[3]

$$\Pr\left(y_{i,t} \leq CoVaR_{i,t}(\beta, \alpha) | y_{m,t} = VaR_{m,t}(\alpha); \mathcal{F}_{t-1}\right) = \beta\,. \tag{1.2}$$

Then ΔCoVaR marks the difference between the conditional VaR (CoVaR) of an institution conditional on the financial system being in distress and the CoVaR conditional on the financial system being in its median state ($y_{m,t} = VaR_{m,t}(0.5)$), i.e.,

$$\Delta\text{CoVaR}_{i,t}(\beta, \alpha) = CoVaR_{i,t}(\beta, \alpha) - CoVaR_{i,t}(\beta, 0.5). \tag{1.3}$$

[2] A different approach consists of defining the financial stress as a situation in which $y_{m,t} \leq VaR_{m,t}(\alpha)$, as in Girardi and Ergün (2013), Banulescu-Radu et al. (2021) or as in Francq and Zakoïan (2025). Here, we follow the original definition introduced by Adrian and Brunnermeier (2016)

[3] If we denote by $F_{y_{i,t}|y_{m,t}=VaR_{m,t}(\alpha)}(.; \mathcal{F}_{t-1})$ the cdf of the distribution of $y_{i,t}$ given $y_{m,t} = VaR_{m,t}(\alpha)$ and $\mathcal{F}_{t-1}$, then $CoVaR_{i,t}(\beta, \alpha)$ is the β-quantile of $y_{i,t}|y_{m,t} = VaR_{m,t}$, i.e., $CoVaR_{i,t}(\beta, \alpha) = F^{-1}_{y_{i,t}|y_{m,t}=VaR_{m,t}(\alpha)}(\beta; \mathcal{F}_{t-1})$.

These measures are designed to project the systemic risk contribution of a given financial institution into a single figure and can be used to rank financial institutions according to their systemic importance.

We propose a semi-parametric approach for dynamic MES in (1.1) and dynamic ΔCoVaR in (1.3) that employs a general multivariate GARCH (MGARCH) specification for the dynamics of the underlying $\boldsymbol{y}_t$ and leaves the innovation term parametrically unspecified.[4] In this setup, we show that the MES and ΔCoVaR have closed form solutions that can be decomposed into a part originating from the conditional variance-covariance matrix of $\boldsymbol{y}_t$, comprising firm's return volatility and the correlation between the firm's and market's returns, and a factor based on the respective univariate SRM for the market innovation. Accordingly, estimation of the SRMs can be separated in two steps. In the first step, we estimate the model parameters in the conditional variance-covariance specification using quasi-maximum likelihood estimation (QMLE) within the general class of multivariate GARCH (MGARCH) models (see Bauwens et al., 2006; Silvennoinen and Teräsvirta, 2009, for a survey). In the second step, the truncated mean and quantiles are estimated using their empirical counterparts from the MGARCH residuals. This approach extends the filtered historical simulation (FHS) method used in univariate GARCH settings for dynamic semi-parametric ES and VaR (see e.g., Gao and Song, 2008; Francq and Zakoïan, 2015) to the multivariate case.

While much of the literature has focused on the limiting behavior of dynamic univariate risk measures, such as VaR and Expected Shortfall (ES) (see e.g., Chen and Tang, 2005; Scaillet, 2005; Chen, 2008; Gao and Song, 2008; Francq and Zakoïan, 2015; Patton et al., 2019), only a few recent papers address estimation theory for dynamic SRMs. Nolde and Zhang (2020) develop a semi-parametric framework based on extreme value theory (EVT) to compute dynamic CoVaR. Hoga (2023) proposes a dynamic MES estimator combining GARCH volatilities with EVT-based joint tail modeling. Francq and Zakoïan (2025) develop a dynamic CoVaR (and ΔCoVaR) estimator based on univariate location-scale processes, using equation-by-equation quasi-maximum likelihood.

Our first main contribution to this literature is the asymptotic theory for SRM under general dynamic MGARCH specifications. In particular, we provide a theoretical and operational bridge between the estimation theory of MGARCH models and the semi-parametric estimation of SRMs such as MES and ΔCoVaR. In particular, we show that the asymptotic variances of the FHS second-step estimators decompose into two components: (*i*) the estimation error of empirical quantiles or truncated means, free of nuisance parameters, and (*ii*) the uncertainty propagated from the first-step QMLE of the MGARCH model. This decomposition clarifies how uncertainty from the two estimation stages interacts in a multivariate setting. As in Francq and Zakoïan (2025), MES and ΔCoVaR are both derived within a common setup, sharing the same first-step estimation, with differences only in the second step. The second-step estimation, based on FHS, makes no distributional assumptions on the innovations, which has been shown to reduce model misspecification risk (see e.g., Kuester et al., 2006, showing this for VaR). Our

[4] For completeness, we also cover the estimation theory for the dynamic CoVaR in (1.2).

parametric MGARCH specification, however, allows for dynamic correlation structures that we empirically confirm to be crucial when modeling financial stock returns. At the same time, dynamics in correlations heavily influence MES and ΔCoVaR estimates. We show that ignoring the time variation in correlations, such as in the CCC-GARCH of Bollerslev (1990) and the ECCC-GARCH of Jeantheau (1998), reduces the precision of SRM estimates. This calls for MGARCH models capturing dynamic correlations, as the DCC-GARCH of Engle (2002), the corrected DCC of Aielli (2013), or the BEKK model of Engle and Kroner (1995).

Our second main contribution is the derivation of statistical inference for SRM-based systemic risk rankings. In particular, we construct tests for the equality of systemic risk across institutions by showing that both $MES_{i,t}$ and $\Delta CoVaR_{i,t}$ can be written as a market factor, which scales with company-specific multipliers $\lambda_{i,t}$. Accordingly, rankings depend solely on $\lambda_{i,t}$ and are invariant across risk measures. We derive the joint asymptotic distribution of $\hat{\boldsymbol{\lambda}}_t = (\hat{\lambda}_{1,t}, \ldots, \hat{\lambda}_{N,t})'$, for a set of N institutions, construct Wald tests for pairwise equality of systemic riskiness between institutions, and extend the procedure to simultaneous inference with strong control of the family-wise error rate (FWER). This enables us to build valid confidence sets for institutions' ranks in systemic risk rankings. Up to our knowledge, this paper is the first to provide such a testing strategy that is of crucial practical importance when interpreting rankings and monitoring SIFIs over time.

A comprehensive simulation study supports the validity of the derived asymptotic results in finite samples. It showcases that the incorporation of time-variation in correlations matters. Imposing static correlations when the true dependence is dynamic leads to substantial distortions in SRM estimates. These distortions are considerably larger in finite samples than those arising from the efficiency loss of an overspecified model that assumes dynamic correlations when the true dependence is static. In a detailed empirical application, we study the systemic risk of 50 large US banks over the period from January 3, 2010, to December 31, 2020. Using daily stock return data from the CRSP database, we compute one-day-ahead out-of-sample forecasts of MES and ΔCoVaR, along with standard errors and confidence intervals of these predictions. We formally test for the presence of dynamic versus static correlations and find that the need for the general dynamic specification is prevalent. Moreover, standard errors in MES and ΔCoVaR estimates increase during market turmoil, such as the COVID-19 crisis, resulting in higher levels of estimation risk in financial downturns. Moreover, we test for differences in systemic risk both marginally and jointly across all distinct pairs of firms and construct corresponding confidence sets for ranks in systemic risk rankings. This analysis yields interesting and novel insights into the uncertainty in commonly used rankings. We show that the informativeness of such rankings varies substantially over time. While there are periods where the systemic riskiness of most banks can be statistically hardly distinguished, we also identify periods in which ranks can be clearly differentiated, rendering the resulting rankings economically informative.

The rest of this paper is organized as follows. In Section 1.2, we present closed-form expressions for computing MES and ΔCoVaR and introduce the semi-parametric

estimators for both. Section 1.3 discusses the asymptotic properties of the MES and ΔCoVaR estimators and develops statistical inference for the ranking of banks based on these measures. Section 1.4 presents a simulation study, illustrating the finite-sample properties of the estimators and analyzing the role of time-varying market-firm return dependence. Section 1.5 applies the methods to real data and provides novel evidence on estimation uncertainty of SRMs and corresponding rankings. Finally, Section 1.6 concludes the paper.

1.2 Estimating Systemic Risk under Multivariate Dynamics

In this section, we introduce closed-form expressions to compute MES and ΔCoVaR under the general class of MGARCH models. These build the basis to construct semi-parametric estimators for the conditional MES and ΔCoVaR.

1.2.1 Closed-Form Representations of MES and ΔCoVaR

SRMs are typically issued from a dynamic parametric or semi-parametric model specified by the researcher, the risk manager, or the regulatory authority. Consequently, the systemic measures are generally specified through a volatility equation for modeling the time-varying second moment (see e.g., Acharya et al., 2012; Girardi and Ergün, 2013; Brownlees and Engle, 2017) and the cdf of the joint distribution of $\boldsymbol{y}_t = (y_{m,t}, y_{i,t})'$ depending upon an unknown model parameter set $\boldsymbol{\theta}_0 \in \Theta \subseteq \mathbb{R}^p$. Let $\boldsymbol{y}_t = (y_{m,t}, y_{i,t})'$ be the (bivariate) vector of log-returns of the market and of the firm and consider the following data generating process (DGP) for the vector of log-returns,

$$\boldsymbol{y}_t = \boldsymbol{H}_t^{1/2}(\boldsymbol{\theta}_0)\,\boldsymbol{\eta}_t, \tag{1.4}$$

where $\boldsymbol{\eta}_t = (\eta_{m,t}, \eta_{i,t})'$ is a sequence of independent and identically distributed (i.i.d) $\mathbb{R}^2$-valued variables with $\mathbb{E}(\boldsymbol{\eta}_t) = \mathbf{0}$ and $\mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t') = \boldsymbol{I}_2$, where $\boldsymbol{I}_2$ is a (2×2) identity matrix, and the conditional variance-covariance matrix

$$\boldsymbol{H}_t(\boldsymbol{\theta}_0) = \boldsymbol{H}_t\left(\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots; \boldsymbol{\theta}_0\right), \tag{1.5}$$

is almost surely a positive definite (2×2) matrix depending on the information set $\mathcal{F}_{t-1}$ generated by the past values of $\boldsymbol{y}_t$. Assuming that (1.4) admits a non-anticipative stationary solution, $\boldsymbol{H}_t$ is the conditional variance-covariance matrix of $\boldsymbol{y}_t$ and takes the following form

$$\boldsymbol{H}_t(\boldsymbol{\theta}_0) = \begin{bmatrix} \sigma_{m,t}^2(\boldsymbol{\theta}_0) & \rho_{i,t}(\boldsymbol{\theta}_0)\,\sigma_{m,t}(\boldsymbol{\theta}_0)\,\sigma_{i,t}(\boldsymbol{\theta}_0) \\ \rho_{i,t}(\boldsymbol{\theta}_0)\,\sigma_{m,t}(\boldsymbol{\theta}_0)\,\sigma_{i,t}(\boldsymbol{\theta}_0) & \sigma_{i,t}^2(\boldsymbol{\theta}_0) \end{bmatrix}, \tag{1.6}$$

where $\sigma_{m,t}^2(\boldsymbol{\theta}_0)$ and $\sigma_{i,t}^2(\boldsymbol{\theta}_0)$ denote the volatilities of the market returns $y_{m,t}$ and firm return $y_{i,t}$, respectively, and $\rho_{i,t}(\boldsymbol{\theta}_0)$ denotes the correlation between the market and the firm returns at time t.

This representation encompasses the class of MGARCH models, allowing for time-varying volatilities for $y_{i,t}$ and $y_{m,t}$ as well as dynamic correlations between them (see Bauwens et al., 2006; Silvennoinen and Teräsvirta, 2009, for a survey). It requires specifying parametric models (based on $\boldsymbol{\theta}_0$) for the evolution of time-varying volatilities and correlations, however, it does not impose parametric assumptions on the distribution of $\boldsymbol{\eta}_t$. Notable examples within this context include the CCC-GARCH model introduced by Bollerslev (1990) and its extensions, such as the ECCC-GARCH model incorporating volatility spillovers (Jeantheau, 1998). Other widely adopted models are the BEKK model by Engle and Kroner (1995) as well as the DCC-GARCH model developed by Engle (2002) or the corrected DCC-GARCH of Aielli (2013). In the following, we will use the BEKK model as the prime example of an MGARCH specification with time-varying correlation and as a special case for our theoretical results but also in the provided simulations and the empirical study. We choose this setup as it allows us to rely on existing asymptotic theory for the associated parameter estimates. In contrast, for general DCC and corrected DCC models formally established asymptotic results are yet to be derived (see, e.g., Francq and Zakoïan, 2016). The conditional variance-covariance matrix $\boldsymbol{H}_t$ of the BEKK model, in its simplest BEKK-GARCH(1,1) form, is given by

$$\boldsymbol{H}_t = \boldsymbol{C}_0\boldsymbol{C}_0' + \boldsymbol{A}_0\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}'\boldsymbol{A}_0' + \boldsymbol{B}_0\boldsymbol{H}_{t-1}\boldsymbol{B}_0', \tag{1.7}$$

where $\boldsymbol{C}_0$ is a lower triangular parameter matrix and $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ are (2×2) parameter matrices. In the sequel, we consider the reduced diagonal BEKK model, where the matrices $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ are diagonal and thus the unknown parameters are given by $\boldsymbol{\theta}_0 = (\text{vech}(\boldsymbol{C}_0)', \text{diag}(\boldsymbol{A}_0)', \text{diag}(\boldsymbol{B}_0)')'$. For simplicity, we will refer to this model as BEKK.

For comparison purposes, we also consider a baseline specification with static dependence. Specifically, we use the CCC model of Bollerslev (1990). This specification allows us to isolate the role of time-varying correlation and to contrast SRMs under static versus dynamic dependence. In this case, the conditional variance-covariance matrix is given by

$$\boldsymbol{H}_t = \boldsymbol{D}_t\boldsymbol{R}\boldsymbol{D}_t, \tag{1.8}$$

where $\boldsymbol{D}_t = \text{diag}(\sigma_{m,t}, \sigma_{i,t})$ and $\boldsymbol{R} = \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}$, with ρ_i denoting the (time-invariant) correlation between firm i and the market returns. The marginal volatilities follow univariate GARCH(1,1) dynamics given by $\sigma^2_{m,t} = \omega_m + \alpha_m y^2_{m,t-1} + \beta_m \sigma^2_{m,t-1}$ and $\sigma^2_{i,t} = \omega_i + \alpha_i y^2_{i,t-1} + \beta_i \sigma^2_{i,t-1}$. The corresponding parameter vector is $\boldsymbol{\theta}_0 = (\omega_m, \alpha_m, \beta_m, \omega_i, \alpha_i, \beta_i, \rho_i)'$.

Generally, to derive analytically tractable representations for MES and ΔCoVaR based on (1.4), we need to specify $\boldsymbol{H}_t^{1/2}$. In this paper, we opt for the widely used Cholesky factorization of $\boldsymbol{H}_t$ which yields

$$\boldsymbol{H}_t^{1/2}(\boldsymbol{\theta}_0) = \begin{bmatrix} \sigma_{m,t}(\boldsymbol{\theta}_0) & 0 \\ \sigma_{i,t}(\boldsymbol{\theta}_0)\,\rho_{i,t}(\boldsymbol{\theta}_0) & \sigma_{i,t}\sqrt{1-\rho_{i,t}(\boldsymbol{\theta}_0)^2} \end{bmatrix}. \tag{1.9}$$

For potentially non-spherical innovation terms $\boldsymbol{\eta}_t$, a factorization $\boldsymbol{H}_t^{1/2}$ is necessary and affects the joint law of (y_{mt}, y_{it}) and thus MES and ΔCoVaR.[5] The Cholesky form in (1.9), however, depends on the ordering of the variables in $\boldsymbol{y}_t$ determining the direction of the effect between y_{mt} and y_{it}. In our setting, it is natural to model the market first, since the market as an aggregate is much more likely to induce shocks on individual firms than vice versa. This ordering is standard in the literature[6] and also consistent with the direction of the considered MES and ΔCoVaR specifications, which condition firm-specific returns on market events.

With the lower triangular form of the Cholesky decomposition, the MGARCH specification (1.4) simplifies into

$$\begin{aligned} y_{m,t} &= \sigma_{m,t}\eta_{m,t}, \\ y_{i,t} &= \sigma_{i,t}\rho_{i,t}\eta_{m,t} + \sigma_{i,t}\sqrt{1-\rho_{i,t}^2}\eta_{it}, \end{aligned}$$

where the market return is driven solely by market shocks, while the institution's return depends on both market and idiosyncratic shocks. Hence, a key advantage of this decomposition is to avoid the need of numerical integration of the joint pdf of $\boldsymbol{y}_t$ yielding simple analytical expressions for both considered SRMs. As shown in the propositions below, the Cholesky factorization of $\boldsymbol{H}_t$ allows us to decompose the two risk measures, MES and ΔCoVaR into dynamic univariate marginal components and time-varying dependence components.

Proposition 1.1 (MES)**.** *Given equations (1.4) and (1.9), the MES is defined as*

$$MES_{i,t}(\alpha, \boldsymbol{\theta}_0) = \sigma_{i,t}(\boldsymbol{\theta}_0)\,\rho_{i,t}(\boldsymbol{\theta}_0)\,\mu_{m,\alpha}, \tag{1.10}$$

where $\sigma_{i,t}(\boldsymbol{\theta}_0)$ is the standard deviation of the i-th firm return at time t, $\rho_{i,t}(\boldsymbol{\theta}_0)$ is the correlation between the i-th firm and the market returns at time t, $\mu_{m,\alpha} = \mathbb{E}(\eta_{m,t}|\eta_{m,t} \leq \xi_{m,\alpha})$ is the (time-invariant) expected shortfall of the market innovation $\eta_{m,t}$ with $\xi_{m,\alpha}$ being the (time-invariant) α-quantile of the market innovation.

Proof. See Appendix B.1. □

Hence, the MES of firm i depends on its time-varying volatility and return correlation with the market, scaled by the (time-invariant) expected shortfall of market innovations. Thus besides idiosyncratic time-varying conditional moments, the systemic riskiness of any company depends on the likelihood and magnitude of extremes in market returns. To estimate MES, the parameters $\boldsymbol{\theta}_0$ determining $\boldsymbol{H}_t$ and the truncated mean of the market innovations, $\mu_{m,\alpha}$ need to be estimated.

As shown by the following proposition, a similar representation can be derived for $\Delta CoVaR$.

[5] When the innovation $\boldsymbol{\eta}_t$ has a spherical distribution, the joint distribution of $\boldsymbol{y}_t$ depends only on $\boldsymbol{H}_t$ and is invariant to the chosen square root. In this case, there is no need to specify an ordering of the variables since any factorization leads to unique solutions for MES and ΔCoVaR.

[6] See, e.g., Acharya et al. (2012), Idier et al. (2014), White et al. (2015), Brownlees and Engle (2017), Perea et al. (2019), Nolde and Zhang (2020) and Armanious (2024), among others.

Proposition 1.2 (ΔCoVaR)**.** *Given the equations (1.4) and (1.9), the $\Delta CoVaR$ is defined as*

$$\Delta CoVaR_{i,t}(\alpha, \boldsymbol{\theta}_0) = \sigma_{i,t}(\boldsymbol{\theta}_0)\, \rho_{i,t}(\boldsymbol{\theta}_0)\, (\xi_{m,\alpha} - \xi_{m,0.5}), \tag{1.11}$$

where $\sigma_{i,t}(\boldsymbol{\theta}_0)$ is the standard deviation of the i-th firm return at time t, $\rho_{i,t}(\boldsymbol{\theta}_0)$ is the correlation between the i-th firm and the market return at time t, and $\xi_{m,\alpha}$ and $\xi_{m,0.5}$, respectively, are the time-invariant α-quantile and the time-invariant median of market innovations $\eta_{m,t}$.

Proof. See Appendix B.2. □

Similarly to MES, ΔCoVaR is driven by $\sigma_{i,t}(\boldsymbol{\theta}_0)$ and $\rho_{i,t}(\boldsymbol{\theta}_0)$, however, its link to the market arises through the (time-invariant) difference between the α-quantile and 0.5-quantile of market innovations $\Delta\xi_{m,\alpha} := \xi_{m,\alpha} - \xi_{m,0.5}$. Note that $\Delta CoVaR_{i,t}(\alpha)$ is independent of the β-level underlying the two CoVaRs in (1.3). Moreover, when the innovations $\eta_{m,t}$ are symmetrically distributed around zero, (1.11) simplifies to $\Delta CoVaR_{i,t}(\alpha) = \sigma_{i,t}\rho_{i,t}\xi_{m,\alpha}$. In this case, ΔCoVaR and MES coincide up to a multiplicative constant, which is either the expected shortfall in case of MES or the VaR of market innovations in case of ΔCoVaR.

Note that the expressions for MES and ΔCoVaR in Propositions 1.1 and 1.2 depend not only on the dynamic univariate marginals but also on the dynamic dependence structure captured by the time-varying correlation $\rho_{i,t}$ between firm and market returns as of (1.9). This extends the existing approaches of Hoga (2023) and Francq and Zakoïan (2025) that also propose closed-form formulas to compute dynamic MES and dynamic ΔCoVaR. In both frameworks, however, all time-variation in SRMs stems from marginal volatility $\sigma_{i,t}$ dynamics, while dependence remains time-invariant, ruling out time-varying spillovers between market and firm-level returns. Conversely, in our framework both marginal and dependence components are time-varying and – due to the Cholesky factorization (1.9) – admit an analytically tractable decomposition. Without such a factorization marginal and dependence dynamics would be intertwined in a single expression, which would strongly reduce analytical tractability and strongly affect the estimation and inference strategy for unknown parameters of the SRMs.

In practice, the innovation distribution $\eta_{m,t}$ can be either parametrically specified or left nonparametrically unspecified. When a parametric distribution is assumed for $\eta_{m,t}$, MES and ΔCoVaR are fully parametric and the quantities $\mu_{m,\alpha}$, and $\xi_{m,\alpha}$ might be available in closed-form. Key examples are the Gaussian distribution with $\mu_{m,\alpha} = -\phi(\Phi^{-1}(\alpha))/\alpha$ and $\xi_{m,\alpha} = \Phi^{-1}(\alpha)$, where $\phi(.)$ and $\Phi^{-1}(.)$ denote the pdf and the inverse of the cdf of a standard normal distribution, respectively, and the standardized Student distribution with $\mu_{m,\alpha} = -\sqrt{\frac{v-2}{v}} \times \frac{v+F^{-1}(\alpha,v)^2}{v-1} \times \frac{f(F^{-1}(\alpha,v),v)}{\alpha}$ and $\xi_{m,\alpha} = \sqrt{\frac{v-2}{v}} F^{-1}(\alpha, v)$, where $f(., v)$ and $F^{-1}(., v)$ denote the pdf and the inverse of the cdf of a Student distribution with v degrees of freedom.[7]

[7] More generally, the parameters $\mu_{m,\alpha}$ and $\xi_{m,\alpha}$ are made available for a large variety of parametric distributions (Laplace, Logistic, Pareto, Generalized Pareto distribution, Generalized extreme value distribution, etc.). See for instance Khokhlov (2016) for more details in elliptical distributions.

In the following section, we propose general estimators for the parameters $\boldsymbol{\theta}_0$ and $\mu_{m,\alpha}$ and $\xi_{m,\alpha}$ treating $\eta_{m,t}$ as an infinite dimensional nuisance parameter and thereby avoiding biases arising from the potential misspecification of a parametric distributional form.

1.2.2 Estimators for MES and ΔCoVaR

Based on the analytical SRM decompositions in Propositions 1.1 and 1.2, we propose a straightforward two-step plug-in quasi maximum likelihood estimator (QMLE) for MES and ΔCoVaR. This approach is empirically well-established for market risk measures such as VaR and ES constructed based on univariate GARCH models (see e.g., Gao and Song, 2008; Francq and Zakoïan, 2015). We extend it to SRMs building on a general MGARCH framework where univariate equation-by-equation modeling does not work.

In the first step, the parameters $\boldsymbol{\theta}_0$ are estimated by the standard QMLE using observations $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$. In the second step, the truncated mean $\mu_{m,\alpha}$ and the quantiles $\Delta\xi_{m,\alpha}$ are estimated using the estimated innovations resulting from the first-step estimator. The procedure is as follows.

1. Estimate the parameters $\boldsymbol{\theta}_0$ of the univariate volatility models for market and firm returns as well as of the corresponding (dynamic) correlation specification by QMLE, minimizing

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(\boldsymbol{y}_t' \boldsymbol{H}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t + \log|\boldsymbol{H}_t(\boldsymbol{\theta})| \right), \tag{1.12}$$

   where $\boldsymbol{\theta}$ denotes the vector of parameters associated with $\boldsymbol{H}_t(\boldsymbol{\theta})$, i.e., of $\sigma_{m,t}(\boldsymbol{\theta})$, $\sigma_{i,t}(\boldsymbol{\theta})$, and $\rho_{i,t}(\boldsymbol{\theta})$. Using the QML estimate $\hat{\boldsymbol{\theta}}_n$, we obtain estimates of (past) variances $\hat{\sigma}^2_{m,t} \equiv \sigma^2_{m,t}(\hat{\boldsymbol{\theta}}_n)$ and $\hat{\sigma}^2_{i,t} \equiv \sigma^2_{i,t}(\hat{\boldsymbol{\theta}}_n)$, and correlations $\hat{\rho}_{i,t} \equiv \rho_{i,t}(\hat{\boldsymbol{\theta}}_n)$, for all $t \leq n$ as well as their one-step-ahead prediction for $t = n+1$.[8]

2. Using the estimated volatilities $\hat{\sigma}^2_{m,t}$, $t = 1, ..., n$, the innovations of market returns are computed as $\hat{\eta}_{m,t} = y_{m,t}/\hat{\sigma}_{m,t}$. Then, $\Delta\xi_{m,\alpha}$ is estimated as

$$\Delta\hat{\xi}_{m,\alpha} = \hat{\xi}_{m,\alpha} - \hat{\xi}_{m,0.5}, \tag{1.13}$$

   where $\hat{\xi}_{m,\alpha} = \inf\{x : \hat{F}_{\eta_m}(x) \geq \alpha\}$ and $\hat{F}_{\eta_m}(x) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}(\hat{\eta}_{m,t} \leq x)$, and $\mu_{m,\alpha}$ is estimated by

$$\hat{\mu}_{m,\alpha} = \frac{\sum_{t=1}^{n} \hat{\eta}_{m,t} \mathbb{1}(\hat{\eta}_{m,t} < \hat{\xi}_{m,\alpha})}{\sum_{t=1}^{n} \mathbb{1}(\hat{\eta}_{m,t} < \hat{\xi}_{m,\alpha})}. \tag{1.14}$$

3. Using (1.10) and (1.11) plug-in estimators for MES and ΔCoVaR are obtained as

$$MES_{i,t}(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_{m,\alpha}) = \hat{\sigma}_{i,t}\hat{\rho}_{i,t}\hat{\mu}_{m,\alpha}, \tag{1.15}$$

$$\Delta CoVaR_{i,t}(\hat{\boldsymbol{\theta}}_n, \Delta\hat{\xi}_{m,\alpha}) = \hat{\sigma}_{i,t}\hat{\rho}_{i,t}\Delta\hat{\xi}_{m,\alpha}. \tag{1.16}$$

[8] In both the simulation study and the empirical application, we use the Oxford (`Matlab`) MFE Toolbox for MGARCH models to estimate $\boldsymbol{\theta}_0$ (Sheppard, 2009), where the objective function for QML estimation is adapted to align with the reduced form in (1.12).

Exogenous variables could, in principle, be incorporated through a two-step procedure. First, GARCH-X volatility equations are estimated separately, for which closed-form asymptotic results are available (see Han and Kristensen, 2014; Francq and Thieu, 2019). Second, the dynamic correlation parameters are estimated by minimizing (1.12), holding the variance parameters fixed at their first-step estimates. Related equation-by-equation approaches are considered by Francq and Sucarrat (2017) for multivariate log-GARCH-X models, while Section 3.2 of Bauwens et al. (2006) illustrates second-step correlation estimation based on a DCC specification. MES and ΔCoVaR could then be computed accordingly.

However, a closed-form asymptotic theory for general MGARCH-X models is not yet available. Developing such results would be necessary to formally justify this approach, and we therefore leave a rigorous treatment of MGARCH-X specifications for future research.

1.3 Statistical Inference

In this section, we provide statistical inference results for the proposed SRM estimators in Section 1.2.2. By allowing for a general time-varying dependence structure, the obtained results extend Francq and Zakoïan (2025). Moreover, using estimates of SRMs for systemic risk rankings, we show how SRM estimation uncertainty affects systemic risk rankings and provide formal inference for such ranks.

1.3.1 Asymptotic Distributions

For sake of clarity, we use the notation $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{0,\sigma_m}, \boldsymbol{\theta}_{0,\sigma_i}, \boldsymbol{\theta}_{0,\rho_i})' \in \mathbb{R}^p$, where $\boldsymbol{\theta}_{0,\sigma_m}$, $\boldsymbol{\theta}_{0,\sigma_i}$, and $\boldsymbol{\theta}_{0,\rho_i}$ are the true unknown parameters in the specifications of the market volatility $\sigma_{m,t}$, the firm volatility $\sigma_{i,t}$, and the correlation $\rho_{i,t}$. All respective components are estimated via QMLE in a first step as outlined in Section 1.2.2 and gathered in $\hat{\boldsymbol{\theta}}_n = \left(\hat{\boldsymbol{\theta}}_{\sigma_m}, \hat{\boldsymbol{\theta}}_{\sigma_i}, \hat{\boldsymbol{\theta}}_{\rho_i}\right)$. The key technical results are contained in Theorems 1.1 and 1.2 that establish the joint asymptotic distribution of all components in ΔCoVaR and MES. These results are then used to derive the conditional confidence intervals of the ΔCoVaR and MES estimates via Corollary 1.1. For all proofs we refer to the Appendix. For completeness, Theorem 1.4 in the Appendix C.3 also provides the asymptotic results of the semi-parametric estimator of the components of $CoVaR_{i,t}(\beta, \alpha)$, as defined in Eq. (1.2), under (1.4) and (1.9), and establishes their joint asymptotic properties.

We require the following standard assumptions on the MGARCH specification in (1.4) that were used in Section 1.2.1 to obtain the decompositions of the SRMs in Propositions 1.1 and 1.2.

Assumption 1.1. *$\boldsymbol{y}_t$ is strictly stationary and the ergodic solution of the MGARCH-type model (1.4), and $\boldsymbol{\eta}_t = (\eta_{m,t}, \eta_{i,t})'$ is a sequence of independent and identically distributed (i.i.d) $\mathbb{R}^2$-valued variables with $\mathbb{E}(\boldsymbol{\eta}_t) = \mathbf{0}$, $\mathbb{E}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') = \boldsymbol{I}_2$, where $\boldsymbol{I}_2$ is a (2×2) identity matrix.*

Moreover, we need the first step QMLE of the MGARCH model to satisfy a second order moment expansion of the log-likelihood function, that can be obtained for most M- and Z-estimators (see Chapter 5 of Van der Vaart (2000)) under granular conditions on the model specification and parameters. Essentially, the assumptions for consistency and asymptotic normality of the respective estimator imply the existence of such an expansion that is often used to show the statistical properties of (functionals of) the estimator. Thus we require that $\hat{\boldsymbol{\theta}}_n$ satisfies

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{m}\left(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0\right) + o_P(1), \tag{1.17}$$

where $\boldsymbol{m}$ is such that $\mathbb{E}[\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$. Using the MGARCH structure, we obtain $\boldsymbol{m}\left(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0\right)$ as

$$\boldsymbol{m}\left(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0\right) = \boldsymbol{\Delta}_{t-1} \operatorname{vec}\left(\boldsymbol{I}_2 - \boldsymbol{\eta}_t \boldsymbol{\eta}_t'\right), \tag{1.18}$$

with $\boldsymbol{\Delta}_{t-1} = -\boldsymbol{J}^{*-1} \frac{\partial \operatorname{vec}' \boldsymbol{H}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \left\{ \boldsymbol{H}_t^{-1/2}\left(\boldsymbol{\theta}_0\right) \otimes \boldsymbol{H}_t^{-1/2}\left(\boldsymbol{\theta}_0\right) \right\}'$, and $\boldsymbol{J}^* = \mathbb{E}(\frac{\partial^2 L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'})$ where $\boldsymbol{I}_2$ denotes the (2×2) identity matrix. The calculation of the explicit form in (1.18) of the inverse of the Fisher information matrix and the score function is standard for the MGARCH case in (1.12). We therefore refer to Francq and Zakoïan (2018) for details. We require that all components in (1.17) and (1.18) are well-defined and that the variable $\boldsymbol{\Delta}_t$ belongs to L^2 and $\mathbb{E}\left[\boldsymbol{\Delta}_t\right] = \boldsymbol{\Lambda}$ is full row rank. This is ensured by the following standard granular conditions:

Assumption 1.2. *(i) The set of potential parameter values $\boldsymbol{\Theta}$ ($\subset \mathbb{R}^p$) is compact. (ii) $\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta})$ is concave over the parameter space for all $\boldsymbol{\eta}_t$. (iii) $\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta})$ is measurable in $\boldsymbol{\eta}_t$ for all $\boldsymbol{\theta}$ in $\boldsymbol{\Theta}$.*

Assumption 1.3. *(a) $\mathbb{E}[\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta})]$ is uniquely maximized at $\boldsymbol{\theta}_0 \in int(\boldsymbol{\Theta})$, where int() marks the interior of the set. (b) $\mathbb{E}[sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0)|] < \infty$.*

Assumption 1.4. *$\mathbb{E}[\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0)\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0)']$ exists, is positive definite and nonsingular. $\mathbb{E}[\sup_{\boldsymbol{\theta} \in \mathcal{B}} ||\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0)\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0)'||] < \infty$, for some neighborhood $\mathcal{B}$ of $\boldsymbol{\theta}_0$.*

Assumption 1.5. *$\mathbb{E}\left|\eta_{k,t}\right|^{4(1+\delta)} < \infty$, for some $\delta > 0$.*

Assumption 1.6. *For any sequence $(\boldsymbol{y}_t)$, the function $\boldsymbol{\theta} \mapsto \boldsymbol{H}_t^{1/2}\left(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots; \boldsymbol{\theta}\right)$ is continuously differentiable over $\boldsymbol{\Theta}$.*

Note that Assumptions 1.1–1.5 imply consistency of the QMLE pre-step. The smoothness in Assumption 1.6 ensures that we can statistically control the linear Taylor expansion (1.17). It is essentially implied by smoothness of its element functions $\sigma_{m,t}$, $\sigma_{i,t}$, and $\rho_{i,t}$ - where smoothness on the latter is only a restriction in time-dynamic correlation cases like the BEKK model.

For specific multivariate cases of the MGARCH model, also the asymptotic properties of QMLE are well established under Assumptions 1.1–1.6. See, for example, Ling and McAleer (2003) and Francq and Zakoïan (2012) for the CCC model, and Comte

and Lieberman (2003) and Hafner and Preminger (2009) for the BEKK model.[9] In these particular cases, we can derive simpler versions of (1.18) using the respective parametric forms of $\boldsymbol{H}_t$. Simple calculations show that for the CCC model $\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0) = -\boldsymbol{J}^{*-1}[2\frac{\partial\sigma_{m,t}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_{0,\sigma_m}}{\sigma_{m,t}(\boldsymbol{\theta}_0)}(1-\eta_{m,t}^2), 2\frac{\rho_i(\boldsymbol{\theta}_0)}{\rho_i(\boldsymbol{\theta}_0)^2-1}(1-\eta_{i,t}^2),\ 2\frac{\partial\sigma_{i,t}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_{0,\sigma_i}}{\sigma_{i,t}(\boldsymbol{\theta}_0)}(1-\eta_{i,t}^2)]'$, while for the BEKK model $\boldsymbol{m}(\boldsymbol{\eta}_t, \boldsymbol{\theta}_0) = -\boldsymbol{J}^{*-1}[2\frac{\partial\sigma_{m,t}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_{0,\sigma_m}}{\sigma_{m,t}(\boldsymbol{\theta}_0)}(1-\eta_{m,t}^2), 2\frac{\rho_{i,t}(\boldsymbol{\theta}_0)\partial\rho_{i,t}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_{0,\rho_i}}{\rho_{i,t}(\boldsymbol{\theta}_0)^2-1}(1-\eta_{i,t}^2),\ 2\frac{\partial\sigma_{i,t}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_{0,\sigma_i}}{\sigma_{i,t}(\boldsymbol{\theta}_0)}(1-\eta_{i,t}^2)]'$.

Moreover, together with Assumption 1.6, the following condition is required to ensure that we can use and statistically control an asymptotic Taylor expansion in the conditional quantile $\hat{\xi}_{m,\alpha}$ and truncated mean $\hat{\mu}_{m,\alpha}$ of the innovation term that enters both SRMs.

Assumption 1.7. *Assume that $\eta_{m,t}$ are i.i.d. random variables with distribution function $F(x)$ and Lebesgue density f. Additionally, assume that $f(x) > 0$ almost surely (a.s.), $\lim_{|x|\to\infty} xf(x) = 0$, and $\sup_x(1+x^2)|f'(x)| < \infty$.*

Assumption 1.7 is essential for characterizing the impact of the first-step estimation on the subsequent second-step estimation.

The following Theorems 1.1 and 1.2 establish the asymptotic properties of the two-step semi-parametric estimators for the unknown parameter vectors $(\boldsymbol{\theta}_0', \Delta\xi_{m,\alpha})'$ and $(\boldsymbol{\theta}_0', \mu_{m,\alpha})'$ of ΔCoVaR and MES.

Theorem 1.1 (Asymptotic distribution of ΔCoVaR components)**.** *Suppose that $\eta_{m,t}$ admits a density f that is continuous and strictly positive in a neighborhood of $\xi_{m,\alpha}$ and that Assumptions 1.1–1.7 hold. Define $\Delta\xi_{m,\alpha} = \xi_{m,\alpha} - \xi_{m,0.5}$ and $\Delta\hat{\xi}_{m,\alpha} = \hat{\xi}_{m,\alpha} - \hat{\xi}_{m,0.5}$ Then, we have under model (1.4) and the Cholesky decomposion $\boldsymbol{H}_t(\boldsymbol{\theta}) = \boldsymbol{H}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{H}_t^{1/2}(\boldsymbol{\theta})'$ from (1.9) for a fixed quantile level α*

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \Delta\hat{\xi}_{m,\alpha} - \Delta\xi_{m,\alpha}\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_\alpha := \begin{pmatrix}\boldsymbol{\Psi} & \boldsymbol{\Xi}_{\theta\xi} \\ \boldsymbol{\Xi}_{\theta\xi}' & \Upsilon_\alpha\end{pmatrix}\right),$$

where the components of $\boldsymbol{\Sigma}_\alpha$ are given by

$$\begin{aligned}
\boldsymbol{\Psi} &= \mathbb{E}\left(\boldsymbol{\Delta}_t Var(\boldsymbol{V}(\boldsymbol{\eta}_t))\boldsymbol{\Delta}_t'\right), \\
\boldsymbol{\Xi}_{\theta\xi} &= -\Delta\xi_{m,\alpha}\boldsymbol{\Psi}\boldsymbol{\Omega}_\xi - \frac{1}{f(\xi_{m,\alpha})}\boldsymbol{\Lambda}\boldsymbol{W}_\alpha + \frac{1}{f(\xi_{m,0.5})}\boldsymbol{\Lambda}\boldsymbol{W}_{0.5}, \\
\Upsilon_\alpha &= \Delta\xi_{m,\alpha}^2\boldsymbol{\Omega}_\xi'\boldsymbol{\Psi}\boldsymbol{\Omega}_\xi + \frac{2\Delta\xi_{m,\alpha}}{f(\xi_{m,\alpha})}\boldsymbol{\Omega}_\xi'\boldsymbol{\Lambda}\boldsymbol{W}_\alpha - \frac{2\Delta\xi_{m,\alpha}}{f(\xi_{m,0.5})}\boldsymbol{\Omega}_\xi'\boldsymbol{\Lambda}\boldsymbol{W}_{0.5} \\
&\quad + \frac{\alpha(1-\alpha)}{f^2(\xi_{m,\alpha})} + \frac{1}{4f^2(\xi_{m,0.5})} - \frac{\alpha}{f(\xi_{m,\alpha})f(\xi_{m,0.5})},
\end{aligned}$$

with $\boldsymbol{V}(\boldsymbol{\eta}_t) = \mathrm{vec}\{\boldsymbol{I}_2 - \boldsymbol{\eta}_t\boldsymbol{\eta}_t'\}$, $\boldsymbol{W}_\alpha = Cov(\boldsymbol{V}(\boldsymbol{\eta}_t), \mathbb{1}(\eta_{m,t} \le \xi_{m,\alpha}) - \alpha))$,
$\boldsymbol{\Omega}_\xi = \mathbb{E}\left[\frac{1}{\sigma_{m,t}(\boldsymbol{\theta})}\frac{\partial\sigma_{m,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]$, *and* $\boldsymbol{\Delta}_{t-1} = -\boldsymbol{J}^{*-1}\frac{\partial\,\mathrm{vec}'\boldsymbol{H}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\left\{\boldsymbol{H}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\boldsymbol{H}_t^{-1/2}(\boldsymbol{\theta}_0)\right\}'$,
where $\boldsymbol{J}^* = \mathbb{E}(\frac{\partial^2 L_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'})$ and $\mathbb{E}(\boldsymbol{\Delta}_{t-1}) = \boldsymbol{\Lambda}$.

[9] Note that in particular for all univariate subcases such asymptotic properties and the required assumptions are standard (see, e.g., Berkes and Horváth, 2003).

Proof. See Appendix C.1. □

Theorem 1.2 (Asymptotic distribution of MES components)**.** *Suppose that $\eta_{m,t}$ admits a density f that is continuous and strictly positive in a neighborhood of $\xi_{m,\alpha}$ and that Assumptions 1.1–1.7 hold. Then we have under model (1.4) and the Cholesky decomposion $\boldsymbol{H}_t(\boldsymbol{\theta}) = \boldsymbol{H}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{H}_t^{1/2}(\boldsymbol{\theta})'$ from (1.9) for a fixed quantile level α,*

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \hat{\mu}_{m,\alpha} - \mu_{m,\alpha}\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \overline{\boldsymbol{\Sigma}}_\alpha := \begin{pmatrix}\boldsymbol{\Psi} & \boldsymbol{\Phi}_{\theta\mu} \\ \boldsymbol{\Phi}'_{\theta\mu} & \zeta_\alpha\end{pmatrix}\right),$$

where the components of $\overline{\boldsymbol{\Sigma}}_\alpha$ are given by

$$\begin{aligned}
\boldsymbol{\Psi} &= \mathbb{E}\left(\boldsymbol{\Delta}_t Var(\boldsymbol{V}(\boldsymbol{\eta}_t))\boldsymbol{\Delta}_t'\right), \\
\boldsymbol{\Phi}_{\theta\mu} &= \frac{1}{\alpha}\boldsymbol{\Lambda}\boldsymbol{M}_\alpha - \frac{\mu_{m,\alpha}}{2}\boldsymbol{\Psi}\boldsymbol{\Omega}_\mu, \\
\zeta_\alpha &= \frac{\mu_{m,\alpha}^2}{4}\boldsymbol{\Omega}_\mu'\boldsymbol{\Psi}\boldsymbol{\Omega}_\mu - \frac{\mu_{m,\alpha}}{\alpha}\boldsymbol{\Omega}_\mu'\boldsymbol{\Lambda}\boldsymbol{M}_\alpha \\
&\quad + \frac{1}{\alpha^2}\left(p_\alpha + \alpha(1-\alpha)\xi_{m,\alpha}^2 - \alpha^2\mu_{m,\alpha}^2 - 2\alpha(1-\alpha)\mu_{m,\alpha}\xi_{m,\alpha}\right),
\end{aligned}$$

with $\boldsymbol{V}(\boldsymbol{\eta}_t) = \mathrm{vec}\{\boldsymbol{I}_2 - \boldsymbol{\eta}_t\boldsymbol{\eta}_t'\}$, $\boldsymbol{M}_\alpha = Cov(\boldsymbol{V}(\boldsymbol{\eta}_t), (\eta_{m,t} - \xi_{m,\alpha})\mathbb{1}(\eta_{m,t} \le \xi_{m,\alpha}))$, $\boldsymbol{\Delta}_{t-1} = -\boldsymbol{J}^{*-1}\frac{\partial \mathrm{vec}'\boldsymbol{H}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\left\{\boldsymbol{H}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\boldsymbol{H}_t^{-1/2}(\boldsymbol{\theta}_0)\right\}'$, *where* $\boldsymbol{J}^* = \mathbb{E}(\frac{\partial^2 L_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'})$, $\mathbb{E}(\boldsymbol{\Delta}_{t-1}) = \boldsymbol{\Lambda}$; $p_\alpha = \mathbb{E}[\eta_{m,t}^2\mathbb{1}(\eta_{m,t} \le \xi_{m,\alpha})]$, *and* $\boldsymbol{\Omega}_\mu = \mathbb{E}\left[\frac{2}{\sigma_{m,t}(\boldsymbol{\theta}_0)}\frac{\partial\sigma_{m,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]$.

Proof. See Appendix C.4. □

In both theorems, the asymptotic variance of the first-step parameter estimates $\hat{\boldsymbol{\theta}}_n$, denoted by $\boldsymbol{\Psi}$, takes the standard sandwich form and enters the variances Υ_α and ζ_α of the second-step estimators $\Delta\hat{\xi}_{m,\alpha}$ and $\hat{\mu}_{m,\alpha}$, respectively. Accordingly, both variances Υ_α and ζ_α consist of two additive components. The first component captures the propagation of estimation risk from the first-step estimator $\hat{\boldsymbol{\theta}}_n$ to the second step, while the second one captures the sampling variability that would arise if the true innovations $\eta_{m,t}(\boldsymbol{\theta}_0)$ were observed, i.e., without nuisance of the first-step estimate. For $\Delta\hat{\xi}_{m,\alpha}$, the second term consists of $\frac{\alpha(1-\alpha)}{f^2(\xi_{m,\alpha})}$, $\frac{1}{4f^2(\xi_{m,0.5})}$, and $-\frac{\alpha}{f(\xi_{m,\alpha})f(\xi_{m,0.5})}$, which are the variances of the α- and 0.5-quantiles of $\eta_{m,t}$ and their covariance, respectively. For $\hat{\mu}_{m,\alpha}$, the corresponding term is $\frac{1}{\alpha^2}\left(p_\alpha + \alpha(1-\alpha)\xi_{m,\alpha}^2 - \alpha^2\mu_{m,\alpha}^2 - 2\alpha(1-\alpha)\mu_{m,\alpha}\xi_{m,\alpha}\right)$, that is the limit variance of the nonparametric expected shortfall using truncated averages (see Chen, 2008). Conversely, the first component of the variances of $\Delta\hat{\xi}_{m,\alpha}$ and $\hat{\mu}_{m,\alpha}$ (capturing the propagation of estimation risk) depends only on the volatility parameters $\boldsymbol{\theta}_{0,\sigma_m}$, which are – as implied by the Cholesky structure – sufficient for estimating η_{mt}. This is ensured by the vectors $\boldsymbol{\Omega}_\xi$, $\boldsymbol{W}_\alpha$, $\boldsymbol{\Omega}_\mu$, and $\boldsymbol{M}_\alpha$, which isolate the components of $\boldsymbol{\theta}_{0,\sigma_m}$ within the matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Lambda}$, respectively.

Finally, differences between MGARCH models allowing for time-varying return cross-dependencies (such as BEKK) and MGARCH models with constant cross-dependencies

(such as CCC) are particularly reflected in $\mathbf{\Delta}_{t-1}$ and thus $\mathbf{\Lambda}$. Thus, the main source of estimation uncertainty in this respect arises through this channel, which influences $\mathbf{\Psi}$ and the cross-covariances $\mathbf{\Xi}_{\theta\xi}$ and $\mathbf{\Phi}_{\theta\mu}$.

To compute (asymptotic) confidence intervals (CIs) for $MES_{i,t}(\alpha)$ and $\Delta CoVaR_{i,t}(\alpha)$, recall that both measures are constructed conditionally on $\mathcal{F}_{t-1}$, see Eqs. (1.1)-(1.3). Hence, their randomness stems from the randomness induced by the estimators $(\hat{\boldsymbol{\theta}}_n', \hat{\mu}_{m,\alpha})'$ and $(\hat{\boldsymbol{\theta}}_n', \Delta\hat{\xi}_{m,\alpha})'$ (inferred from Theorems 1.1 and 1.2), while the sample $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{t-1}\}$ is treated as fixed.[10] Accordingly, consider the out-of-sample period $t = n+1$, i.e., the first period following the estimation sample. The conditional limit distributions and corresponding CIs are then obtained by conditioning on past information $\mathcal{F}_n \subseteq \boldsymbol{y}_n, \boldsymbol{y}_{n-1}, \ldots$ and propagating the randomness induced by $(\hat{\boldsymbol{\theta}}_n, \hat{\mu}_{m,\alpha})'$ and $(\hat{\boldsymbol{\theta}}_n, \Delta\hat{\xi}_{m,\alpha})'$. From Propositions 1.1 and 1.2 we have $\text{MES}_{i,n+1}(\alpha) = \text{MES}_{i,n+1}(\alpha; \boldsymbol{\theta}_0, \mu_{m,\alpha}) = \sigma_{i,n+1}(\boldsymbol{\theta}_0)\, \rho_{i,n+1}(\boldsymbol{\theta}_0)\, \mu_{m,\alpha}$ and $\Delta\text{CoVaR}_{i,n+1}(\alpha) = \Delta\text{CoVaR}_{i,n+1}(\alpha; \boldsymbol{\theta}_0, \Delta\xi_{m,\alpha}) = \sigma_{i,n+1}(\boldsymbol{\theta}_0)\, \rho_{i,n+1}(\boldsymbol{\theta}_0)\, (\xi_{m,\alpha} - \xi_{m,0.5})$. Corollary 1.1 below gives the corresponding asymptotic distributions which are essential for constructing $(1-\tau)100\%$ CIs for $\widehat{MES}_{i,n+1}(\alpha)$ and $\widehat{\Delta\text{CoVaR}}_{i,n+1}(\alpha)$.

Corollary 1.1 (Asymptotic distributions of $\widehat{\Delta\text{CoVaR}}_{i,n+1}(\alpha)$ and $\widehat{\text{MES}}_{i,n+1}(\alpha)$)**.**

1. *Under the assumptions of Theorem 1.1, the estimated risk measure* $\widehat{\Delta\text{CoVaR}}_{i,n+1}(\alpha) = \Delta\text{CoVaR}_{i,n+1}(\alpha; \hat{\boldsymbol{\theta}}_n, \Delta\hat{\xi}_{m,\alpha})$ *of* $\Delta\text{CoVaR}_{i,n+1}(\alpha)$ *satisfies*

$$\sqrt{n}\left(\widehat{\Delta\text{CoVaR}}_{i,n+1}(\alpha) - \Delta\text{CoVaR}_{i,n+1}(\alpha)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^{\xi}_{\alpha,n+1}),$$

where $\gamma^{\xi}_{\alpha,n+1} = \left[\frac{\partial\Delta\text{CoVaR}_{i,n+1}(\alpha;\boldsymbol{\theta},\Delta\xi)}{\partial(\boldsymbol{\theta}',\Delta\xi)}\right]_{(\boldsymbol{\theta}_0,\Delta\xi_{m,\alpha})} \mathbf{\Sigma}_\alpha \left[\frac{\partial\Delta\text{CoVaR}_{i,n+1}(\alpha;\boldsymbol{\theta},\Delta\xi)}{\partial(\boldsymbol{\theta}',\Delta\xi)}\right]'_{(\boldsymbol{\theta}_0,\Delta\xi_{m,\alpha})}$, *with* $\mathbf{\Sigma}_\alpha$ *as of Theorem 1.1.*

2. *Under the assumptions of Theorem 1.2, the estimated risk measure* $\widehat{\text{MES}}_{i,n+1}(\alpha) = \text{MES}_{i,n+1}(\alpha; \hat{\boldsymbol{\theta}}_n, \hat{\mu}_{m,\alpha})$ *of* $\text{MES}_{i,n+1}(\alpha)$ *satisfies*

$$\sqrt{n}\left(\widehat{\text{MES}}_{i,n+1}(\alpha) - \text{MES}_{i,n+1}(\alpha)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^{\mu}_{\alpha,n+1}),$$

where $\gamma^{\mu}_{\alpha,n+1} = \left[\frac{\partial\text{MES}_{i,n+1}(\alpha;\boldsymbol{\theta},\mu)}{\partial(\boldsymbol{\theta}',\mu)}\right]_{(\boldsymbol{\theta}_0,\mu_{m,\alpha})} \overline{\mathbf{\Sigma}}_\alpha \left[\frac{\partial\text{MES}_{i,n+1}(\alpha;\boldsymbol{\theta},\mu)}{\partial(\boldsymbol{\theta}',\mu)}\right]'_{(\boldsymbol{\theta}_0,\mu_{m,\alpha})}$, *with* $\overline{\mathbf{\Sigma}}_\alpha$ *as of Theorem 1.2.*

Proof. We essentially use that asymptotic variances of $\widehat{MES}_{i,n+1}(\alpha)$ and $\widehat{\Delta\text{CoVaR}}_{i,n+1}(\alpha)$ can be derived from the limiting joint distribution of $(\hat{\boldsymbol{\theta}}_n', \hat{\mu}_{m,\alpha})'$ for MES and $(\hat{\boldsymbol{\theta}}_n', \Delta\hat{\xi}_{m,\alpha})'$ for ΔCoVaR (see Theorems 1.1 and 1.2) and applications of the delta method utilizing Propositions 1.1 and 1.2. For details, see Appendix C.5. □

[10] This corresponds to the typical dual treatment of the sample $\{\boldsymbol{y}_t\}_{t=1}^n$, as it is standard in time-series analysis: CIs for statistics in $t = n+1$, which condition on $\mathcal{F}_n \subseteq \boldsymbol{y}_n, \boldsymbol{y}_{n-1}, \ldots$, treat $\mathcal{F}_n$ as fixed, while any uncertainty stemming from estimates of $\boldsymbol{\theta}_0$ is captured by asymptotic distributions of $\hat{\boldsymbol{\theta}}_n$, treating $\{\boldsymbol{y}_t\}_{t=1}^n$ as random, see, e.g., Beutner et al. (2021).

Note that the variances of $\widehat{\Delta\text{CoVaR}}_{i,n+1}(\alpha)$ and $\widehat{\text{MES}}_{i,n+1}(\alpha)$ are model-dependent, and may therefore be either time-invariant or time-varying depending on the model specification and on the application of the delta method. For example, under the CCC model, we have $\partial\rho_{i,n+1}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_{0,\sigma_i} = \mathbf{0}$ and $\partial\rho_{i,n+1}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_{0,\sigma_m} = \mathbf{0}$, whereas this is not the case for the BEKK model due to the presence of standardized residuals in the correlation equation.

For feasible asymptotic confidence intervals we require consistent estimates of all theoretical quantities entering the asymptotic variance-covariance matrices in Theorems 1.1 and 1.2. These are straightforward sample analogues and provided for completeness in Appendix C.6. With these estimates $\hat{\boldsymbol{\Sigma}}_\alpha$ and $\hat{\overline{\boldsymbol{\Sigma}}}_\alpha$ and Corollary 1.1 we obtain the estimated asymptotic CI for $\Delta\text{CoVaR}_{i,n+1}(\alpha)$ and $\text{MES}_{i,n+1}(\alpha)$, respectively, at $(1-\tau)100\%$ confidence level as

$$\widehat{\Delta\text{CoVaR}}_{i,n+1}(\alpha) \;\pm\; z_{1-\tau/2}\sqrt{\hat{\gamma}^{\xi}_{\alpha,n+1}/n}, \tag{1.19}$$

$$\widehat{\text{MES}}_{i,n+1}(\alpha) \;\pm\; z_{1-\tau/2}\sqrt{\hat{\gamma}^{\mu}_{\alpha,n+1}/n}, \tag{1.20}$$

where $z_{1-\tau/2}$ is the quantile of the standard normal distribution at level $1-\tau/2$, $\hat{\gamma}^{\xi}_{\alpha,n+1} = \left[\frac{\partial\Delta\text{CoVaR}_{i,n+1}(\alpha;\boldsymbol{\theta},\Delta\xi)}{\partial(\boldsymbol{\theta}',\Delta\xi)}\right]_{(\hat{\boldsymbol{\theta}}_n,\Delta\hat{\xi}_{m,\alpha})} \hat{\boldsymbol{\Sigma}}_\alpha \left[\frac{\partial\Delta\text{CoVaR}_{i,n+1}(\alpha;\boldsymbol{\theta},\Delta\xi)}{\partial(\boldsymbol{\theta}',\Delta\xi)}\right]'_{(\hat{\boldsymbol{\theta}}_n,\Delta\hat{\xi}_{m,\alpha})}$, and $\hat{\gamma}^{\mu}_{\alpha,n+1} = \left[\frac{\partial\text{MES}_{i,n+1}(\alpha;\boldsymbol{\theta},\mu)}{\partial(\boldsymbol{\theta}',\mu)}\right]_{(\hat{\boldsymbol{\theta}}_n,\hat{\mu}_{m,\alpha})} \hat{\overline{\boldsymbol{\Sigma}}}_\alpha \left[\frac{\partial\text{MES}_{i,n+1}(\alpha;\boldsymbol{\theta},\mu)}{\partial(\boldsymbol{\theta}',\mu)}\right]'_{(\hat{\boldsymbol{\theta}}_n,\hat{\mu}_{m,\alpha})}$, with $\hat{\boldsymbol{\Sigma}}_\alpha$ and $\hat{\overline{\boldsymbol{\Sigma}}}_\alpha$ consistent estimators of $\boldsymbol{\Sigma}_\alpha$ and $\overline{\boldsymbol{\Sigma}}_\alpha$, respectively, as provided in Appendix C.6.

1.3.2 Testing for Equality of Banks' Systemic Riskiness

To statistically assess whether differences in systemic risk rankings are significant, we develop pairwise inference for systemic risk comparisons across banks. Consider two institutions j and k with $j,k \in I \equiv \{1,\dots,N\}$ selected among N institutions. A test of equality between their MES and ΔCoVaR at time t are given by the null hypothesises $H_0 : MES_{j,t}(\alpha,\boldsymbol{\theta}_{0,j}) = MES_{k,t}(\alpha,\boldsymbol{\theta}_{0,k})$, and $H_0 : \Delta CoVaR_{j,t}(\alpha,\boldsymbol{\theta}_{0,j}) = \Delta CoVaR_{k,t}(\alpha,\boldsymbol{\theta}_{0,k})$. From (1.10) and (1.11) it becomes obvious that $MES_{i,t}(\alpha,\boldsymbol{\theta}_{0,i})$ and $\Delta CoVaR_{i,t}(\alpha,\boldsymbol{\theta}_{0,i})$ share a common market factor $\mu_{m,\alpha}$ or $\Delta\xi_{m,\alpha}$, respectively and only differ by multipliers $\lambda_{i,t} \equiv \lambda_{i,t}(\boldsymbol{\theta}_{0,i}) = \sigma_{i,t}(\boldsymbol{\theta}_{0,i})\,\rho_{i,t}(\boldsymbol{\theta}_{0,i})$. Accordingly, systemic risk rankings depend solely on the institution-specific component $\lambda_{i,t}$, which can be seen as bank i's exposure to systemic risk. Moreover, rankings are invariant between MES and ΔCoVaR since $\lambda_{i,t}$ is scaled by a common factor across all institutions, which does not affect the ordering. Accordingly, testing for the equality of MES (or ΔCoVaR, respectively) between institutions j and k boils down to the hypothesis

$$H_0 :\; \lambda_{j,t}(\boldsymbol{\theta}_{0,j}) = \lambda_{k,t}(\boldsymbol{\theta}_{0,k}). \tag{1.21}$$

With a slight refinement of notation define $\boldsymbol{\theta}_{0,i}$ as the parameter vector associated with MES or ΔCoVaR of institution i.[11] Define $\boldsymbol{G}_{i,t}(\boldsymbol{\theta}_{0,i}) := \nabla_{\boldsymbol{\theta}_i}\lambda_{i,t}(\boldsymbol{\theta}_{0,i}) =$

[11] This indexing was not necessary before.

$\rho_{i,t}(\boldsymbol{\theta}_{0,i}) \dfrac{\partial \sigma_{i,t}(\boldsymbol{\theta}_{0,i})}{\partial \boldsymbol{\theta}_i} + \sigma_{i,t}(\boldsymbol{\theta}_{0,i}) \dfrac{\partial \rho_{i,t}(\boldsymbol{\theta}_{0,i})}{\partial \boldsymbol{\theta}_i} \in \mathbb{R}^p$ as the gradient vector of $\lambda_{i,t}$ with respect to $\boldsymbol{\theta}_{0,i}$. The following three corollaries build on the asymptotic distribution of $\boldsymbol{\lambda}_t(\hat{\boldsymbol{\theta}}_n)$ as derived in Appendix G. Throughout this subsection we assume that the granular conditions 1.1–1.7 hold. Moreover we require that a linear expansion of the Bahadur type (1.17) also exists for all parameters when estimating the MGARCH dynamics jointly for all N institution market pairs.

Then, Corollary 1.2 below provides a Wald test for hypothesis (1.21).

Corollary 1.2 (Wald test for $\lambda_{j,t}(\boldsymbol{\theta}_{0,j}) = \lambda_{k,t}(\boldsymbol{\theta}_{0,k})$)**.** *A Wald statistic for the null hypothesis (1.21) is given by*

$$\hat{Z}_{j,k,t} = \frac{\lambda_{j,t}(\hat{\boldsymbol{\theta}}_{n,j}) - \lambda_{k,t}(\hat{\boldsymbol{\theta}}_{n,k})}{n^{-1/2}\sqrt{\hat{\iota}_{j,k,t}}}, \tag{1.22}$$

where $\hat{\iota}_{j,k,t} = \boldsymbol{G}_{j,t}(\hat{\boldsymbol{\theta}}_{n,j})' \hat{\boldsymbol{\Psi}}_{jj} \boldsymbol{G}_{j,t}(\hat{\boldsymbol{\theta}}_{n,j}) + \boldsymbol{G}_{k,t}(\hat{\boldsymbol{\theta}}_{n,k})' \hat{\boldsymbol{\Psi}}_{kk} \boldsymbol{G}_{k,t}(\hat{\boldsymbol{\theta}}_{n,k}) - 2\, \boldsymbol{G}_{j,t}(\hat{\boldsymbol{\theta}}_{nj})' \hat{\boldsymbol{\Psi}}_{jk} \boldsymbol{G}_{k,t}(\hat{\boldsymbol{\theta}}_{n,k})$. *Under Eq. (1.21), we have* $\hat{Z}_{j,k,t} \xrightarrow{d} \mathcal{N}(0,1)$.

Testing equality across *all* institutions at t yields the hypothesis

$$H_0 : \ \lambda_{1,t}(\boldsymbol{\theta}_{0,1}) = \lambda_{2,t}(\boldsymbol{\theta}_{0,2}) = \ldots = \lambda_{N,t}(\boldsymbol{\theta}_{0,N}). \tag{1.23}$$

Corollary 1.3 formulates a test for simultaneous pairwise inference with strong control of the family-wise error rate (FWER).

Corollary 1.3 (Joint pairwise tests)**.** *Let* $\hat{\boldsymbol{Z}}_t$ *denote the stacked vector of Wald statistics from Corollary 1.2, ordered lexicographically,* $\hat{\boldsymbol{Z}}_t = \left(\hat{Z}_{1,2,t}, \hat{Z}_{1,3,t}, \ldots, \hat{Z}_{1,N,t}, \hat{Z}_{2,3,t}, \ldots, \hat{Z}_{N-1,N,t}\right)'$, *according to the set* $\mathcal{P} = \{(j,k) : 1 \le j < k \le N\}$ *with* $|\mathcal{P}| = N(N-1)/2$. *Define the contrast matrix* $\hat{\boldsymbol{A}}_t \in \mathbb{R}^{|\mathcal{P}| \times N}$ *by*

$$\left(\hat{\boldsymbol{A}}_t\right)_{(j,k),m} = \frac{\mathbf{1}\{m = j\} - \mathbf{1}\{m = k\}}{n^{-1/2}\sqrt{\hat{\iota}_{j,k,t}}}, \qquad (j,k) \in \mathcal{P}, \ \ m = 1, \ldots, N,$$

so that $\hat{\boldsymbol{Z}}_t = \hat{\boldsymbol{A}}_t \boldsymbol{\lambda}_t(\hat{\boldsymbol{\theta}}_n) = \left(\hat{Z}_{j,k,t} : (j,k) \in \mathcal{P}\right)'$. *Define the maximum statistic*

$$\hat{M}_t \ = \ \max_{(j,k) \in \mathcal{P}} |\hat{\boldsymbol{Z}}_t|. \tag{1.24}$$

Then, the null hypothesis (1.23) is rejected whenever $\hat{M}_t > c_{1-\tau,t}$, *where* $c_{1-\tau,t}$ *denotes the* $(1-\tau)$*-quantile of the null distribution of* $\hat{M}_t$. *Consequently, under multiple testing, the asymptotic* $100(1-\tau)\%$ *CIs for* $\lambda_{j,t}(\boldsymbol{\theta}_{0,j}) - \lambda_{k,t}(\boldsymbol{\theta}_{0,k})$ *with* $(j,k) \in \mathcal{P}$ *are given by*

$$\lambda_{j,t}(\hat{\boldsymbol{\theta}}_{n,j}) - \lambda_{k,t}(\hat{\boldsymbol{\theta}}_{n,k}) \ \pm \ c_{1-\tau,t} n^{-1/2} \sqrt{\hat{\iota}_{j,k,t}}.$$

For the proof see Appendix G, in particular the computation of critical values $c_{1-\tau,t}$ relies on the max-t method of Westfall and Young (1993), which controls the FWER. While Westfall and Young (1993) implement this approach via permutation resampling

under i.i.d. assumptions, our setting is non-i.i.d. We therefore rely on the conditional joint asymptotic distribution of $\boldsymbol{\lambda}_t(\hat{\boldsymbol{\theta}}_n)$ and recover the asymptotic distribution of $\hat{\boldsymbol{Z}}_t$ under (1.23) by simulation.

Hence, if $\hat{M}_t > c_{1-\tau,t}$, then there exists at least one pair $(j,k) \in \mathcal{P}$ such that $|\hat{Z}_{j,k,t}| > c_{1-\tau,t}$. Thus, under multiple testing, the null hypothesis (1.21) is rejected whenever $|\hat{Z}_{j,k,t}| > c_{1-\tau,t}$. Corollary 1.3 exploits the dependence structure across tests instead of using ad hoc conservative corrections like the Bonferroni method. Since subset pivotality holds for the joint distribution of $\hat{\boldsymbol{Z}}_t$ under the true null hypotheses, the method asymptotically ensures strong control of the FWER at level τ across all $|\mathcal{P}|$ tests.

The corollary below uses the $100(1-\tau)\%$ joint confidence sets for the pairwise differences $(\lambda_{j,t} - \lambda_{k,t})$, $(j,k) \in \mathcal{P}$, derived in Corollary 1.3, to construct $100(1-\tau)\%$ (two-sided) simultaneous joint confidence sets for the ranks of $\lambda_{i,t}(\boldsymbol{\theta}_{0,i})$, with $i \in I$.

Corollary 1.4 (Two-sided confidence set for systemic risk ranks)**.** *Let $C_{n,t}(1-\tau,\mathcal{P})$ denote the two-sided $100(1-\tau)\%$ joint confidence region for $\lambda_{j,t}(\boldsymbol{\theta}_{0,j}) - \lambda_{k,t}(\boldsymbol{\theta}_{0,k})$, $(j,k) \in \mathcal{P}$, defined as the Cartesian product of the marginal intervals $C_{n,t}(1-\tau,\mathcal{P}) \equiv \prod_{(j,k)\in\mathcal{P}} C_{n,t}(1-\tau,\mathcal{P},(j,k))$ where each component is given by $C_{n,t}(1-\tau,\mathcal{P},(j,k)) \equiv \left[\lambda_{j,t}(\hat{\boldsymbol{\theta}}_{n,j}) - \lambda_{k,t}(\hat{\boldsymbol{\theta}}_{n,k}) \;\pm\; c_{1-\tau,t} n^{-1/2} \sqrt{\hat{\iota}_{j,k,t}}\right]$. To construct a confidence set for the rank of institution i define $\pi(i,k) = (\min\{i,k\}, \max\{i,k\})$, $\operatorname{sgn}(i,k) = \begin{cases} +1, & i < k, \\ -1, & i > k, \end{cases}$ and $C_{n,t}^{(i,k)} = \operatorname{sgn}(i,k) C_{n,t}(1-\tau,\mathcal{P},\pi(i,k))$. Then, for each $i \in I$ $\boldsymbol{R}_- \equiv (-\infty, 0)$ and $\boldsymbol{R}_+ \equiv (0,\infty)$,*

$$C_{i,t}^- \;\equiv\; \{\, j \in I \setminus \{i\} :\; C_{n,t}^{(i,j)} \subseteq \boldsymbol{R}_- \,\},$$

$$C_{i,t}^+ \;\equiv\; \{\, j \in I \setminus \{i\} :\; C_{n,t}^{(i,j)} \subseteq \boldsymbol{R}_+ \,\},$$

contain all institutions j whose $\lambda_{j,t}(\hat{\boldsymbol{\theta}}_{n,j})$ are significantly larger, respectively smaller, than that of institution i on day t. The $100(1-\tau)\%$ confidence set for the ranking of institution i is then obtained as $R_{i,t} \equiv \left\{\, |C_{i,t}^-| + 1,\; \ldots,\; N - |C_{i,t}^+| \,\right\}$.

The proof of Corollary 1.4 follows from the results of Corollary 1.3 and Mogstad et al. (2024) Section 3.3.1, where the set $\mathcal{P}$ includes only $N(N-1)/2$ non-redundant pairs, and the rank confidence sets in our setting are time-varying.

All three corollaries of this section provide a basis for assessing whether differences in systemic risk ranks are statistically significant and thus how much information is provided by such a ranking.

1.4 Simulation Experiments

1.4.1 The Role of Time-Varying Dependence

We investigate the finite-sample properties of the proposed MES and ΔCoVaR estimators through Monte Carlo simulations under two alternative dynamic DGPs for $\boldsymbol{H}_t$. First, we

analyze these measures in a BEKK-MGARCH setup that is characterized by dynamic variances and dynamic cross-correlations.[12] Second, we consider a CCC specification with dynamic variances but time-invariant correlation, which serves as a benchmark to assess the behavior of MES and ΔCoVaR estimators under static versus dynamic dependence structures. Complementing these presented results, Appendix H.1 provides the simulation results for the fully time-invariant setting ($\boldsymbol{H}_t \equiv \boldsymbol{H}$ for all t), along with a description of the setup and a discussion of the findings. Furthermore, Appendix H.2 provides simulation evidence and implementation details for the pairwise rank test introduced in Section 1.3.2 in a BEKK setup.

For both DGPs, the vector process of de-meaned returns is generated as $\boldsymbol{y}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t$, where $y_{m,t}$ and $y_{i,t}$ denote the market and company returns observed at time t and $\boldsymbol{\eta}_t$ is an i.i.d. vector error process with $\mathbb{E}(\boldsymbol{\eta}_t) = 0$ and $\mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t') = \boldsymbol{I}_2$. For the distribution of $\boldsymbol{\eta}_t$ we use three alternative choices, the standard normal distribution (labeled by design N) and the normalized Student t distribution with 5 and 10 degrees of freedom, respectively (labeled by design $S5$ and $S10$, respectively). The specification of $\boldsymbol{H}_t$ under the BEKK and CCC models are given in (1.7) and (1.8), respectively. The MGARCH parameters $\boldsymbol{\theta}_0$ are calibrated at their QMLE using daily log-returns of Bank of America (firm return) and the CRSP market value-weighted index (market return) over the period January 2, 2015 to December 31, 2020. For the CCC model, this yields $\boldsymbol{\theta}_0 = (\omega_m, \alpha_m, \beta_m, \omega_i, \alpha_i, \beta_i, \rho_i)' = (0.038,\ 0.212,\ 0.757,\ 0.247,\ 0.145,\ 0.786,\ 0.657)'$. For the BEKK model, $\boldsymbol{\theta}_0 = (\text{vech}(\boldsymbol{C}_0)', \text{diag}(\boldsymbol{A}_0)', \text{diag}(\boldsymbol{B}_0)')' =$(0.183, 0.282, 0.31, 0.3863, 0.341, 0.902, 0.916)′.

For each Monte Carlo replication, we simulate the time series $\{y_{m,t}, y_{i,t}\}_{t=1}^{n+1}$ from one of the two specifications for $\boldsymbol{H}_t$ (BEKK or CCC) combined with one of the three innovation distributions (N, $S5$, or $S10$). Using the first n simulated observations $\{y_{m,t}, y_{i,t}\}_{t=1}^{n}$, we estimate $\boldsymbol{\theta}_0$ by minimizing (1.12). We then compute $\hat{\mu}_{m,\alpha}$ and $\Delta\hat{\xi}_{m,\alpha}$ as in (1.13) and (1.14) using the standardized residuals $\{\hat{\eta}_{m,t}\}_{t=1}^{n}$. Finally, we compute the one-step-ahead conditional volatilities $\sigma_{i,n+1}$ and $\sigma_{m,n+1}$, as well as the conditional correlation $\rho_{i,n+1}$, implied by $\hat{\boldsymbol{\theta}}_n$. Based on these quantities, we then evaluate MES and ΔCoVaR for period $t = n+1$ according to (1.15) and (1.16), i.e., $MES_{i,n+1}(\alpha, \hat{\boldsymbol{\theta}}_n, \hat{\mu}_{m,\alpha})$ and $\Delta CoVaR_{i,n+1}(\alpha, \hat{\boldsymbol{\theta}}_n, \Delta\hat{\xi}_{m,\alpha})$.

The simulation study is conducted under both correct specification and misspecification of the first-step dynamics for $\boldsymbol{H}_t$. Under correct specification, the same MGARCH model is used for data generation and estimation (BEKK-BEKK and CCC-CCC). Under misspecification, data are generated from one specification for $\boldsymbol{H}_t$ and estimated under the other (BEKK-CCC and CCC-BEKK) to assess the impact of first-step misspecification of $\boldsymbol{H}_t$ on the MES and ΔCoVaR estimators. We use sample sizes $n = 500$ and $n = 1000$, and coverage levels $\alpha = 1\%$ and 10%. Each simulation scenario uses 1,000 Monte Carlo replications.

[12] Though the GARCH-DCC model is widely used in the systemic risk literature (see e.g., Brownlees and Engle, 2017; Girardi and Ergün, 2013), formally established asymptotic theory for its QMLE is not yet available.

Table 1.1 reports the corresponding results for $\alpha = 10\%$. The results for $\alpha = 1\%$ are reported in Appendix H.1, Table 1.6. Each column corresponds to one of the quantities $\sigma_{i,n+1}$, $\sigma_{m,n+1}$, $\rho_{i,n+1}$, $\Delta\xi_{m,\alpha}$, $\mu_{m,\alpha}$, $MES_{i,n+1}$, and $\Delta CoVaR_{i,n+1}$. For each quantity, we report the RMSE, the average bias (in parentheses), and the empirical 90% coverage probability of the corresponding estimated asymptotic CI (in brackets).[13] Panel A corresponds to the case where the DGP follows the BEKK specification. Panel B reports results when the DGP is given by the CCC model. Within each panel, subpanel 1) refers to the sample size $n = 500$, and subpanel 2) to $n = 1000$. In each subpanel, the first column indicates the parametric specification imposed for $\boldsymbol{H}_t$. Correct specification occurs when this first-step specification matches the DGP, otherwise, the setting is misspecified.

We start by examining the correctly specified settings. For both BEKK and CCC specifications, increasing the sample size from $n = 500$ to $n = 1000$ leads to a reduction in RMSE for $\mu_{m,\alpha}$ and $\Delta\xi_{m,\alpha}$, as well as for $MES_{i,n+1}$ and $\Delta CoVaR_{i,n+1}$ in all correctly specified cases. For the underlying first-step quantities entering their construction, RMSE decreases with n, except for $\sigma_{m,n+1}$ under the CCC specification with $S5$ and $S10$ innovations. The average bias is negligible across all configurations. The empirical coverage probabilities are close to the nominal 90% level in the correctly specified cases. The results are qualitatively similar under BEKK and CCC when the first-step dynamics are correctly specified.

Results for $\alpha = 1\%$ (reported in Appendix H.1) confirm this pattern. The first-step quantities $\sigma_{i,n+1}$, $\sigma_{m,n+1}$, and $\rho_{i,n+1}$ are unaffected by the choice of α, since their estimation does not depend on the tail probability level used to construct MES and ΔCoVaR. In contrast, the second-step quantities $\mu_{m,\alpha}$ and $\Delta\xi_{m,\alpha}$, and the resulting $MES_{i,n+1}$ and $\Delta CoVaR_{i,n+1}$, exhibit larger RMSE and slightly increased biases at $\alpha = 1\%$. Coverage probabilities deteriorate mildly, but primarily in the most demanding configuration with $n = 500$ and $\alpha = 1\%$. For $n = 1000$, coverage remains close to the nominal level. Overall, these simulation results demonstrate that the asymptotic results of Section 1.3.1 offer reasonable approximations in finite samples, with the approximations improving for larger sample sizes and larger α.

We now turn to misspecified first-step dynamics. RMSEs for $\sigma_{i,n+1}$, $\sigma_{m,n+1}$, and $\rho_{i,n+1}$ are almost uniformly higher than under correct specification. For $n = 500$, finite-sample variability may obscure this pattern, but for $n = 1000$ the increase in RMSE under misspecification is evident. The average bias remains moderate across configurations. Coverage probabilities fall below the nominal 90% level in all misspecified settings. In detail, coverage lies in the 65%-84% range for $\sigma_{i,n+1}$ and $\sigma_{m,n+1}$, but drops to 14%-59% for $\rho_{i,n+1}$. Interestingly, the asymmetry across misspecification directions is particularly marked for $\rho_{i,n+1}$. Coverage lies between 33% and 59% when data are generated under CCC and estimated using BEKK, but falls to 14%-27% when data are generated under BEKK and estimated under CCC. Thus, misspecifying dynamic correlation as static

[13] The two-sided estimated asymptotic CIs for $\Delta CoVaR_{i,n+1}$ and $MES_{i,n+1}$, constructed for $\tau = 0.10$, are shown in 1.19 and 1.20, respectively, and form the basis for the empirical 90% coverage probabilities reported in brackets.

Table 1.1: Simulation results under BEKK and CCC-type MGARCH specifications ($\alpha = 10\%$).

Panel A: Data generating process is a BEKK

Estimated Model	Dist	$\sigma_{i,n+1}$	$\sigma_{m,n+1}$	$\rho_{\mathbf{i,n+1}}$	$\Delta\xi_{m,\alpha}$	$\mu_{m,\alpha}$	$\mathbf{MES_{i,n+1}}$	$\mathbf{\Delta CoVaR_{i,n+1}}$
1) $n = 500$								
BEKK	*N*	0.464(-0.012)[0.90]	0.238(-0.011)[0.90]	**0.037(0.000)[0.90]**	0.067(0.001)[0.91]	0.065(-0.005)[0.90]	**0.157(0.000)[0.91]**	**0.199(-0.008)[0.90]**
CCC	*N*	0.688(0.075)[0.79]	0.222(0.013)[0.81]	**0.212(-0.006)[0.21]**	0.067(-0.002)[0.90]	0.065(-0.004)[0.90]	**0.623(-0.031)[0.35]**	**0.851(-0.044)[0.35]**
BEKK	*S10*	0.724(-0.007)[0.89]	0.216(-0.012)[0.90]	**0.040(-0.001)[0.89]**	0.071(0.004)[0.90]	0.080(-0.010)[0.90]	**0.173(-0.001)[0.89]**	**0.233(-0.021)[0.90]**
CCC	*S10*	0.825(0.082)[0.79]	0.269(0.028)[0.82]	**0.205(-0.012)[0.21]**	0.071(0.002)[0.89]	0.080(-0.010)[0.90]	**0.582(-0.022)[0.36]**	**0.843(-0.046)[0.35]**
BEKK	*S5*	2.578(-0.151)[0.89]	0.938(-0.044)[0.87]	**0.050(0.002)[0.89]**	0.078(0.013)[0.89]	0.100(0.000)[0.90]	**0.187(0.005)[0.90]**	**0.293(-0.016)[0.89]**
CCC	*S5*	2.106(0.016)[0.83]	0.781(0.025)[0.81]	**0.199(0.007)[0.27]**	0.077(0.013)[0.90]	0.098(0.002)[0.89]	**0.611(-0.042)[0.45]**	**0.965(-0.083)[0.45]**
2) $n = 1000$								
BEKK	*N*	0.300(0.007)[0.91]	0.110(-0.006)[0.90]	**0.026(0.000)[0.90]**	0.046(-0.001)[0.92]	0.048(-0.002)[0.88]	**0.108(-0.003)[0.93]**	**0.137(-0.006)[0.91]**
CCC	*N*	0.675(0.102)[0.73]	0.161(0.014)[0.75]	**0.210(-0.005)[0.14]**	0.046(-0.003)[0.91]	0.047(-0.002)[0.89]	**0.617(-0.039)[0.28]**	**0.846(-0.051)[0.25]**
BEKK	*S10*	0.457(-0.002)[0.90]	0.139(-0.009)[0.91]	**0.028(-0.001)[0.90]**	0.049(0.002)[0.90]	0.058(-0.004)[0.90]	**0.114(-0.001)[0.91]**	**0.161(-0.008)[0.91]**
CCC	*S10*	0.706(0.098)[0.78]	0.176(0.020)[0.80]	**0.205(-0.011)[0.15]**	0.049(-0.000)[0.90]	0.057(-0.004)[0.90]	**0.572(-0.026)[0.28]**	**0.831(-0.041)[0.28]**
BEKK	*S5*	1.584(-0.096)[0.90]	0.323(-0.010)[0.89]	**0.034(0.000)[0.90]**	0.056(0.009)[0.89]	0.074(0.002)[0.90]	**0.144(0.008)[0.89]**	**0.238(0.001)[0.89]**
CCC	*S5*	1.488(0.021)[0.82]	0.507(0.057)[0.80]	**0.200(0.005)[0.20]**	0.056(0.009)[0.89]	0.072(0.003)[0.90]	**0.616(-0.036)[0.34]**	**0.956(-0.067)[0.36]**

Panel B: Data generating process is a CCC

Estimated Model	Dist	$\sigma_{i,n+1}$	$\sigma_{m,n+1}$	$\rho_{\mathbf{i,n+1}}$	$\Delta\xi_{m,\alpha}$	$\mu_{m,\alpha}$	$\mathbf{MES_{i,n+1}}$	$\mathbf{\Delta CoVaR_{i,n+1}}$
1) $n = 500$								
CCC	*N*	0.482(-0.009)[0.90]	0.675(-0.026)[0.91]	**0.025(0.000)[0.90]**	0.067(0.000)[0.90]	0.065(-0.004)[0.90]	**0.144(0.002)[0.90]**	**0.178(-0.004)[0.91]**
BEKK	*N*	0.763(0.053)[0.73]	1.055(-0.002)[0.82]	**0.119(0.042)[0.48]**	0.069(-0.001)[0.90]	0.066(-0.003)[0.91]	**0.297(-0.080)[0.68]**	**0.405(-0.112)[0.65]**
CCC	*S10*	0.607(-0.009)[0.89]	0.436(0.003)[0.89]	**0.028(0.000)[0.90]**	0.070(0.005)[0.90]	0.079(-0.009)[0.89]	**0.154(0.003)[0.89]**	**0.210(-0.014)[0.89]**
BEKK	*S10*	1.015(0.084)[0.75]	0.832(-0.007)[0.84]	**0.129(0.051)[0.48]**	0.071(0.002)[0.89]	0.080(-0.011)[0.90]	**0.337(-0.105)[0.64]**	**0.484(-0.167)[0.63]**
CCC	*S5*	2.960(-0.091)[0.89]	1.977(-0.029)[0.88]	**0.034(0.002)[0.89]**	0.078(0.018)[0.90]	0.099(0.006)[0.89]	**0.165(0.011)[0.91]**	**0.256(-0.007)[0.89]**
BEKK	*S5*	3.535(-0.024)[0.77]	2.044(-0.023)[0.84]	**0.120(0.036)[0.59]**	0.078(0.012)[0.89]	0.099(-0.002)[0.89]	**0.316(-0.059)[0.73]**	**0.494(-0.111)[0.73]**
2) $n = 1000$								
CCC	*N*	0.381(0.005)[0.91]	0.397(-0.023)[0.91]	**0.018(-0.000)[0.91]**	0.045(-0.001)[0.91]	0.047(-0.002)[0.89]	**0.099(-0.002)[0.90]**	**0.126(-0.004)[0.90]**
BEKK	*N*	0.744(0.070)[0.65]	0.455(0.013)[0.78]	**0.120(0.044)[0.33]**	0.047(-0.002)[0.91]	0.048(-0.001)[0.89]	**0.272(-0.086)[0.56]**	**0.369(-0.116)[0.54]**
CCC	*S10*	0.448(-0.008)[0.89]	0.707(-0.032)[0.92]	**0.020(-0.000)[0.90]**	0.049(0.002)[0.90]	0.057(-0.003)[0.90]	**0.106(0.002)[0.90]**	**0.149(-0.004)[0.89]**
BEKK	*S10*	0.857(0.084)[0.66]	1.274(-0.021)[0.84]	**0.129(0.052)[0.37]**	0.049(0.000)[0.90]	0.058(-0.006)[0.90]	**0.301(-0.104)[0.57]**	**0.434(-0.157)[0.55]**
CCC	*S5*	2.462(-0.100)[0.90]	2.022(-0.014)[0.89]	**0.025(0.001)[0.90]**	0.056(0.011)[0.89]	0.072(0.005)[0.90]	**0.131(0.011)[0.90]**	**0.204(0.005)[0.90]**
BEKK	*S5*	2.524(-0.017)[0.71]	2.007(-0.003)[0.84]	**0.119(0.036)[0.47]**	0.057(0.006)[0.89]	0.073(-0.002)[0.89]	**0.294(-0.056)[0.65]**	**0.463(-0.099)[0.66]**

Simulation results for MES and ΔCoVaR estimators at $\alpha = 10\%$. Panel A (Panel B) reports results when the DGP follows the BEKK (CCC) specification. Subpanel 1) corresponds to $n = 500$, and subpanel 2) to $n = 1000$. Entries are reported as RMSE(bias)[coverage], where coverage denotes the empirical 90% coverage probability of the estimated asymptotic CI. Correct specification occurs when the first-step MGARCH specification coincides with the DGP, otherwise, the setting is misspecified. Each design is based on 1,000 Monte Carlo replications.

induces larger distortion than the reverse case.

We next examine $\mu_{m,\alpha}$ and $\Delta\xi_{m,\alpha}$ under misspecified $\boldsymbol{H}_t$. In terms of RMSE, bias, and coverage probability, their performance remains broadly comparable to the correctly specified case. Hence, misspecification of $\boldsymbol{H}_t$, including the omission of dynamic correlation, does not materially propagate to the second-step estimators $\hat{\mu}_{m,\alpha}$ and $\Delta\hat{\xi}_{m,\alpha}$. This conclusion, however, does not extend to $MES_{i,n+1}$ and $\Delta CoVaR_{i,n+1}$. These measures depend directly on the first-step quantities $\sigma_{i,n+1}$ and $\rho_{i,n+1}$, and therefore inherit the misspecification affecting these components. Accordingly, RMSE increases under misspecification, while average bias remains moderate. Coverage deteriorates substantially, particularly in the BEKK-CCC case, mirroring the severe degradation observed for the correlation parameter. Overall, allowing for dynamic correlation through a BEKK-type specification provides greater robustness for systemic risk measurement under misspecification.

1.4.2 Alternative SRM Specifications

In this subsection, we investigate the robustness of our previous simulation results with respect to a different inequality conditioning set in the SRM specification. In our framework, CoVaR is defined conditionally on the restriction $y_{m,t} = VaR_{m,t}(\alpha)$ according to Adrian and Brunnermeier (2016) (see Eq. (1.2)) and in line with the extensive finance literature. Alternatively, however, CoVaR and ΔCoVaR can be defined conditionally on the inequality $y_{m,t} \leq VaR_{m,t}(\alpha)$, yielding SRMs that we denote as $CoVaR^{\leq}$ and $\Delta CoVaR^{\leq}$, respectively. In particular, for the time-invariant CCC type benchmark of Section 1.4.1 the statistical properties of SRMs in the $\Delta CoVaR^{\leq}$ setting have actually been derived by Francq and Zakoïan (2025) (thereafter FZ).

To make the different settings comparable, we adapt our estimator according to the SRM framework of FZ. We therefore define the version $CoVaR^{\leq}_{i,t}(\beta,\alpha)$ of our estimator through $\Pr(y_{i,t} \leq CoVaR^{\leq}_{i,t}(\beta,\alpha) | y_{m,t} \leq VaR_{m,t}(\alpha); \mathcal{F}_{t-1}) = \beta$. Accordingly, CoVaR conditionally on the financial system being in its median state, denoted by $CoVaR^{\leq}_{i,t}(\beta,\alpha_{\text{inf}},\alpha_{\text{sup}})$, is defined through $\Pr(y_{i,t} \leq CoVaR^{\leq}_{i,t}(\beta,\alpha_{\text{inf}},\alpha_{\text{sup}}) | VaR_{m,t}(\alpha_{\text{inf}}) \leq y_{m,t} \leq VaR_{m,t}(\alpha_{\text{sup}}); \mathcal{F}_{t-1}) = \beta$, where the probabilities α_{inf} and α_{sup} with $\alpha < \alpha_{\text{inf}} < \alpha_{\text{sup}}$ are around 0.5 representing a median state. Consequently, the ΔCoVaR of institution i at time t is defined as $\Delta CoVaR^{\leq}_{i,t}(\beta,\alpha,\alpha_{\text{inf}},\alpha_{\text{sup}}) = CoVaR^{\leq}_{i,t}(\beta,\alpha) - CoVaR^{\leq}_{i,t}(\beta,\alpha_{\text{inf}},\alpha_{\text{sup}})$.

Setting $\tilde{\eta}_{i,t}(\boldsymbol{\theta}_0) \equiv \frac{y_{i,t}}{\sigma_{i,t}(\boldsymbol{\theta}_0)} = \rho_{i,t}(\boldsymbol{\theta}_0)\,\eta_{m,t} + \sqrt{1-\rho^2_{i,t}(\boldsymbol{\theta}_0)}\,\eta_{i,t}$ using equations (1.4) and (1.9), we can show that the multiplicative decomposition of ΔCoVaR in Proposition 1.2 also extends to $\Delta CoVaR^{\leq}_{i,t}(\beta,\alpha,\alpha_{\text{inf}},\alpha_{\text{sup}},\boldsymbol{\theta}_0)$ as

$$\Delta CoVaR^{\leq}_{i,t}(\beta,\alpha,\alpha_{\text{inf}},\alpha_{\text{sup}},\boldsymbol{\theta}_0) = \sigma_{i,t}(\boldsymbol{\theta}_0)\left(q_{\beta,\alpha,t}(\boldsymbol{\theta}_0) - q_{\beta,\alpha_{\text{inf}},\alpha_{\text{sup}},t}(\boldsymbol{\theta}_0)\right), \tag{1.25}$$

where $q_{\beta,\alpha,t}(\boldsymbol{\theta}_0)$ and $q_{\beta,\alpha_{\text{inf}},\alpha_{\text{sup}},t}(\boldsymbol{\theta}_0)$ denote the β-quantiles of the (conditional) distributions of $\tilde{\eta}_{i,t}(\boldsymbol{\theta}_0) \mid \eta_{m,t} \leq \xi_{m,\alpha}$ and $\tilde{\eta}_{i,t}(\boldsymbol{\theta}_0) \mid \xi_{m,\alpha_{\text{inf}}} \leq \eta_{m,t} \leq \xi_{m,\alpha_{\text{sup}}}$, conditionally on $\mathcal{F}_{t-1}$, respectively. The derivation of (1.25) follows along the lines of the proof of Proposition 1.2 accounting for the different conditioning set. The semi-parametric implementation of

the resulting two-step estimator is provided in Appendix D.1. Note that (1.25) provides an extension of Proposition 2.1 in FZ, where the resulting $\Delta CoVaR^{\leq}$ is dynamic not only through the firm's volatility but also through the correlation, as captured by the (time-varying) difference in truncated quantiles $q_{\beta,\alpha,t}(\boldsymbol{\theta}_0) - q_{\beta,\alpha_{\text{inf}},\alpha_{\text{sup}},t}(\boldsymbol{\theta}_0)$.

We re-use the same DGPs as in Section 1.4.1. In the first setting, $\boldsymbol{y}_t$ is generated from the BEKK specification in (1.7), with the parameter calibration reported in Section 1.4.1. In the second setting, $\boldsymbol{y}_t$ follows the CCC model in (1.8), together with the univariate GARCH(1,1) variance dynamics using the parameter calibration as in Section 1.4.1. In each replication, we generate $\{y_{i,t}, y_{m,t}\}_{t=1}^{n+1}$ from both DGPs, utilize the first n observations to estimate the model parameters and compute (i) $\Delta CoVaR^{\leq}$ as of (1.25) and (ii) the CoVaR metric according to FZ based on equation-by-equation GARCH models[14], denoted by $\Delta\text{CoVaR}^{\leq}$(FZ), at period $t = n + 1$. We set $\beta = 10\%$, $\alpha = 20\%$, $\alpha_{\text{inf}} = 25\%$ and $\alpha_{\text{sup}} = 75\%$, in line with the simulation study in FZ. We repeat the experiment 1,000 times for sample sizes $n \in 500, 1000, 2500$ and for the three choices of distributions of $\boldsymbol{\eta}_t$ N, $S10$ and $S5$.

Table 1.2: Average bias and RMSE for $\Delta\text{CoVaR}^{\leq}$ estimators.

Panel A: Data generating process is a BEKK

Estimator	N (bias)	N (RMSE)	$S10$ (bias)	$S10$ (RMSE)	$S5$ (bias)	$S5$ (RMSE)
1) $n = 500$						
$\Delta\text{CoVaR}^{\leq}$	-0.0140	0.3227	-0.0160	0.3699	-0.0166	0.4738
$\Delta\text{CoVaR}^{\leq}$(FZ)	-0.0030	0.7097	0.0147	0.7746	-0.0385	0.8934
2) $n = 1000$						
$\Delta\text{CoVaR}^{\leq}$	-0.0156	0.2286	0.0025	0.2609	0.0047	0.3373
$\Delta\text{CoVaR}^{\leq}$(FZ)	-0.0096	0.6972	0.0319	0.7537	-0.0189	0.8875
3) $n = 2500$						
$\Delta\text{CoVaR}^{\leq}$	-0.0130	0.1518	-0.0075	0.1716	0.0017	0.1992
$\Delta\text{CoVaR}^{\leq}$(FZ)	-0.0051	0.6871	0.0226	0.7356	-0.0118	0.8688

Panel B: Data generating process is a CCC

Estimator	N (bias)	N (RMSE)	$S10$ (bias)	$S10$ (RMSE)	$S5$ (bias)	$S5$ (RMSE)
1) $n = 500$						
$\Delta\text{CoVaR}^{\leq}$	-0.0026	0.3098	0.0045	0.3530	0.0052	0.4491
$\Delta\text{CoVaR}^{\leq}$(FZ)	-0.0133	0.3087	-0.0077	0.3554	-0.0091	0.4502
2) $n = 1000$						
$\Delta\text{CoVaR}^{\leq}$	-0.0123	0.2231	0.0033	0.2518	0.0109	0.3219
$\Delta\text{CoVaR}^{\leq}$(FZ)	-0.0196	0.2247	-0.0020	0.2558	0.0062	0.3224
3) $n = 2500$						
$\Delta\text{CoVaR}^{\leq}$	-0.0109	0.1476	-0.0035	0.1627	0.0070	0.2002
$\Delta\text{CoVaR}^{\leq}$(FZ)	-0.0137	0.1503	-0.0054	0.1628	0.0050	0.2006

N, $S10$, and $S5$ denote Normal, Student-t with $\nu = 10$, and Student-t with $\nu = 5$ innovations, respectively. Panel A (Panel B) reports results when the DGP follows the BEKK (CCC) specification. Subpanels 1)–3) correspond to sample sizes $n = 500$, $n = 1000$, and $n = 2500$, respectively. Results are reported for the $\Delta\text{CoVaR}^{\leq}$ and $\Delta\text{CoVaR}^{\leq}$(FZ) estimators. For each design, we report the bias and RMSE, based on 1,000 Monte Carlo replications.

[14] See Francq and Zakoïan (2025), Eqs. (3) and (13).

Table 1.2 reports the average bias and RMSE of one-step-ahead predictions of $t = n+1$. In Panel A, the DGP originates from the BEKK model, where alternatively $\Delta\text{CoVaR}^{\leq}$ and $\Delta\text{CoVaR}^{\leq}(\text{FZ})$ are computed. The simulation results show that under this DGP $\Delta\text{CoVaR}^{\leq}$ yields a comparable bias but a substantially lower RMSE than $\Delta\text{CoVaR}^{\leq}(\text{FZ})$. This is true across all sample sizes and innovation distributions.

In Panel B, the data originates from the CCC model, under which $\Delta\text{CoVaR}^{\leq}(\text{FZ})$ is nested within our framework (see Appendix D.1 for details). In this case, both estimators deliver similar biases and RMSEs. They do not, however, coincide exactly, as the methods used to estimate the parameters of the CCC model differ.[15] Hence, allowing for dynamic cross-dependence in a MGARCH framework improves estimation accuracy in more general settings without losing estimation performance under more simple conditions.

1.5 Empirical Application

We analyze a panel of 50 large US financial institutions. Tickers and company names are provided in Appendix E. The dataset spans from January 2, 2008, to December 31, 2020. For each firm $i = 1, \ldots, 50$, we compute daily logarithmic returns $y_{i,t}$ using data from the CRSP (Center for Research in Security Prices) database. The market return $y_{m,t}$ is the daily logarithmic return of the CRSP market value-weighted index.

We employ the diagonal BEKK specification for $\boldsymbol{H}_t$ in (1.7). Parameters ($\boldsymbol{\theta}_0$, $\mu_{m,\alpha}$ and $\Delta\xi_{m,\alpha}$) are estimated using a rolling window of $n = 500$ daily observations, with the estimation window ending at the last trading day of every month. Then, for each day $t = n+1, n+2, \ldots$, of the subsequent month, we compute one-step-ahead forecasts of $\text{MES}_{i,t}\left(\alpha; \hat{\boldsymbol{\theta}}_n, \hat{\mu}_{m,\alpha}\right)$, and $\Delta\text{CoVaR}_{i,t}(\alpha; \hat{\boldsymbol{\theta}}_n, \Delta\hat{\xi}_{m,\alpha})$ for $\alpha = 0.1$, conditionally on the information sets $\mathcal{F}_n, \mathcal{F}_{n+1}, \ldots$, respectively, while keeping the parameter estimates originating from the most recent rolling window (ending at the end of the most recent month). This procedure is repeated each month until the end of the sample period, yielding a time series of MES and ΔCoVaR forecasts for the 50 institutions from January 3, 2010, to December 31, 2020. We compute the corresponding CIs for MES and ΔCoVaR for these institutions for $\tau = 0.1$, based on 1.19 and 1.20. We report MES and ΔCoVaR forecasts with opposite signs to associate higher values with larger systemic risk contributions.

Furthermore, we perform a sensitivity analysis of ΔCoVaR with respect to the market conditioning scheme ($y_{m,t} = VaR_{m,t}(\alpha)$ vs. $y_{m,t} \leq VaR_{m,t}(\alpha)$) and the specification of firm-market dependence. We display $\Delta\text{CoVaR}^{\leq}(\text{FZ})$ (Francq and Zakoïan, 2025), which is defined under inequality conditioning and static dependence (see Section 1.4.2). We set $\beta = \alpha = 10\%$, $\alpha_{\text{inf}} = 25\%$, and $\alpha_{\text{sup}} = 75\%$, in line with their paper. Furthermore, we

[15] Our approach relies on a joint MGARCH estimation based on the minimization of (1.12), whereas $\Delta\text{CoVaR}^{\leq}(\text{FZ})$ relies on an equation-by-equation GARCH estimation, with cross-sectional dependence being recovered from empirical residuals. Additional simulations (not reported) show that both methods coincide pointwise when $\Delta\text{CoVaR}^{\leq}$ is estimated using an equation-by-equation GARCH specification with the correlation estimated from empirical residuals.

compute ΔCoVaR (under equality conditioning) assuming a constant conditional correlation (CCC) structure for $\boldsymbol{H}_t$ (Bollerslev, 1990), which nests our estimation framework, as all parameters of $\boldsymbol{H}_t$ are jointly estimated by minimizing (1.12). Conditional variances follow GARCH(1,1) dynamics. This specification is hereafter denoted ΔCoVaR(CCC).

In summary, MES, ΔCoVaR, and ΔCoVaR(CCC) are the measures developed in this paper (Propositions 1.1 and 1.2), with the difference between ΔCoVaR and ΔCoVaR(CCC) due to the use of a BEKK or CCC dependence specification for $\boldsymbol{H}_t$, while $\Delta\text{CoVaR}^{\leq}$(FZ) refers to the inequality-conditioned measure with static dependence of Francq and Zakoïan (2025).

1.5.1 Forecasting Dynamic SRMs and Their Uncertainty

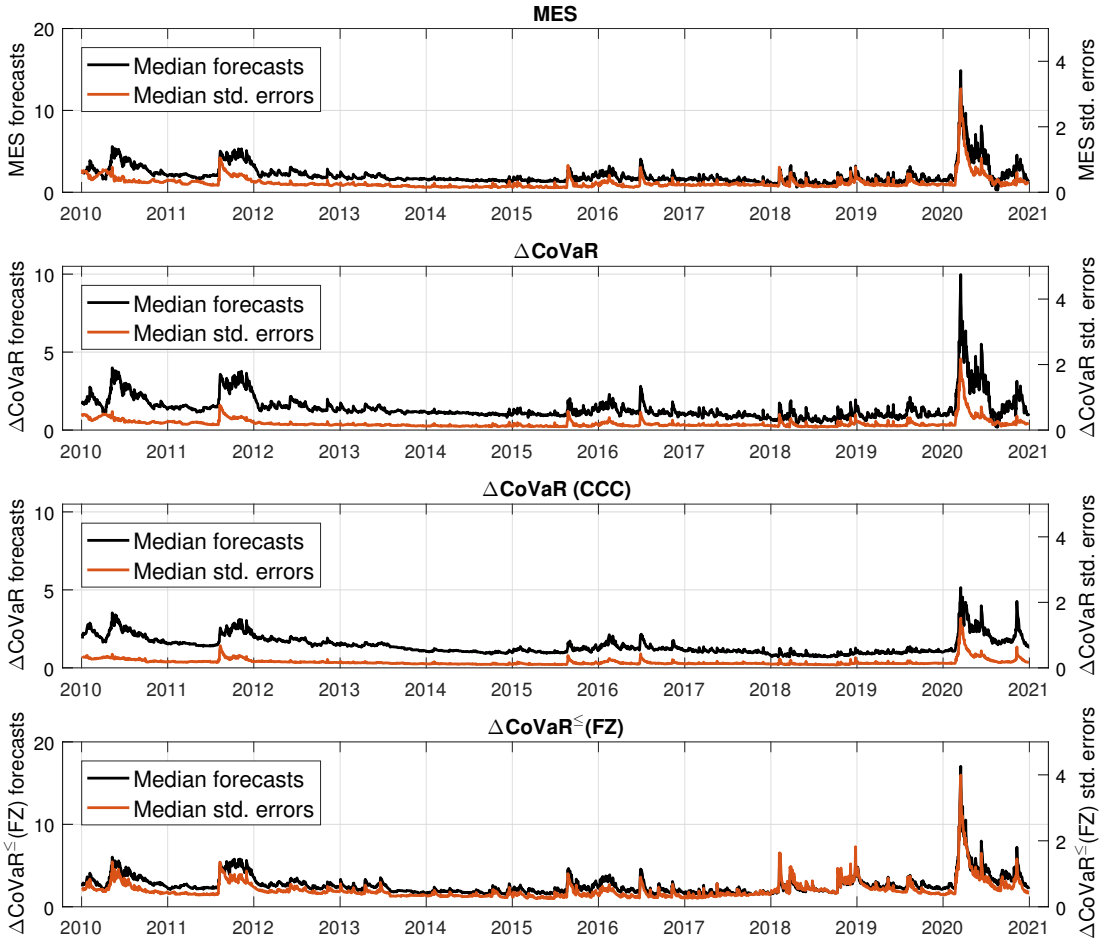
To examine the system level of time series dynamics of systemic risk forecasts and their uncertainty, we report the evolution of the pointwise medians of respective forecasts and estimated standard errors across all 50 institutions for the metrics MES, ΔCoVaR, ΔCoVaR(CCC), and $\Delta\text{CoVaR}^{\leq}$(FZ) (see Figure 1.1). In Appendix I, we also provide a graphical comparison between the CoVaR estimator of White et al. (2015) and our CoVaR estimator defined in Appendix C.3 focusing on point predictions of the CoVaR pre-step.[16]

Two observations emerge from Figure 1.1. First, the evolution of cross-sectional median values of MES, ΔCoVaR, ΔCoVaR (CCC), and $\Delta\text{CoVaR}^{\leq}$(FZ) exhibits similar patterns, rising during periods of market instability - such as the European sovereign debt crisis (2011-2012) and the COVID-19 pandemic (2020). Note that systemic risk responses differ in magnitude across specifications, with increases in ΔCoVaR being more pronounced than in ΔCoVaR(CCC), reflecting the greater responsiveness induced by time-varying correlations. The robustness of our approach for SRM point predictions is also confirmed by the comparison in Appendix I. Second, rising systemic risk forecasts imply roughly proportional increases in estimation errors. This is also the case for the FZ-benchmark, but notable differences occur in shape during the years 2018-2020 and in scale after 2020. This suggests that on average across institutions, high systemic risk forecasts coincide with high uncertainty across time. Note, however, that due to the roughly proportional increase of uncertainty, *relative* precision of predictions does not substantially deteriorate during periods of financial distress.

To assess how static versus dynamic dependence structures affect ΔCoVaR forecasts, we compare estimates obtained under constant (CCC) and time-varying (BEKK) dependence specifications across the 50 banks, referred as ΔCoVaR(CCC) and ΔCoVaR, respectively.

[16] Note that for the CoVaR estimator of White et al. (2015) there are only recent predictive asymptotic results available (Dimitriadis and Hoga, 2024). For an adequate formal evaluation in their case, however, joint elicitability of VaR and CoVaR via scores or tests requires specific forms of these criteria and the assumption that all competing CoVaR forecasts to have a similar VaR-level. Also moving from CoVaR to ΔCoVaR, is in their case non-trivial and requires additional modeling choices. Therefore, we decided to only provide a graphical comparison on the level of the CoVaR pre-step for studying its time-series robustness.

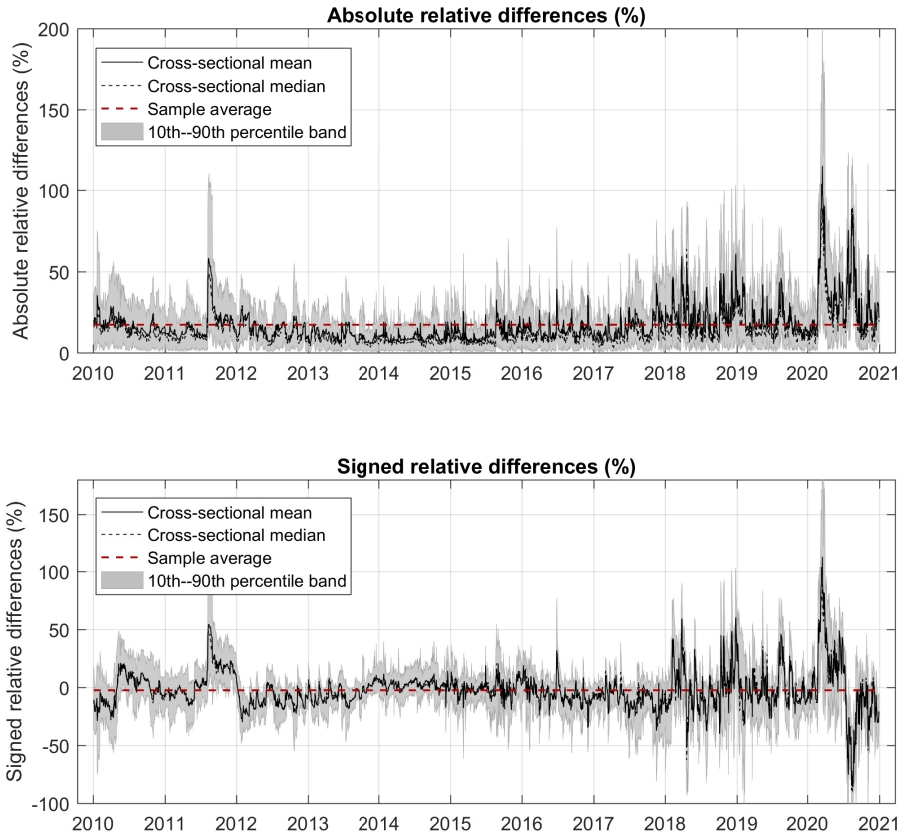
Figure 1.1: Cross-sectional medians of forecasts and asymptotic standard errors.



Medians of MES, ΔCoVaR, ΔCoVaR (CCC), and $\Delta\text{CoVaR}^{\leq}$(FZ) forecasts (black line), computed at each period t as the median across all 50 banks, along with the median of their asymptotic standard errors (orange line). Underlying estimates based on rolling windows of size 500, updated on a monthly basis. The estimated pointwise asymptotic standard errors for MES, ΔCoVaR and ΔCoVaR(CCC) are computed for each institution as of Corollary 1.1, with feasible counterparts shown in Appendix C.6. For the computation of asymptotic standard errors of $\Delta\text{CoVaR}^{\leq}$(FZ), we refer readers to Section G of the online supplemental material in Francq and Zakoïan (2025).

Figure 1.2 shows absolute (upper panel) and signed (lower panel) relative differences between ΔCoVaR and ΔCoVaR(CCC). The absolute deviation between the two specifications amounts to 17% on average and the width of the 10th-90th percentile band shows sizeable dispersion across banks. The signed relative differences show large fluctuations in both magnitude and sign. During the European sovereign debt crisis in 2011 and the COVID-19 episode in 2020, the allowance for dynamic dependencies yields larger SRM forecasts on average. These episodes coincide with the largest differences between

Figure 1.2: ΔCoVaR forecasts under static and dynamic dependence specifications.



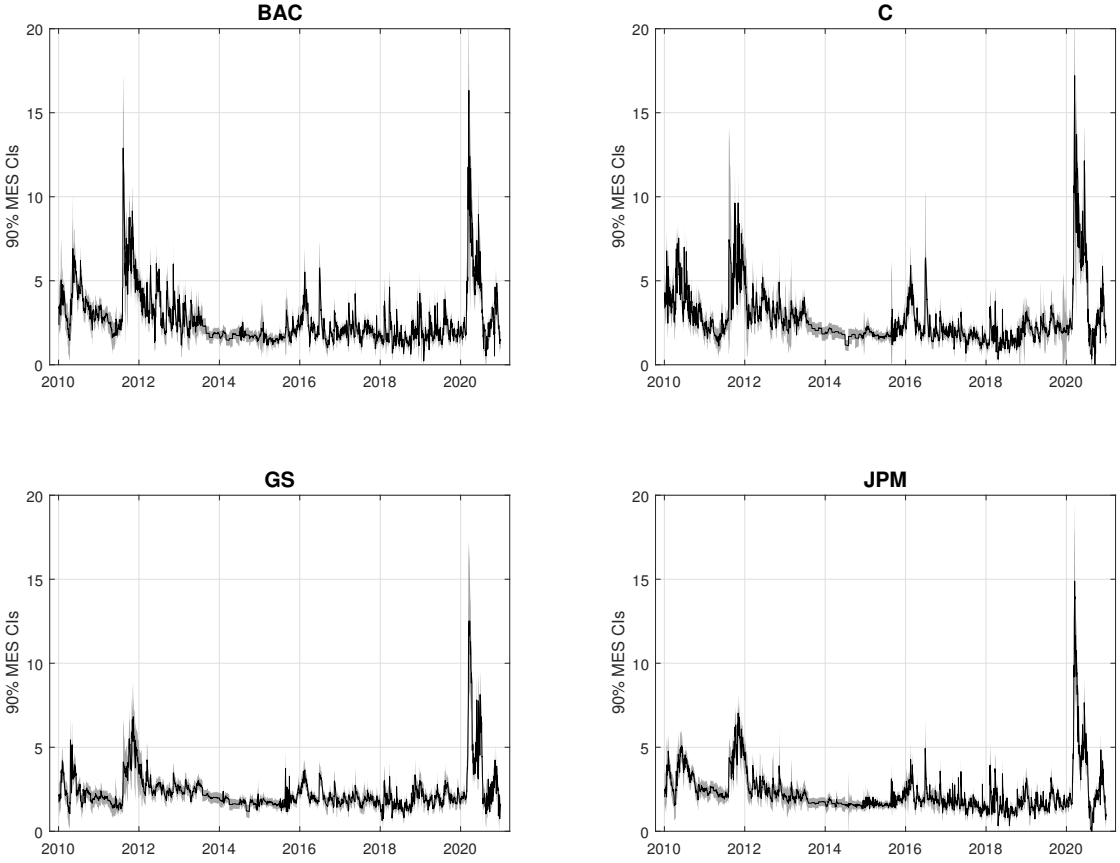
For the 50 banks, we depict the cross-sectional absolute relative difference (in %) of ΔCoVaR and ΔCoVaR(CCC) as $\frac{1}{50}\sum_{i=1}^{50}(|\Delta\text{CoVaR}_{i,t} - \Delta\text{CoVaR}_{i,t}(\text{CCC})|)/|\Delta\text{CoVaR}_{i,t}(\text{CCC})|$ in the upper panel, and their direction using the signed relative difference, $\frac{1}{50}\sum_{i=1}^{50}(\Delta\text{CoVaR}_{i,t} - \Delta\text{CoVaR}_{i,t}(\text{CCC}))/\Delta\text{CoVaR}_{i,t}(\text{CCC})$ in the lower panel, where for notational simplicity, estimated forecasts for institution i at time t are written as $\Delta\text{CoVaR}_{i,t}$ and $\Delta\text{CoVaR}_{i,t}(\text{CCC})$. The solid black line denotes the daily cross-sectional mean as defined above, while the dashed black line denotes the corresponding median (i.e., the 50th percentile across institutions). The shaded area represents the 10th-90th percentile range. The dashed red line marks the overall average. Positive signed differences indicate higher ΔCoVaR forecasts under the BEKK specification relative to the CCC benchmark.

both measures, illustrating the greater responsiveness of time-varying dependence during periods of financial stress.

To illustrate systemic risk dynamics of individual institutions, we consider four representative banks: Bank of America (BAC), Citigroup (C), Goldman Sachs (GS), and JP Morgan Chase (JPM). These companies are recognized as the top US G-SIBs (Global Systemically Important Banks) in the Financial Stability Board's systemic ranking (2022).

As already depicted in the aggregate dynamics in Figure 1.1, also on the firm-level the dynamics of MES and ΔCoVaR are closely aligned (see Figure 1.9 in the Appendix F) as both measures follow the same underlying structure determined by the product of $\sigma_{i,t}$ and $\rho_{i,t}$. However, their scales differ. While MES is influenced by the multiplicative constant $\mu_{m,\alpha}$, ΔCoVaR depends on $\xi_{m,\alpha} - \xi_{m,0.5}$. Again, both measures increase (in absolute terms) during periods of financial instability, such as the European debt crisis (2011-2012) and the COVID-19 pandemic (2020) reflecting higher capital shortfalls during major crises.

Figure 1.3: MES forecasts and 90% confidence intervals.



Daily MES forecasts (black line) and the corresponding 90% CIs (grey shaded area) for BAC, C, GS, and JPM. Parameters are estimated using a rolling window of size 500 and are updated on a monthly basis. The estimated CIs are obtained from (1.20) and feasible quantities are shown in Appendix C.6.

Figure 1.3 presents corresponding MES forecasts (solid black lines) along with their corresponding estimated 90% asymptotic CIs (gray shaded areas). The CIs are relatively tight over time for $n = 500$, suggesting that estimation errors are comparably low. However, when comparing the CIs across the four panels, frequent overlaps occur. This

overlap complicates identifying which institution is more systemic than another. This motivates statistical inference as derived in Section 1.3.2 and empirically implemented in Section 1.5.3. Similar findings hold for ΔCoVaR with corresponding results provided in Figure 1.10 in Appendix F.

1.5.2 Testing for Time-Varying Correlations

A key advantage of MES and ΔCoVaR estimators relying on MGARCH frameworks as discussed in Section 1.2.1 is to allow for the possibility of time-varying correlations across individual return series. To evaluate the importance of this flexibility in real applications, we formally test for constant correlations between market returns and bank-specific returns employing the Lagrange Multiplier (LM) test of Tse (2000). This allows to test for constant correlations through the hypothesis $H_0: \delta_i = 0$ in

$$\rho_{i,t} = \rho_i + \delta_i \, y_{i,t-1} y_{m,t-1}. \tag{1.26}$$

Consider the parameters $\boldsymbol{\theta}_{\sigma_m}$, $\boldsymbol{\theta}_{\sigma_i}$, and $\boldsymbol{\theta}_{\rho_i}$ associated with $\sigma_{m,t}$, $\sigma_{i,t}$, and $\rho_{i,t}$, respectively, with $\boldsymbol{\theta}_{\rho_i} = (\rho_i, \delta_i)$ collecting the parameters of (1.26). Then, $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\sigma_m}, \boldsymbol{\theta}_{\sigma_i}, \boldsymbol{\theta}_{\rho_i})'$ with $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ and the log-likelihood as of (1.12) can be formulated as $L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \ell_t(\boldsymbol{\theta})$ with $\ell_t(\boldsymbol{\theta}) = \boldsymbol{y}_t' \boldsymbol{H}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t + \log|\boldsymbol{H}_t(\boldsymbol{\theta})|$, where $\boldsymbol{H}_t = \boldsymbol{D}_t \, \boldsymbol{R}_t \, \boldsymbol{D}_t$, $\boldsymbol{D}_t = \mathrm{diag}(\sigma_{m,t}, \sigma_{i,t})$ and $\boldsymbol{R}_t$ denotes the (2×2) correlation matrix with off-diagonal element $\rho_{i,t}$. Minimizing $L_n(\boldsymbol{\theta})$ under the restriction $\delta_i = 0$ yields the restricted estimator $\hat{\boldsymbol{\theta}}_{n,c} = \big(\hat{\boldsymbol{\theta}}_{\sigma_m}, \hat{\boldsymbol{\theta}}_{\sigma_i}, \hat{\rho}_i, 0\big)'$. Denote by $\hat{\boldsymbol{S}}$ the $n \times p$ matrix whose rows equal the (estimated) partial derivatives $\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{n,c}}$, for $t = 1, ..., n$ and define $\hat{\boldsymbol{s}} = \frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{n,c}}$. Then, the LM statistic is computed as

$$\mathrm{LM} = \hat{\boldsymbol{s}}' (\hat{\boldsymbol{S}}' \hat{\boldsymbol{S}})^{-1} \hat{\boldsymbol{s}}, \tag{1.27}$$

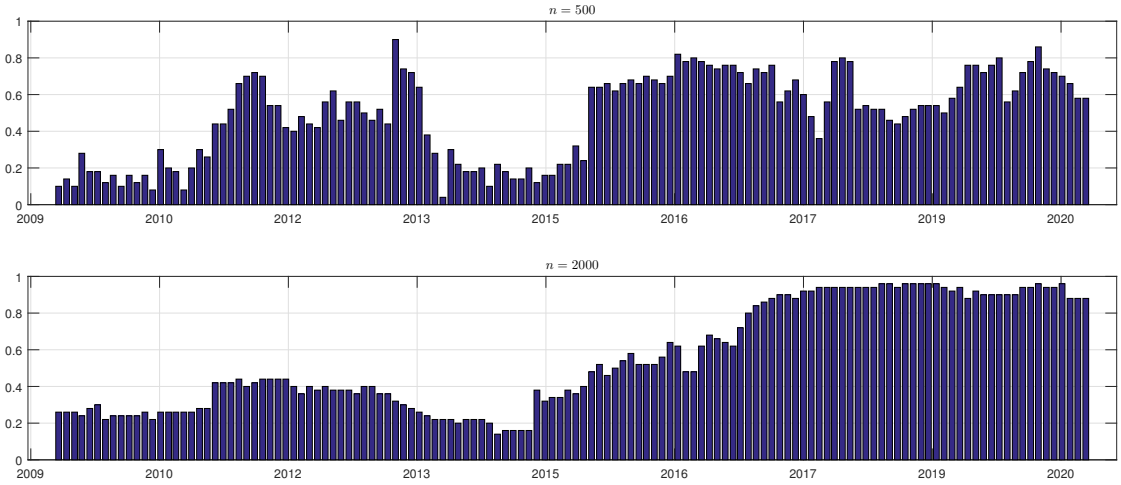
and under H_0 is asymptotically chi-square distributed with one degree of freedom.

We apply the test to each sample of market-bank pairs $\{y_{m,t}, y_{i,t}\}_{t=1}^{n}$, $i = 1, \ldots, 50$, where we opt for GARCH(1,1) equations for the conditional variances with $\sigma_{m,t}^2 = w_m + a_m \sigma_{m,t-1}^2 + b_m y_{m,t-1}^2$ and $\sigma_{i,t}^2 = w_i + a_i \sigma_{i,t-1}^2 + b_i y_{i,t-1}^2$, with $\boldsymbol{\theta}_{\sigma_m} = (w_m, a_m, b_m)$ and $\boldsymbol{\theta}_{\sigma_i} = (w_i, a_i, b_i)$ collecting the GARCH parameters of bank i and the market m, respectively. This setting implies testing the adequacy of the CCC model of Bollerslev (1990) against time-varying correlation structure according to (1.26) with $\delta_i \neq 0$.

Since our setting requires to jointly test $N = 50$ hypotheses (one per bank), we apply the Bonferroni correction. Let τ denote the significance level and $p_{i,k}$ the p-value of the LM test for the pair of assets (m, i) associated with estimation window k. Here, $i = 1, \ldots, N$, and $k = 1, \ldots, K = 132$ indexes the rolling estimation windows of size n, which end month by month from December 2009 to November 2020. Then, the rejection rate, measuring the fraction of institutions for which the null hypothesis of constant correlation with the market is rejected in window k, is given by $\frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(p_{i,k} < \tau/N)$.

Figure 1.4 shows these rejection rates based on estimation windows of size $n = 500$ (top panel) and $n = 2000$ (bottom panel). On average, the null of constant correlation

Figure 1.4: Rejections of constant correlations on the 5% significance level.



Rejection rates of the LM test on the 5% significance level across the sample of all 50 banks. Underlying estimates of $\hat{\boldsymbol{\theta}}_{n,c}$ based on rolling windows of size $n = 500$ (top panel) or $n = 2000$ observations (bottom panel), updated on a monthly basis. We apply a Bonferroni correction for multiple testing.

with the market is rejected in 49% or 55% of all cases, respectively. This confirms that for a large fraction of banks, correlations with the market are time-varying. In particular, for sample sizes of $n = 2000$, rejection rates reach almost 100% after 2017 highlighting the importance of accounting for time-varying correlations in SRMs.

1.5.3 Systemic Risk Rankings and Statistical Inference

Average Rankings over Sub-Periods

We rank U.S. financial institutions according to their systemic risk contributions measured by the time-averaged MES and ΔCoVaR. The time-aggregated measures are defined as $\overline{M}_i = \frac{1}{T}\sum_{t=1}^{T} MES_{i,t}(\alpha)$, $\overline{\Delta C}_i = \frac{1}{T}\sum_{t=1}^{T} \Delta CoVaR_{i,t}(\alpha)$, $\overline{\Delta C}_i(\text{CCC}) = \frac{1}{T}\sum_{t=1}^{T} \Delta CoVaR_{i,t}(\text{CCC})(\alpha)$, and $\overline{\Delta C}_i^{\leq}(\text{FZ}) = \frac{1}{T}\sum_{t=1}^{T} \Delta CoVaR_{i,t}^{\leq}(\text{FZ})(\beta, \alpha, \alpha_{\text{inf}}, \alpha_{\text{sup}})$, where $\overline{M}_i$, $\overline{\Delta C}_i$, $\overline{\Delta C}_i^{\leq}(\text{FZ})$, $\overline{\Delta C}_i(\text{CCC})$ denote the time-series averages of MES, ΔCoVaR, ΔCoVaR$^{\leq}$(FZ), and ΔCoVaR(CCC), respectively, and T is the number of trading days used for averaging. We report the time-averaged volatility $\overline{\sigma}_i = \frac{1}{T}\sum_{t=1}^{T} \sigma_{i,t}$, and the time-averaged correlation with market returns $\overline{\rho}_i = \frac{1}{T}\sum_{t=1}^{T} \rho_{i,t}$ as key components in SRMs (see (1.10) and (1.11). Averages are computed over two sub-periods: a low-risk period (January 2012 to December 2019) and a high-risk period (January 2020 to December 2020), as suggested by systemic risk levels in Figure 1.1.

Table 1.3 provides the rankings of the corresponding metrics for the 50 banks and the two periods. Institutions are ordered according to $\overline{\Delta C}_i$. For $\overline{M}_i$, $\overline{\Delta C}_i^{\leq}(\text{FZ})$, $\overline{\Delta C}_i(\text{CCC})$,

Table 1.3: Systemic risk rankings of financial institutions.

	Period Jan. 2012 to Dec. 2019							Period Jan. 2020 to Dec. 2020						
Tickers	$\overline{\Delta CI}_i$ (r)	$\overline{M}_i$ (r)	$\overline{\rho}_i$; $\overline{\sigma}_i$	$\overline{\Delta CI}_i^{\leq}$(FZ) (r)	$\overline{\Delta CI}_i$(CCC) (r)	$\tilde{r}^{\Delta CI_i}$	$\tilde{r}^{M_i}$	$\overline{\Delta CI}_i$ (r)	$\overline{M}_i$ (r)	$\overline{\rho}_i$; $\overline{\sigma}_i$	$\overline{\Delta CI}_i^{\leq}$(FZ) (r)	$\overline{\Delta CI}_i$(CCC) (r)	$\tilde{r}^{\Delta CI_i}$	$\tilde{r}^{M_i}$
LNC	1.737 (1)	2.569 (1)	0.71 ; 1.89	3.414 (1)	1.776 (1)	1	1	4.008 (1)	5.962 (1)	0.64 ; 4.73	8.407 (1)	3.664 (1)	1	1
MS	1.643 (2)	2.438 (2)	0.70 ; 1.85	2.889 (7)	1.665 (3)	2	2	2.549 (16)	3.775 (17)	0.71 ; 2.70	4.765 (24)	2.341 (16)	10	10
MBI	1.592 (3)	2.324 (3)	0.41 ; 3.05	3.090 (4)	1.707 (2)	4	4	1.780 (42)	2.622 (42)	0.38 ; 3.45	3.646 (38)	1.986 (34)	38	38
SCHW	1.545 (4)	2.286 (4)	0.66 ; 1.79	3.120 (3)	1.608 (4)	3	3	2.056 (36)	3.034 (35)	0.55 ; 2.72	4.695 (27)	2.062 (32)	32	32
C	1.498 (5)	2.215 (5)	0.70 ; 1.68	2.765 (9)	1.544 (5)	5	5	3.047 (6)	4.566 (4)	0.63 ; 3.67	7.155 (2)	3.214 (2)	4	4
BAC	1.459 (6)	2.164 (6)	0.65 ; 1.77	2.692 (12)	1.497 (7)	7	7	2.730 (12)	4.051 (12)	0.67 ; 3.11	6.479 (4)	2.685 (11)	10	10
BEN	1.438 (7)	2.128 (7)	0.74 ; 1.54	3.080 (5)	1.502 (6)	5	5	2.254 (27)	3.320 (27)	0.60 ; 2.82	3.911 (35)	2.278 (21)	21	21
RF	1.409 (8)	2.085 (8)	0.61 ; 1.80	3.173 (2)	1.485 (8)	10	10	3.108 (3)	4.567 (3)	0.60 ; 3.95	6.398 (5)	2.911 (5)	3	3
BLK	1.391 (9)	2.065 (9)	0.76 ; 1.42	2.750 (11)	1.452 (9)	7	7	2.261 (26)	3.356 (26)	0.73 ; 2.54	5.801 (8)	2.253 (23)	18	18
SNV	1.372 (10)	2.028 (11)	0.62 ; 1.78	2.624 (14)	1.372 (12)	12	12	3.088 (4)	4.557 (5)	0.55 ; 4.48	6.753 (3)	2.959 (4)	4	5
GS	1.371 (11)	2.034 (10)	0.71 ; 1.52	2.623 (16)	1.371 (13)	9	9	2.553 (15)	3.809 (15)	0.67 ; 2.97	5.209 (13)	2.549 (13)	15	16
ZION	1.358 (12)	2.003 (12)	0.62 ; 1.71	2.624 (15)	1.425 (10)	10	10	2.296 (25)	3.389 (25)	0.54 ; 3.24	4.947 (20)	2.296 (19)	21	21
CMA	1.323 (13)	1.959 (13)	0.63 ; 1.62	2.789 (8)	1.367 (15)	14	14	2.930 (9)	4.316 (9)	0.54 ; 4.11	5.218 (12)	2.981 (3)	4	7
TROW	1.308 (14)	1.938 (14)	0.74 ; 1.36	2.471 (19)	1.388 (11)	12	12	2.140 (33)	3.169 (32)	0.74 ; 2.28	4.565 (29)	2.091 (30)	19	19
KEY	1.298 (15)	1.920 (15)	0.63 ; 1.59	2.547 (18)	1.355 (16)	15	15	2.961 (8)	4.376 (8)	0.57 ; 4.15	5.872 (7)	2.888 (6)	9	9
UNM	1.281 (16)	1.902 (17)	0.66 ; 1.55	2.765 (10)	1.329 (18)	17	17	3.288 (2)	4.871 (2)	0.58 ; 4.23	6.044 (6)	2.826 (9)	2	2
STT	1.280 (17)	1.908 (16)	0.66 ; 1.55	3.018 (6)	1.370 (14)	15	15	2.191 (30)	3.248 (29)	0.57 ; 2.96	4.879 (21)	2.315 (18)	17	17
HBAN	1.274 (18)	1.890 (18)	0.63 ; 1.59	2.585 (17)	1.347 (17)	18	18	2.712 (13)	4.013 (14)	0.56 ; 3.63	5.486 (9)	2.568 (12)	14	15
JPM	1.255 (19)	1.864 (19)	0.71 ; 1.40	2.278 (27)	1.290 (22)	21	21	2.437 (22)	3.609 (22)	0.65 ; 2.84	4.076 (32)	2.276 (22)	24	24
SEIC	1.255 (20)	1.854 (21)	0.70 ; 1.43	2.637 (13)	1.320 (20)	18	18	2.125 (34)	3.161 (34)	0.66 ; 2.52	4.698 (26)	1.985 (35)	26	24
HIG	1.253 (21)	1.846 (22)	0.63 ; 1.53	2.197 (31)	1.326 (19)	23	23	2.210 (28)	3.313 (28)	0.52 ; 3.36	3.768 (36)	2.012 (33)	33	33
COF	1.245 (22)	1.857 (20)	0.63 ; 1.55	2.319 (24)	1.302 (21)	18	18	2.887 (10)	4.301 (10)	0.61 ; 3.65	5.313 (11)	2.849 (7)	7	6
BK	1.207 (23)	1.788 (23)	0.68 ; 1.39	2.310 (25)	1.274 (24)	22	21	1.590 (46)	2.352 (46)	0.50 ; 2.27	3.046 (45)	1.516 (45)	42	44
NTRS	1.200 (24)	1.783 (24)	0.68 ; 1.37	2.411 (20)	1.257 (26)	23	24	2.142 (32)	3.167 (33)	0.63 ; 2.58	5.177 (15)	2.287 (20)	27	27
FITB	1.199 (25)	1.780 (25)	0.62 ; 1.52	2.389 (21)	1.278 (23)	25	24	3.075 (5)	4.544 (6)	0.60 ; 3.98	5.113 (16)	2.694 (10)	7	7
SLM	1.169 (26)	1.743 (26)	0.50 ; 1.85	2.334 (23)	1.219 (27)	28	28	1.980 (37)	2.936 (38)	0.46 ; 3.36	3.721 (37)	1.933 (36)	35	35
AIG	1.167 (27)	1.718 (27)	0.60 ; 1.53	2.336 (22)	1.261 (25)	26	26	2.963 (7)	4.411 (7)	0.56 ; 3.96	4.811 (23)	2.450 (15)	12	12
PNC	1.137 (28)	1.682 (28)	0.67 ; 1.32	2.275 (28)	1.177 (29)	26	27	2.460 (20)	3.626 (21)	0.60 ; 3.14	4.859 (22)	2.251 (24)	30	29
WFC	1.118 (29)	1.653 (29)	0.68 ; 1.30	2.238 (29)	1.210 (28)	28	30	2.452 (21)	3.638 (20)	0.56 ; 3.13	4.107 (31)	2.070 (31)	24	24
TFC	1.086 (30)	1.608 (31)	0.65 ; 1.30	2.281 (26)	1.145 (30)	31	30	2.762 (11)	4.094 (11)	0.60 ; 3.46	4.998 (18)	2.250 (25)	12	12
AXP	1.079 (31)	1.609 (30)	0.66 ; 1.31	2.238 (30)	1.115 (31)	28	28	2.682 (14)	4.034 (13)	0.65 ; 3.45	5.451 (10)	2.827 (8)	15	12
MTB	1.009 (32)	1.496 (32)	0.62 ; 1.28	2.108 (32)	1.065 (32)	32	32	2.300 (24)	3.399 (24)	0.52 ; 3.30	4.950 (19)	2.149 (28)	31	31
AFL	0.977 (33)	1.442 (34)	0.64 ; 1.20	1.792 (39)	1.062 (33)	37	37	1.975 (39)	2.923 (40)	0.58 ; 2.89	4.345 (30)	2.175 (26)	39	39
USB	0.975 (34)	1.446 (33)	0.69 ; 1.10	2.034 (35)	1.017 (34)	33	33	2.544 (17)	3.777 (16)	0.60 ; 3.18	5.007 (17)	2.335 (17)	20	19
L	0.949 (35)	1.408 (35)	0.71 ; 1.04	1.760 (40)	0.989 (36)	34	34	2.384 (23)	3.524 (23)	0.61 ; 3.09	4.718 (25)	2.134 (29)	28	29
GL	0.925 (36)	1.380 (36)	0.70 ; 1.05	2.094 (33)	0.963 (38)	40	39	2.528 (18)	3.748 (18)	0.67 ; 2.90	5.183 (14)	2.163 (27)	23	21
CNA	0.921 (37)	1.366 (39)	0.59 ; 1.24	2.054 (34)	1.005 (35)	34	34	1.972 (40)	2.931 (39)	0.54 ; 2.64	3.348 (40)	1.616 (40)	40	40
PBCT	0.914 (38)	1.367 (38)	0.60 ; 1.19	1.844 (37)	0.946 (39)	40	41	2.463 (19)	3.654 (19)	0.51 ; 3.62	3.937 (33)	2.505 (14)	28	28
BRK	0.913 (39)	1.367 (37)	0.75 ; 0.96	1.751 (41)	0.921 (43)	39	39	1.715 (43)	2.551 (43)	0.74 ; 1.83	3.922 (34)	1.566 (43)	42	41
CI	0.906 (40)	1.344 (40)	0.47 ; 1.56	1.733 (43)	0.980 (37)	34	34	2.198 (29)	3.235 (30)	0.53 ; 2.95	3.065 (44)	1.811 (38)	33	33
AON	0.892 (41)	1.327 (41)	0.62 ; 1.12	2.033 (36)	0.925 (41)	37	37	1.460 (49)	2.174 (48)	0.51 ; 2.33	4.576 (28)	1.597 (41)	46	47
MMC	0.881 (42)	1.305 (42)	0.69 ; 0.99	1.660 (45)	0.904 (44)	40	41	1.469 (48)	2.170 (49)	0.62 ; 1.88	3.226 (42)	1.362 (48)	46	46
UNH	0.852 (43)	1.262 (43)	0.50 ; 1.34	1.739 (42)	0.945 (40)	45	45	2.058 (35)	3.023 (36)	0.58 ; 2.69	3.251 (41)	1.708 (39)	36	36
ALL	0.844 (44)	1.244 (44)	0.58 ; 1.13	1.646 (47)	0.875 (46)	43	43	1.640 (45)	2.448 (45)	0.56 ; 2.15	3.069 (43)	1.484 (46)	46	47
CINF	0.833 (45)	1.236 (45)	0.62 ; 1.06	1.655 (46)	0.922 (42)	43	43	2.142 (31)	3.190 (31)	0.48 ; 3.31	3.390 (39)	1.918 (37)	36	36
PGR	0.809 (46)	1.203 (46)	0.55 ; 1.16	1.412 (49)	0.858 (47)	47	47	1.282 (50)	1.896 (50)	0.46 ; 1.99	2.566 (49)	1.249 (50)	49	49
NYCB	0.799 (47)	1.192 (47)	0.49 ; 1.33	1.825 (38)	0.900 (45)	46	46	1.509 (47)	2.228 (47)	0.37 ; 2.89	2.065 (50)	1.272 (49)	49	50
TRV	0.770 (48)	1.134 (48)	0.57 ; 1.04	1.700 (44)	0.820 (49)	48	48	1.979 (38)	2.959 (37)	0.56 ; 2.71	2.639 (48)	1.415 (47)	41	41
HUM	0.750 (49)	1.109 (49)	0.35 ; 1.69	1.580 (48)	0.858 (48)	48	48	1.715 (44)	2.538 (44)	0.45 ; 2.88	2.723 (47)	1.590 (42)	42	44
WRB	0.686 (50)	1.019 (50)	0.54 ; 0.98	1.324 (50)	0.739 (50)	50	50	1.818 (41)	2.711 (41)	0.57 ; 2.43	2.770 (46)	1.547 (44)	42	41

Institutions are ordered according to $\overline{\Delta CI}_i$. The metrics $\overline{M}_i$, $\overline{\Delta CI}_i$, $\overline{\Delta CI}_i^{\leq}$(FZ), and $\overline{\Delta CI}_i$(CCC), denote institution-specific time averages for MES, ΔCoVaR, ΔCoVaR$^{\leq}$(FZ), and ΔCoVaR(CCC), respectively. Numbers in parentheses indicate cross-sectional ranks for the corresponding metrics. The metrics $\overline{\rho}_i$ and $\overline{\sigma}_i$ denote institution-specific time averages for market correlation and volatility, respectively. The quantities $\tilde{r}^{\Delta CI_i}$ and $\tilde{r}^{\Delta M_i}$ denote the time median of daily cross-sectional ranks for ΔCoVaR and MES, respectively. Metrics and rankings are reported for January 2012-December 2019 (left panel) and January 2020-December 2020 (right panel).

numbers reported in parentheses correspond to their cross-sectional ranks. Three main observations emerge. First, rankings based on $\overline{\Delta C}_i$ and $\overline{M}_i$ are close but not identical. This does not contradict the pointwise ranking equivalence of ΔCoVaR and MES (Section 1.3.2), since empirical rankings are constructed from time-averaged ΔCoVaR and MES estimates. Using the median of daily cross-sectional ranks, $\tilde{r}^{\Delta C_i}$ and $\tilde{r}^{\Delta M_i}$, provides a more robust ordering. The few remaining differences only reflect finite-sample estimation uncertainty, which vanishes asymptotically. Hence, the choice of an aggregated ranking metric is not innocuous. Second, shifts in rankings during the COVID-19 pandemic become very evident. For instance, during this period, Unum Group (UNM) moves from rank 16 to rank 2 under $\overline{C}_i$. This shift aligns with its role as the leading disability insurer in the US, where COVID-19 became a top cause of disability for US workers. Meanwhile, Lincoln National (LNC) consistently emerges as the top SIFI in both the low-risk and high-risk periods, driven by its high volatility (1.89 and 4.73, respectively) and strong market correlation (0.71 and 0.64, respectively) compared to other companies. This finding aligns with the long-run MES ranking from the New York University V-Lab website, where LNC was ranked 3rd among 601 US institutions in Feb. 2020.[17] Third, ΔCoVaR computed under alternative market conditioning schemes and dependence structures reveals substantial variation in systemic risk rankings. Under inequality conditioning, rankings can differ markedly from those obtained under equality conditioning. For the period 01/2020 to 12/2020, Citigroup (C) ranks second according to $\overline{\Delta C}_i^{\leq}$(FZ) but sixth under $\overline{\Delta C}_i$, while AIG is ranked seventh under $\overline{\Delta C}_i$ and only twenty-third under $\overline{\Delta C}_i^{\leq}$(FZ). Ranking differences also emerge when contrasting static and dynamic dependence structures. For 01/2020 to 12/2020, UNM ranks second under BEKK but ninth under CCC. Such inversions are more frequent during the crisis period, consistent with a more reactive market environment in which time-varying correlations become particularly relevant, a finding supported by Figures 1.1, 1.2, and 1.4. Table 1.4 in Appendix F reports the rankings over the whole period 01/2010 to 12/2020. The results are overall similar.

Significance of Differences in Systemic Risk

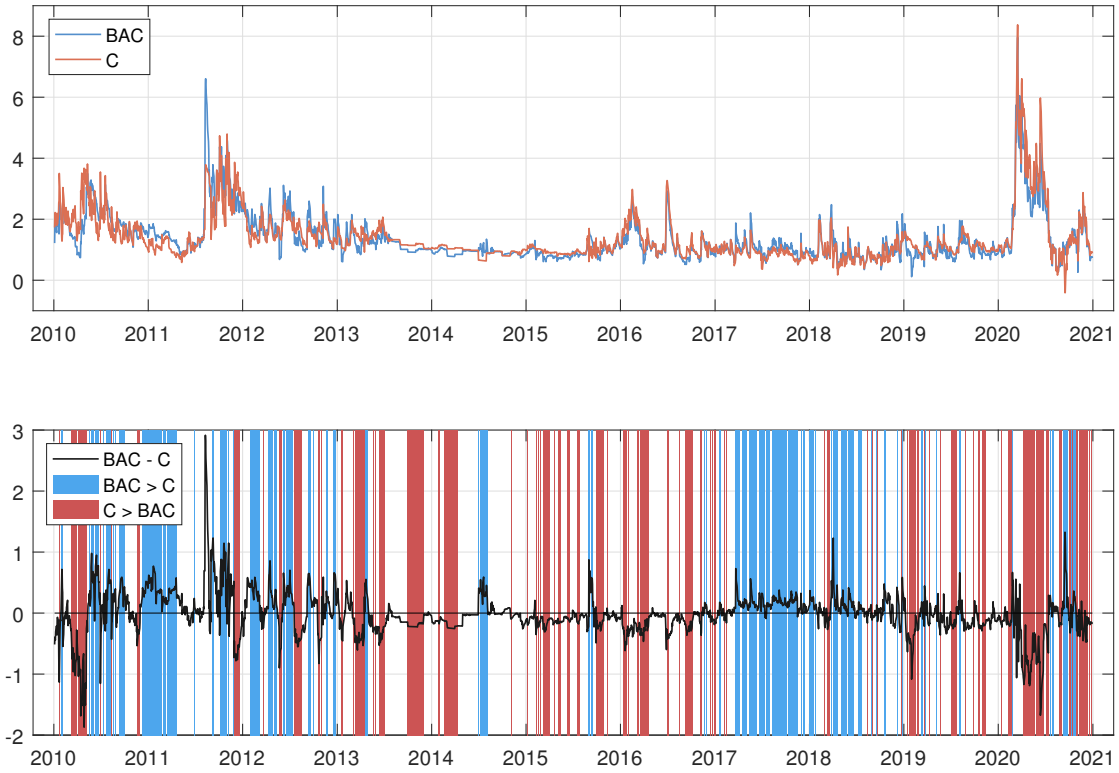
To statistically assess whether differences in systemic risk rankings are significant, we apply the tests developed in Section 1.3.2. Figure 1.5 illustrates outcomes of the Wald test in Corollary 1.2 for the hypothesis $H_0 : \lambda_{j,t} = \lambda_{k,t}$[18] for the case of Bank of America (BAC), indexed by j, and Citigroup (C), indexed by k. The top panel displays the estimates of $\lambda_{j,t}$ and $\lambda_{k,t}$ while the bottom panel depicts their differences with shaded areas representing periods where H_0 is rejected at the 95% confidence level (with blue shading the cases of BAC being more risky and red otherwise). We observe that significant differences between the two institutions are clustered over time. Hence, systemic risk (rankings) tend to persist as also confirmed by a first-order autocorrelation of the rejections

[17] Note that our rankings do not necessarily highlight the largest banks, unlike the FSB's list of G-SIBs. For instance, in the period 01/2020 to 12/2020, JP Morgan (JPM) ranks 22nd under MES and ΔCoVaR, despite being regularly ranked as the top G-SIB. This result is expected, as neither MES nor ΔCoVaR explicitly accounts for size or leverage, unlike SRISK or SES (building on MES).

[18] We test against right-tailed and left-tailed alternatives.

of 0.69. Nevertheless, on 63% of the days, the null is not rejected, i.e., systemic risk rankings of BAC and C are statistically not distinguishable.

Figure 1.5: Systemic risk differences between Bank of America and Citigroup.

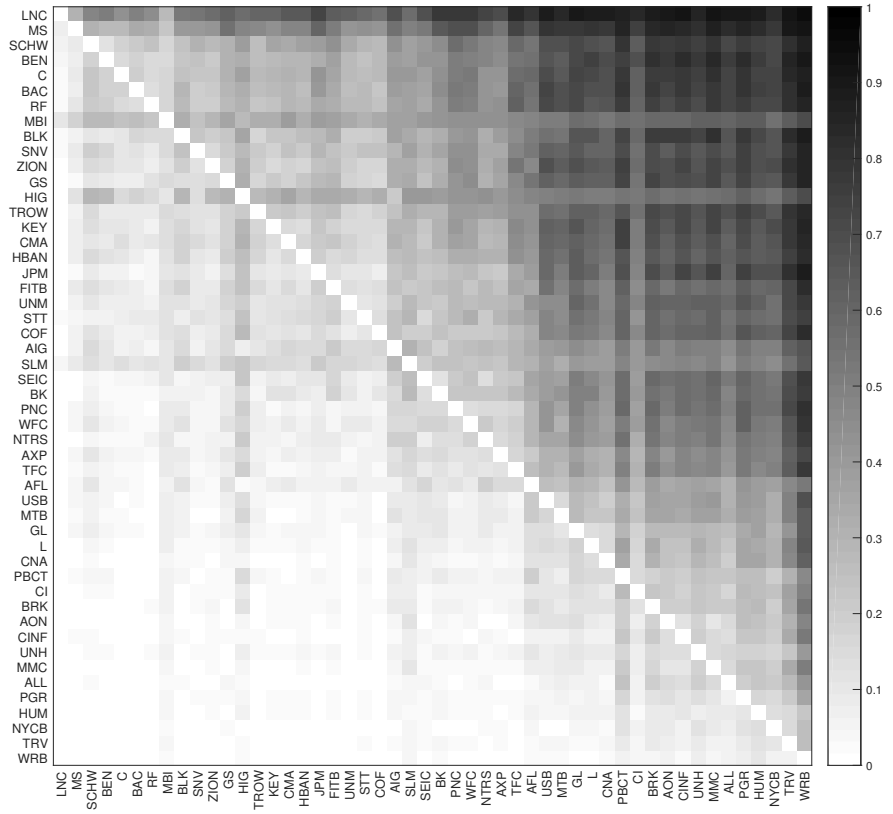


The top panel displays $\hat{\lambda}_{j,t}$ and $\hat{\lambda}_{k,t}$ for BAC (denoted by j) and Citigroup (denoted by k). The bottom panel plots the difference $\hat{\lambda}_{j,t} - \hat{\lambda}_{k,t}$ (BAC − C). Shaded regions indicate periods where the test statistic 1.22 rejects H_0 at the 95% confidence level: blue shading indicates $H_1 : \lambda_{j,t} > \lambda_{k,t}$ (BAC > C), red shading indicates $H_1 : \lambda_{j,t} < \lambda_{k,t}$ (C > BAC). BEKK parameters are estimated using a rolling window of size $n = 500$ and are updated on a monthly basis.

Figure 1.6 summarizes the results for all 1225 distinct pairs of firms. The heatmap displays the frequency of rejections of $H_0 : \lambda_{j,t} = \lambda_{k,t}$ in favor of $H_1 : \lambda_{j,t} > \lambda_{k,t}$ at the 95% confidence level, with j on the y-axis and k on the x-axis. Overall, we find that in 48% of the cases, neither of the one-sided hypotheses $H_1 : \lambda_{j,t} > \lambda_{k,t}$ nor $H_1 : \lambda_{j,t} < \lambda_{k,t}$ is rejected, i.e., in roughly half of the cases the differences are statistically significant, while in the other half the systemic riskiness of institutions cannot be distinguished. The most distinct case is the comparison of LNC versus WRB, where LNC is shown to be more systemically risky in 94% of all cases.

To *jointly* test all bank-market pairs we utilize again test statistic (1.22) of Corollary 1.2 but derive the critical values from the distribution of the maximum statistic (1.24) guaranteeing strong FWER control (see Corollary 1.3). Critical values are computed by

Figure 1.6: Rejection rates of $H_0 : \lambda_{j,t} = \lambda_{k,t}$ for all pairwise comparisons.

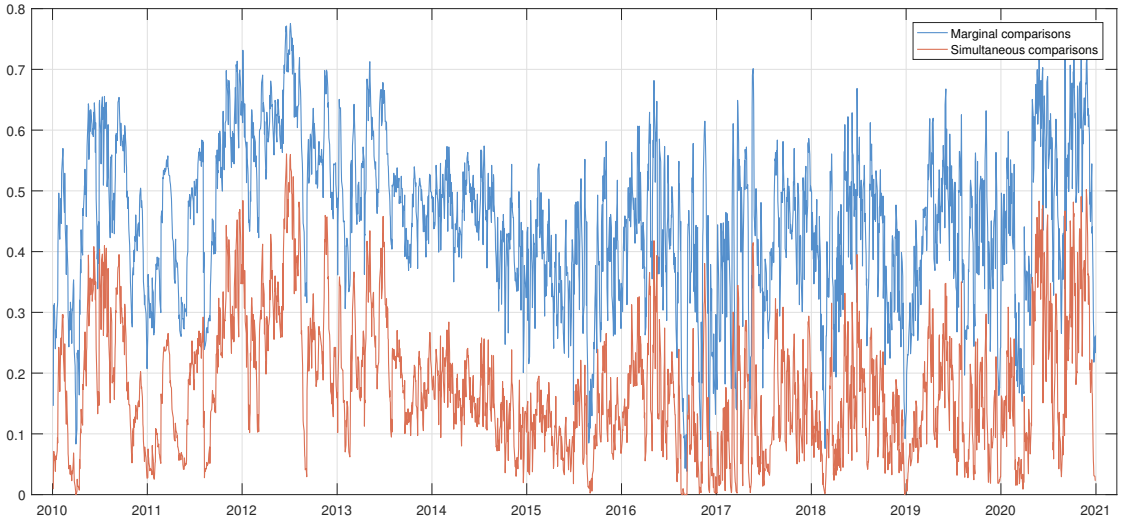


Rejection frequencies of the null hypothesis $H_0 : \lambda_{j,t} = \lambda_{k,t}$ against the one-sided alternative $H_1 : \lambda_{j,t} > \lambda_{k,t}$ at the 95% confidence level, with j on the y-axis and k on the x-axis based on test statistic 1.22. E.g., a value of 0.8 indicates that the firm on the y-axis has a significantly higher MES/ΔCoVaR than the firm on the x-axis on 80% of the days. BEKK parameters are estimated using a rolling window of size $n = 500$ and are updated on a monthly basis.

simulation with 1,000 replications, from the procedure outlined in Appendix G. Figure 1.7 displays the rejection rates of the two-sided test of the null $H_0 : \lambda_{j,t} = \lambda_{k,t}$ across all pairwise comparisons, under both marginal (blue) and joint (red) inference on the 95% confidence level. Two results emerge. First, the informativeness of rankings is strongly time-varying. Under joint inference, at some days systemic risk rankings are not informative at all (with a rejection rate of 0%). On these days, risks are statistically indistinguishable and thus all banks share the same rank of systemic importance. On other days, however, rankings reveal clear separations between institutions' systemic risk (with a maximum rejection rate of 56%). Second, rejection rates are higher under marginal inference since FWER control in the joint case is conservative when 1,225 hypotheses are tested jointly. However, even without the FWER adjustment, statistical inference remains necessary as many institutions still exhibit statistically similar levels of systemic

risk.

Figure 1.7: Rejection rates under marginal and joint inference.

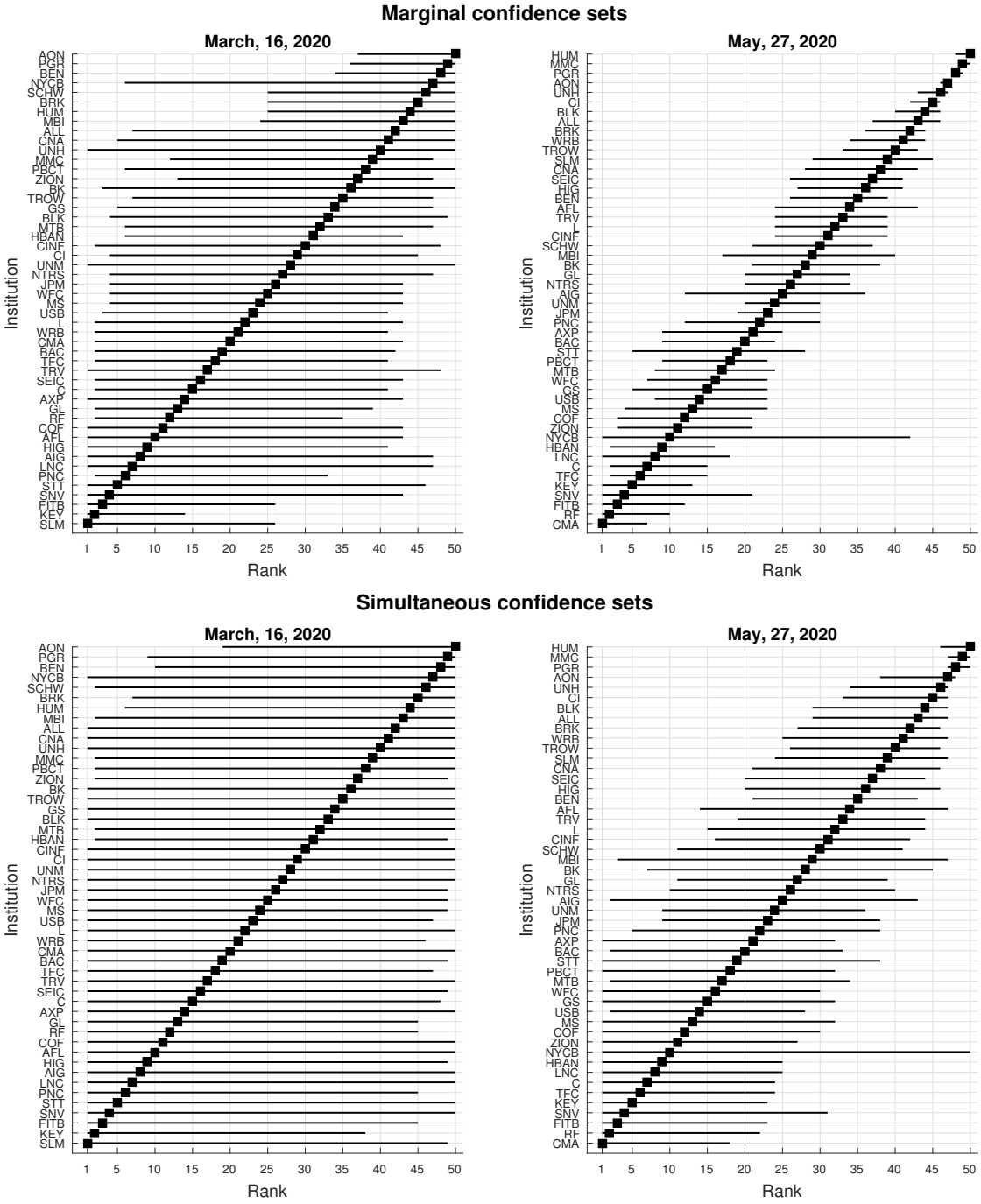


Rejection frequencies of the null hypothesis $H_0 : \lambda_{j,t} = \lambda_{k,t}$ across all bank-market pairs based on marginal (blue) and joint (red) two-sided tests on the 95% confidence level. BEKK parameters are estimated with a rolling window of size $n = 500$, updated monthly.

Figure 1.8 shows the corresponding confidence intervals for systemic risk ranks of all 50 companies on March 16, 2020 (left panel) and May 27, 2020 (right panel). The (95%) confidence intervals in the upper panels are based on marginal inference (as of Corollory 1.2, while the lower panels are based on joint inference as of Corollary 1.4. On March 16, joint inference rejects 4% of pairwise equalities, resulting in wide confidence sets and high uncertainty in rankings. Conversely, on May 27, the rejection rate rises to 48%, yielding narrower sets and more informative rankings. These two snapshots, taken 50 trading days apart, illustrate the time variation in the precision of systemic risk rankings. As expected, marginal confidence sets are narrower than their simultaneous counterparts, since they ignore FWER control.

Obviously, larger estimation samples increase the precision of ranking signals. This is shown in Figures 1.11 and 1.12 in Appendix F, which replicate Figures 1.7 and 1.8 but for larger (rolling) sampling windows of $n = 2000$. As expected, the rejection rates are larger compared to the setting with $n = 500$. Nevertheless, many firms remain statistically indistinguishable, suggesting that the null of equal systemic risk contributions still holds for a substantial fraction of firms at the population level. Figure 1.12 reveals that even based on $n = 2000$, there are days of uninformative systemic risk rankings, such as, e.g., March 16, 2020. Hence, even in large samples, statistical inference is key for interpreting systemic risk rankings.

Figure 1.8: 95% confidence sets based on marginal and joint inference.



Each square indicates the point estimate of the rank of a given institution on the corresponding date. Horizontal lines represent 95% confidence sets for the ranks for March 16, 2020, and May 27, 2020. The upper panels show marginal confidence sets and the lower panels show joint confidence sets. BEKK parameters are estimated with a rolling window of size $n = 500$, updated monthly.

1.6 Conclusions

This paper derives statistical inference for a two-step semi-parametric procedure to estimate MES and ΔCoVaR measures, which are driven by multivariate GARCH dynamics. Model parameters are estimated in a first step under general assumptions of the multivariate volatility process using QMLE, while in the second step, corresponding tail components are estimated nonparametrically.

Exploiting closed-form expressions for the underlying measures we prove the consistency and derive the asymptotic distribution of the resulting estimators. Moreover, we propose Wald tests for pairwise equality in systemic risk between different banks and extend the procedure to joint inference with control of the family-wise error rate. This methodology allows us to construct confidence sets for ranks in systemic risk rankings.

This statistical framework opens up new possibilities to assess the reliability of systemic risk estimates. Based on simulation evidence and a thorough empirical application based on 50 US banks, we demonstrate that (i) the informativeness of systemic risk measures strongly varies over time and can be temporally comparably low and (ii) the incorporation of time variation in return cross-correlations strongly affects the precision of systemic risk estimates.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank for the support by funds of Oesterreichische Nationalbank (Austrian Central Bank, Anniversary Fund, Project No. 18669) and the DFG for support through research unit RU 5583 "Asset Pricing and Asset Allocation under Regulatory Uncertainty". Finally, we would like to thank the participants of the Workshop "Recent Advances in Econometrics" (UC Louvain, June 2, 2023), 18th International Conference on Computational and Financial Econometrics (King's College London, December 14-16, 2024), 7th Quantitative Finance and Financial Econometrics Conference (Aix Marseille School of Economics, June 5-6, 2025), 17th Annual Conference of the Society for Financial Econometrics (ESSEC Business School, June 10-12, 2025), and Finance Seminar at ENSAE-CREST (November 6, 2025).

Appendix to Chapter 1

A - Definitions and Expressions for SRISK and SES

Acharya et al. (2012) and Brownlees and Engle (2017) define SRISK as the expected capital shortfall of a financial institution conditionally on a systemic crisis. The capital

shortfall $CS_{i,t}$ represents the amount of capital a firm would need to raise to meet a prudential capital ratio k, given its liabilities and equity:

$$CS_{i,t} = k(L_{i,t} + W_{i,t}) - W_{i,t},$$

where $L_{i,t}$ and $W_{i,t}$ denote the book value of debt and the market value of equity, respectively. Hence,

$$\begin{aligned} SRISK_{it} &= \mathbb{E}(CS_{it}|y_{mt} \leq VaR_{mt}(\alpha); \mathcal{F}_{t-1}) \\ &= \mathbb{E}(k(L_{it} + W_{it}) - W_{it}|y_{mt} \leq VaR_{mt}(\alpha); \mathcal{F}_{t-1}) \\ &= kL_{it} + (k-1)W_{it} + (k-1)W_{it}MES_{it}(\alpha), \end{aligned}$$

showing that SRISK extends MES by incorporating the institution's leverage and size.

The SES proposed by Acharya et al. (2017) similarly measures the extent to which a bank's equity falls below a target ratio k of assets during systemic distress:

$$SES_{i,t} = W_{i,t}(kLV_{i,t} + \kappa_1 - 1) + W_{i,t}\kappa_2 MES_{i,t}(\alpha),$$

where $LV_{i,t} = (L_{i,t} + W_{i,t})/W_{i,t}$ is the quasi-leverage ratio, and κ_1 and κ_2 are constants.

These two formulations show that SRISK and SES are both extensions of MES that account for liabilities and firm size. Since debt levels are known at time t, the only source of estimation uncertainty in SRISK and SES comes from MES itself.

B - Proof of Propositions 1 and 2

B.1. Proof of Proposition 1. Given the model structure of Section 1.2.1, the firm return at time t is

$$y_{i,t} = \sigma_{i,t}\rho_{i,t}\eta_{m,t} + \sigma_{i,t}\sqrt{1-\rho_{i,t}^2}\eta_{i,t}.$$

Substituting the expression of $y_{i,t}$ into the definition of MES, we have

$$\begin{aligned} \mathbb{E}\left(y_{i,t}|y_{m,t} \leq VaR_{m,t}(\alpha); \mathcal{F}_{t-1}\right) = \sigma_{i,t}\rho_{i,t}\mathbb{E}\left(\eta_{m,t}|y_{m,t} \leq VaR_{m,t}(\alpha); \mathcal{F}_{t-1}\right) \\ + \sigma_{i,t}\sqrt{1-\rho_{i,t}^2}\mathbb{E}\left(\eta_{i,t}|y_{m,t} \leq VaR_{m,t}(\alpha); \mathcal{F}_{t-1}\right). \end{aligned}$$

As the stress event can be equivalently defined as $y_{m,t} \leq VaR_{m,t}(\alpha)$ or $\eta_{m,t} \leq \xi_{m,\alpha}$, where $\xi_{m,\alpha}$ is the α-level quantile of $\eta_{m,t}$, it follows

$$\begin{aligned} MES_{i,t}(\alpha) = \sigma_{i,t}\rho_{i,t}\mathbb{E}\left(\eta_{m,t}|\eta_{m,t} \leq \xi_{m,\alpha}; \mathcal{F}_{t-1}\right) \\ + \sigma_{i,t}\sqrt{1-\rho_{i,t}^2}\mathbb{E}\left(\eta_{i,t}|\eta_{m,t} \leq \xi_{m,\alpha}; \mathcal{F}_{t-1}\right). \end{aligned}$$

As $\mathbb{E}\left(\eta_{i,t}\right) = 0$, $\eta_{i,t}$ and $\eta_{m,t}$ are independent random variables, we have $\mathbb{E}\left(\eta_{i,t}|\eta_{m,t} \leq \xi_{m,\alpha}; \mathcal{F}_{t-1}\right) = 0$ and $\mathbb{E}\left(\eta_{m,t}|\eta_{m,t} \leq \xi_{m,\alpha}; \mathcal{F}_{t-1}\right) = \mathbb{E}\left(\eta_{m,t}|\eta_{m,t} \leq \xi_{m,\alpha}\right)$. Hence, we have

$$MES_{i,t}(\alpha) = \sigma_{i,t}\rho_{i,t}\mu_{m,\alpha}.$$

■

B.2. Proof of Proposition 2. We start with y_{it} defined according to the model structure of Section 1.2.1,

$$y_{i,t} = \sigma_{i,t}\rho_{i,t}\eta_{m,t} + \sigma_{i,t}\sqrt{1-\rho_{i,t}^2}\eta_{i,t}. \tag{1.28}$$

By definition, we verify that,

$$Pr\left(y_{i,t} \leq CoVaR_{i,t}(\beta,\alpha)|y_{m,t} = VaR_{m,t}(\alpha); \mathcal{F}_{t-1}\right) = \beta.$$

Substituting $y_{i,t}$ by the expression in (1.28), we have

$$Pr\left(\sigma_{i,t}\rho_{i,t}\eta_{m,t} + \sigma_{i,t}\sqrt{1-\rho_{i,t}^2}\eta_{i,t} \leq CoVaR_{i,t}(\beta,\alpha)|y_{m,t} = VaR_{m,t}(\alpha); \mathcal{F}_{t-1}\right) = \beta.$$

As the stress event can be equivalently defined as $y_{m,t} = VaR_{m,t}(\alpha)$ or $\eta_{m,t} = \xi_{m,\alpha}$, where $\xi_{m,\alpha}$ is the α-level quantile of $\eta_{m,t}$, it follows that

$$Pr\left(\sigma_{i,t}\rho_{i,t}\xi_{m,\alpha} + \sigma_{i,t}\sqrt{1-\rho_{i,t}^2}\eta_{i,t} \leq CoVaR_{i,t}(\beta,\alpha)|\mathcal{F}_{t-1}\right) = \beta.$$

Rearranging the terms and applying the cdf of $\eta_{i,t}$, written as $F_{\eta_i}(.)$, we obtain

$$F_{\eta_i}\left(\frac{CoVaR_{i,t}(\beta,\alpha) - \sigma_{i,t}\rho_{i,t}\xi_{m,\alpha}}{\sigma_{i,t}\sqrt{1-\rho_{i,t}^2}}\right) = \beta. \tag{1.29}$$

Applying the inverse cdf of η_i on the right and left hand part of Equation (1.29), and rearranging the terms yields

$$CoVaR_{i,t}(\beta,\alpha) = \sigma_{i,t}\sqrt{1-\rho_{i,t}^2}\xi_{i,\beta} + \sigma_{i,t}\rho_{i,t}\xi_{m,\alpha}, \tag{1.30}$$

where $\xi_{i,\beta}$ denotes the β-level quantile of $\eta_{i,t}$. Following the same reasoning, $CoVaR_{i,t}(\beta, 0.5)$ is given by $\sigma_{i,t}\sqrt{1-\rho_{i,t}^2}\xi_{i,\beta} + \sigma_{i,t}\rho_{i,t}\xi_{m,0.5}$. We end with $\Delta\text{CoVaR}_{i,t}(\beta,\alpha) \equiv \Delta CoVaR_{i,t}(\alpha)$ defined as

$$\Delta CoVaR_{i,t}(\alpha) = \sigma_{i,t}\rho_{i,t}\left(\xi_{m,\alpha} - \xi_{m,0.5}\right).$$

■

C - Proofs of Asymptotic Results

C.1. Proof of Theorem 1.1. The quantile function can be obtained by solving an optimization problem. Here, we focus on the estimated quantile $\hat{\xi}_{m,\alpha}$, which represents the α-level quantile of the estimated market innovations $\hat{\eta}_{m,t}$. The procedure for $\hat{\xi}_{i,\alpha}$ is identical. The estimated quantile $\hat{\xi}_{m,\alpha}$ is obtained by

$$\hat{\xi}_{m,\alpha} = \arg\min_{z\in\mathbb{R}} \frac{1}{n}\sum_{t=1}^{n} \rho_\alpha\left(\hat{\eta}_{m,t} - z\right), \tag{1.31}$$

where $\rho_\alpha(u)$ is the quantile loss function defined as $\rho_\alpha(u) = u(\alpha - \mathbb{1}_{\{u<0\}})$, with $\mathbb{1}_{\{u<0\}}$ being the indicator function.

Similar to Van Der Vaart and Wellner (2007) we define

$$P\rho_\alpha = \int \rho_\alpha dP, \quad \mathbb{P}_n\rho_\alpha = \frac{1}{n}\sum_{i=1}^{n} \rho_\alpha(X_i), \quad \mathbb{G}_n\rho_\alpha = \sqrt{n}(\mathbb{P}_n - P)\rho_\alpha,$$

and have the following expansion:

$$\sqrt{n}\left(\mathbb{P}_n\rho_{\alpha,\hat{\boldsymbol{\theta}}_n} - P\rho_{\alpha,\boldsymbol{\theta}}\right) = \mathbb{G}_n\left(\rho_{\alpha,\hat{\boldsymbol{\theta}}_n} - \rho_{\alpha,\boldsymbol{\theta}}\right) + \mathbb{G}_n\rho_{\alpha,\boldsymbol{\theta}} + \sqrt{n}P\left(\rho_{\alpha,\hat{\boldsymbol{\theta}}_n} - \rho_{\alpha,\boldsymbol{\theta}}\right), \quad (1.32)$$

for a collection $\{\rho_{\alpha,\boldsymbol{\theta}} : \alpha \in [0,1], \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ of measurable functions. From Assumptions 1.1-1.3, we know that $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}_0$. Thus, $\sup_\alpha P\left(\rho_{\alpha,\hat{\boldsymbol{\theta}}_n} - \rho_{\alpha,\boldsymbol{\theta}}\right)^2 \to_p 0$.

As for the check function, it is not immediately clear that it is Donsker; we establish this property in Appendix C.2. (See also Example 19.25 of Van der Vaart (2000)). By the Theorem 2.1 of Van Der Vaart and Wellner (2007) we have, as $n \to \infty$,

$$\sup_{\alpha\in\boldsymbol{\Theta}} \left|\mathbb{G}_n\left(\rho_{\alpha,\hat{\boldsymbol{\theta}}_n} - \rho_{\alpha,\boldsymbol{\theta}}\right)\right| \to_p 0. \quad (1.33)$$

Therefore, we obtain

$$\sqrt{n}\left(\mathbb{P}_n\rho_{\alpha,\hat{\boldsymbol{\theta}}_n} - P\rho_{\alpha,\boldsymbol{\theta}}\right) = \mathbb{G}_n\rho_{\alpha,\boldsymbol{\theta}} + \sqrt{n}P\left(\rho_{\alpha,\hat{\boldsymbol{\theta}}_n} - \rho_{\alpha,\boldsymbol{\theta}}\right) + o_p(1). \quad (1.34)$$

The first term on the right-hand side converges to a Gaussian process according to the functional central limit theorem under Assumptions 1.1–1.5. The behavior of the second term depends on the estimators $\hat{\boldsymbol{\theta}}_n$, and would typically follow from an application of the (functional) delta method. In our case, the check function makes the situation a bit more involved.

Therefore we first require convexity of the check function which is easy to see as it can be rewritten as a weighted average of two absolute value functions,

$$\rho_\alpha(\hat{\eta}_{m,t} - z) = \alpha\{(\hat{\eta}_{m,t} - z)_+ - (\hat{\eta}_{m,t})_+\} + (1-\alpha)\{(z - \hat{\eta}_{m,t})_+ - (-\hat{\eta}_{m,t})_+\}.$$

Here, $(\hat{\eta}_{m,t} - z)_+$ and $(z - \hat{\eta}_{m,t})_+$ are both convex in z (as they are maximum functions of linear expressions), while $(\hat{\eta}_{m,t})_+$ and $(-\hat{\eta}_{m,t})_+$ do not depend on z. Since α and $1-\alpha$ are in $[0,1]$, the convexity is preserved. The expected value $\mathbb{E}[\rho_\alpha(\eta - z)]$ is minimized at $z = \xi_\alpha$, the true α-quantile of the innovation distribution. Under the assumptions of Theorem 1.1 (specifically, that the density f_η is continuous and strictly positive in a neighborhood of ξ_α), we have the quadratic approximation

$$\mathbb{E}\{\rho_\alpha(\eta - z) - \rho_\alpha(\eta - \xi_\alpha)\} = \frac{1}{2}f_\eta(\xi_\alpha)(z - \xi_\alpha)^2 + o((z - \xi_\alpha)^2).$$

Assumptions 1.4,1.5, and 1.6 also ensure the QMLE estimator has a finite covariance matrix, therefore, we could apply Theorem 2.1 in Hjort and Pollard (2011). First, we

expand the second term of Equation (1.34) using a Taylor expansion around $\boldsymbol{\theta}$. Then, by employing convexity arguments, referring to Theorem 2.1 in Hjort and Pollard (2011), and taking the argmin of both sides, we obtain

$$\sqrt{n}(\hat{\xi}_{m,\alpha}(\widehat{\boldsymbol{\theta}}_n)-\xi_{m,\alpha}(\boldsymbol{\theta}_0)) = \sqrt{n}(\hat{\xi}_{m,\alpha}(\boldsymbol{\theta}_0)-\xi_{m,\alpha}(\boldsymbol{\theta}_0))+\frac{\partial}{\partial\boldsymbol{\theta}}\mathbb{E}\left[\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})\right]'_{|\tilde{\boldsymbol{\theta}}}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)+o_p(1). \tag{1.35}$$

Finally, by the Continuous Mapping Theorem, we have

$$\frac{\partial}{\partial\boldsymbol{\theta}}\mathbb{E}\left[\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})\right]'_{|\tilde{\boldsymbol{\theta}}} = lim_{n\to\infty}\frac{\partial}{\partial\boldsymbol{\theta}}\mathbb{E}\left[\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})\right]'_{|\boldsymbol{\theta}_0} + o_p(1),$$

which, using Slutsky's theorem, leads to

$$\begin{aligned}\sqrt{n}(\hat{\xi}_{m,\alpha}(\widehat{\boldsymbol{\theta}}_n) - \xi_{m,\alpha}(\boldsymbol{\theta}_0)) = &\underbrace{\sqrt{n}(\hat{\xi}_{m,\alpha}(\boldsymbol{\theta}_0) - \xi_{m,\alpha}(\boldsymbol{\theta}_0))}_{A}\\ &+\underbrace{lim_{n\to+\infty}\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})_{|\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]'\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)}_{B}+o_p(1).\end{aligned}$$

Combining the results of Term A and Term B, we obtain

$$\begin{aligned}\sqrt{n}(\hat{\xi}_{m,\alpha}(\hat{\boldsymbol{\theta}}_n) - \xi_{m,\alpha}(\boldsymbol{\theta}_0)) = &\frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t}(\boldsymbol{\theta}_0) < \xi_{m,\alpha}) - \alpha)\\ &+\xi_{m,\alpha}(\boldsymbol{\theta}_0)\mathbb{E}\left[\frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_0)}\frac{\partial\sigma_{m,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]'\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1).\end{aligned} \tag{1.36}$$

We next explain how this expansion is obtained by treating Term A and Term B separately.
Term A: Term A is given by

$$\sqrt{n}(\hat{\xi}_{m,\alpha}(\boldsymbol{\theta}_0) - \xi_{m,\alpha}(\boldsymbol{\theta}_0)) = \frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t}(\boldsymbol{\theta}_0) < \xi_{m,\alpha}) - \alpha).$$

This corresponds to the expansion of the distribution of the α-quantile of $\eta_{m,t}$ when the residuals are not affected by the estimation of $\boldsymbol{\theta}_0$. Such a distribution is quite common, as indicated in Corollary 21.5 of Van der Vaart (2000). We can easily show that: $\sqrt{n}(\hat{\xi}_{m,\alpha}(\boldsymbol{\theta}_0) - \xi_{m,\alpha}(\boldsymbol{\theta}_0)) \to N\left(0, \frac{\alpha(1-\alpha)}{f_\eta(\xi_{m,\alpha}(\boldsymbol{\theta}_0))^2}\right)$.

Term B:
Term B represents the residual component due to the estimation of $\boldsymbol{\theta}_0$. In the following we will show

$$lim_{n\to+\infty}\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})_{|\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]' = \xi_{m,\alpha}(\boldsymbol{\theta}_0)\mathbb{E}\left[\frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_0)}\frac{\partial\sigma_{m,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]' + o_p(1). \tag{1.37}$$

The right-hand side of this result is also connected to the findings of Francq and Zakoïan (2015).
Proof for Term B: First, define the empirical cdf of $\eta_{m,t}(\boldsymbol{\theta})$ evaluated at $\xi_{m,\alpha} = \xi_{m,\alpha}(\boldsymbol{\theta}_0) \in \mathbb{R}$: $F_n(\xi_{m,\alpha}, \boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}(\eta_{m,t}(\boldsymbol{\theta}) \leq \xi_{m,\alpha})$. Next, we need to find an expression for $\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})$. Denote by $F^{-1}(x)$ the inverse cdf of $\eta_{m,t}(\boldsymbol{\theta}_0)$ evaluated at x. We use $\hat{\xi}_{m,\alpha}(\hat{\boldsymbol{\theta}}_n) = \xi_{m,\alpha} + e_0$ as a short representation of Equation (1.35). From Assumptions 1.1–1.6 and the resulting Bahadur expansion (1.17), we have $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) = O_p(1)$ as $n \to \infty$. Then under Assumption 1.7, the proof follows from Vyazilov (1999); Boldin (2000); Berkes and Horváth (2003). As $n \to \infty$, it follows that

$$\sup_{x \in \mathbf{R}} \left| \sqrt{n}\left(\hat{F}_n(x) - F_n(x)\right) - \frac{1}{2}\sqrt{n} x f(x) \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)^T \boldsymbol{\Omega}_{\boldsymbol{\mu}} \right| = o_p(1), \tag{1.38}$$

where $\boldsymbol{\Omega}_{\boldsymbol{\mu}}{}' = \mathbb{E}\left[\frac{1}{\sigma^2_{m,t}(\boldsymbol{\theta}_0)} \frac{\partial \sigma^2_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right]$ and $\hat{F}_n(x)$ and $F_n(x)$ are empirical distributions of estimated and true innovations, respectively. Therefore, we have the following expansion:

$$\begin{aligned} F_n(\xi_{m,\alpha}, \widehat{\boldsymbol{\theta}}_n) &= \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}(\eta_{n,t}(\widehat{\boldsymbol{\theta}}_n) \leq \xi_{m,\alpha}) \\ &= F_n(\xi_{m,\alpha}, \boldsymbol{\theta}_0) + \frac{1}{2}\xi_{m,\alpha} f(\xi_{m,\alpha})(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' \boldsymbol{\Omega}_{\boldsymbol{\mu}} + o_p(1). \end{aligned} \tag{1.39}$$

We know that the second part on right hand of expansion in Equation (1.39) is also $o_p(1)$, we thus note that $F_n(\xi_{m,\alpha}, \widehat{\boldsymbol{\theta}}_n) = \alpha + e_1$, where e_1 is $o_p(1)$, and have $F^{-1}(F_n(\xi_{m,\alpha}, \widehat{\boldsymbol{\theta}}_n)) = F^{-1}(\alpha + e_1)$. A Taylor expansion gives

$$F^{-1}(F_n(\xi_{m,\alpha}, \widehat{\boldsymbol{\theta}}_n)) = F^{-1}(\alpha) + e_1 \frac{\partial F^{-1}(u)}{\partial u}_{|_\epsilon},$$

where ϵ is an intermediate point between α and $\alpha + e_1$ which simplifies to

$$F^{-1}(F_n(\xi_{m,\alpha}, \widehat{\boldsymbol{\theta}}_n)) = \xi_{m,\alpha} + o_p(1).$$

Substituting $\xi_{m,\alpha}$ by $\hat{\xi}_{m,\alpha}(\widehat{\boldsymbol{\theta}}_n) - e_0$, we get

$$F^{-1}(F_n(\xi_{m,\alpha}, \widehat{\boldsymbol{\theta}}_n)) = \hat{\xi}_{m,\alpha}(\widehat{\boldsymbol{\theta}}_n) - e_0 + o_p(1),$$

and thus,

$$\hat{\xi}_{m,\alpha}(\widehat{\boldsymbol{\theta}}_n) = F^{-1}(F_n(\xi_{m,\alpha}, \widehat{\boldsymbol{\theta}}_n)) + o_p(1). \tag{1.40}$$

Next, we use the expression in (1.40). Taking expectations and computing derivatives, we have

$$\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \hat{\xi}_{m,\alpha}(\boldsymbol{\theta})\right] = \int_{-\infty}^{+\infty} \frac{\partial}{\partial \boldsymbol{\theta}} F^{-1}\left(\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}(\eta_{m,t}(\boldsymbol{\theta}) \leq \xi_{m,\alpha})\right) f(\eta_{m,t}(\boldsymbol{\theta})) d\eta_{m,t}(\boldsymbol{\theta}) + o_p(1), \tag{1.41}$$

where $f(.)$ is the probability density function of $\eta_{m,t}$. By denoting $I_t(\boldsymbol{\theta}, \xi_{m,\alpha}) = \mathbb{1}(\eta_{m,t}(\boldsymbol{\theta}) \leq \xi_{m,\alpha})$ the chain rule implies

$$\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}}\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})\right] = \int_{-\infty}^{+\infty} \frac{\partial \frac{1}{n}\sum_{t=1}^{n} I_t(\boldsymbol{\theta},\xi_{m,\alpha})}{\partial \eta_{m,t}(\boldsymbol{\theta})} \frac{\partial \eta_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial F^{-1}(u)}{\partial u}_{|_{F_n(\xi_{m,\alpha},\boldsymbol{\theta})}} f(\eta_{m,t}(\boldsymbol{\theta}))d\eta_{m,t}(\boldsymbol{\theta}) + o_p(1).$$

The sum is put outside the integral leading to

$$\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}}\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})\right] = \frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{+\infty} \frac{\partial I_t(\boldsymbol{\theta}, \xi_{m,\alpha})}{\partial \eta_{m,t}(\boldsymbol{\theta})} \frac{\partial \eta_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial F^{-1}(u)}{\partial u}_{|_{F_n(\xi_{m,\alpha},\boldsymbol{\theta})}} f(\eta_{m,t}(\boldsymbol{\theta}))d\eta_{m,t}(\boldsymbol{\theta}) + o_p(1).$$

As

$$\frac{\partial \eta_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{y_t}{\sigma_{m,t}(\boldsymbol{\theta})^2}\frac{\partial \sigma_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{\eta_t}{\sigma_{m,t}(\boldsymbol{\theta})}\frac{\partial \sigma_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

we have

$$\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}}\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})\right] = \frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{+\infty} \frac{\partial I_t(\boldsymbol{\theta}, x)}{\partial \eta_{m,t}(\boldsymbol{\theta})} \left(-\frac{\eta_{m,t}(\boldsymbol{\theta})}{\sigma_{m,t}(\boldsymbol{\theta})}\frac{\partial \sigma_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \frac{\partial F^{-1}(u)}{\partial u}_{|_{F_n(\xi_{m,\alpha},\boldsymbol{\theta})}} f(\eta_{m,t}(\boldsymbol{\theta}))d\eta_{m,t}(\boldsymbol{\theta}) + o_p(1).$$

Following the same argument as in Lemma A.1 of Lambert et al. (2012), the quantity $\frac{\partial I_t(\boldsymbol{\theta},\xi_{m,\alpha})}{\partial \eta_{m,t}(\boldsymbol{\theta})} \equiv \frac{\partial \mathbb{1}(\eta_{m,t}(\boldsymbol{\theta})-\xi_{m,\alpha}\leq 0)}{\partial \eta_{m,t}(\boldsymbol{\theta})}$ is a Dirac function which is equal to $-\infty$ at $\eta_{m,t}(\boldsymbol{\theta}) = \xi_{m,\alpha}$ and 0 elsewhere. As a result, the integral may be simplified to an evaluation at the point $\eta_{m,t}(\boldsymbol{\theta}) = \xi_{m,\alpha}$, thus

$$\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}}\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})\right] = \frac{1}{n}\sum_{t=1}^{n} \frac{\xi_{m,\alpha}}{\sigma_{m,t}(\boldsymbol{\theta})}\frac{\partial \sigma_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial F^{-1}(u)}{\partial u}_{|_{F_n(\xi_{m,\alpha},\boldsymbol{\theta})}} f(\xi_{m,\alpha}) + o_p(1).$$

Finally, we have $\frac{\partial F^{-1}(u)}{\partial u}_{|_{F_n(\xi_{m,\alpha},\boldsymbol{\theta})}} = \frac{1}{f(F^{-1}(F_n(\xi_{m,\alpha},\boldsymbol{\theta})))}$, which leads to

$$\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}}\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})\right] = \frac{1}{n}\sum_{t=1}^{n} \frac{\xi_{m,\alpha}}{\sigma_{m,t}(\boldsymbol{\theta})}\frac{\partial \sigma_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{f(F^{-1}(F_n(\xi_{m,\alpha},\boldsymbol{\theta})))} f(\xi_{m,\alpha}) + o_p(1),$$

as $lim_{n\to\infty}F_n(\xi_{m,\alpha},\boldsymbol{\theta}_0) = F(\xi_{m,\alpha},\boldsymbol{\theta}_0)$ and by the law of large number, we can substitute the sample mean by the expectation operator. Then, we have

$$lim_{n\to+\infty}\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}}\hat{\xi}_{m,\alpha}(\boldsymbol{\theta})\right] = \xi_{m,\alpha}\mathbb{E}\left[\frac{1}{\sigma_{m,t}(\boldsymbol{\theta})}\frac{\partial \sigma_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}_{|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}}\right].$$

□

Combining term A and term B expands Equation (1.35) to

$$\begin{aligned}\sqrt{n}(\hat{\xi}_{m,\alpha}(\hat{\boldsymbol{\theta}}_n) - \xi_{m,\alpha}(\boldsymbol{\theta}_0)) = \frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t}(\boldsymbol{\theta}_0) < \xi_{m,\alpha}) - \alpha) \\ + \xi_{m,\alpha}(\boldsymbol{\theta}_0)\mathbb{E}\left[\frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_0)}\frac{\partial \sigma_{m,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}_{|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}}\right]' \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1),\end{aligned} \tag{1.42}$$

which concludes the proof as both terms are consistent estimators. The convergence in distribution from Theorem 1.1 follows from the Central Limit Theorem of Billingsley (1961) for ergodic, stationary (Assumption 1.1), and square integrable martingale differences (Assumption 1.5), applied to the sequence $\binom{\boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\boldsymbol{\eta}_t)}{\mathbb{1}(\eta_{m,t}(\boldsymbol{\theta}_0)<\xi_{m,\alpha})-\alpha)}$, where $\boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\boldsymbol{\eta}_t) = \boldsymbol{\Delta}_{t-1}\operatorname{vec}(\boldsymbol{I}_2 - \boldsymbol{\eta}_t\boldsymbol{\eta}_t')$ is defined as Equation (1.18). We present the formulation of the asymptotic variance from Theorem 1.1 as follows: Using Equation (1.42), we obtain the expansion

$$\begin{aligned}\sqrt{n}(\Delta\hat{\xi}_m(\alpha) - \Delta\xi_m(\alpha)) = -\Delta\xi_m(\alpha)\mathbb{E}\left[\frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_0)}\frac{\partial\sigma_{m,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)\\ - \frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t}\le\xi_{m,\alpha}) - \alpha)\\ + \frac{1}{f(\xi_{m,0.5})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t}\le\xi_{m,0.5}) - 0.5) + o_p(1).\end{aligned}$$

To simplify the result, we define $\mathbb{E}(\boldsymbol{\Delta}_{t-1}) = \Lambda$, $\boldsymbol{W}_\alpha = Cov(\boldsymbol{V}(\eta_t), \mathbb{1}(\eta_{m,t}\le\xi_\alpha) - \alpha))$, $\boldsymbol{\Psi} = \mathbb{E}\left(\boldsymbol{\Delta}_t Var(\boldsymbol{V}(\boldsymbol{\eta}_t))\boldsymbol{\Delta}_t'\right)$, and $\boldsymbol{\Omega}_{\boldsymbol{\xi}} = \mathbb{E}\left[\frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_0)}\frac{\partial\sigma_{m,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]$. Utilizing the expansion derived above, we obtain

$$\begin{aligned}&Cov_{asy}(\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right), \sqrt{n}(\Delta\hat{\xi}_m(\alpha) - \Delta\xi_m(\alpha)))\\ &= -Cov_{asy}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))\boldsymbol{\Omega}_\xi\Delta\xi_m(\alpha)\\ &- \frac{1}{f(\xi_{m,\alpha})}Cov_{asy}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t}\le\xi_\alpha) - \alpha))\\ &+ \frac{1}{f(\xi_{m,0.5})}Cov_{asy}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t}\le\xi_{0.5}) - 0.5))\\ &= -\Delta\xi_m(\alpha)\boldsymbol{\Psi}\boldsymbol{\Omega}_{\boldsymbol{\xi}} - \frac{1}{f(\xi_{m,\alpha})}\Lambda\boldsymbol{W}_\alpha + \frac{1}{f(\xi_{m,0.5})}\Lambda\boldsymbol{W}_{0.5} =: \boldsymbol{\Xi}_{\theta\xi}.\end{aligned}$$

To compute $V_{asy}\left(\sqrt{n}(\Delta\hat{\xi}_m(\alpha) - \Delta\xi_m(\alpha)\right)$, we exploit

$$\begin{aligned}&Cov_{asy}(\frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t}\le\xi_{m,\alpha}) - \alpha), \frac{1}{f(\xi_{m,0.5})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t}\le\xi_{m,0.5}) - 0.5))\\ &= \frac{1}{f(\xi_{m,\alpha})f(\xi_{m,0.5})}\mathbb{E}((\mathbb{1}(\eta_{m,t}\le\xi_\alpha) - \alpha)(\mathbb{1}(\eta_{m,t}\le\xi_{m,0.5}) - 0.5))\\ &= \frac{1}{f(\xi_{m,\alpha})f(\xi_{m,0.5})}\mathbb{E}((\mathbb{1}(\eta_{m,t}\le\xi_\alpha)\mathbb{1}(\eta_{m,t}\le\xi_{m,0.5}) - \alpha\mathbb{1}(\eta_{m,t}\le\xi_{m,0.5}) - 0.5\mathbb{1}(\eta_{m,t}\le\xi_{m,\alpha}) + \frac{1}{2}\alpha))\\ &= \frac{1}{f(\xi_{m,\alpha})f(\xi_{m,0.5})}\mathbb{E}(\mathbb{1}(\eta_{m,t}\le\xi_{m,\alpha})) - \alpha\mathbb{1}(\eta_{m,t}\le\xi_{m,0.5}) - 0.5\mathbb{1}(\eta_{m,t}\le\xi_{m,\alpha}) + \frac{1}{2}\alpha))\\ &= \frac{1}{f(\xi_{m,\alpha})f(\xi_{m,0.5})}(\alpha - \frac{1}{2}\alpha - \frac{1}{2}\alpha + \frac{1}{2}\alpha) = \frac{\alpha}{2f(\xi_{m,\alpha})f(\xi_{m,0.5})}.\end{aligned}$$

Finally, we have:

$$\Upsilon_\alpha := V_{asy}\big(\sqrt{n}(\Delta\hat{\xi}_m(\alpha) - \Delta\xi_m(\alpha)\big) = V_{asy}\big(- \Delta\xi_m(\alpha)\boldsymbol{\Omega_\xi}'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)\big)$$
$$+V_{asy}\big(- \frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} \le \xi_{m,\alpha})\alpha)\big) + V_{asy}\big(\frac{1}{f(\xi_{m,0.5})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} \le \xi_{m,0.5}) - 0.5)\big)$$
$$+\frac{2\Delta\xi_m(\alpha)}{f(\xi_{m,\alpha})}Cov_{asy}\big(\boldsymbol{\Omega_\xi}'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right), \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} \le \xi_{m,\alpha}) - \alpha)\big)$$
$$-\frac{2\Delta\xi_m(\alpha)}{f(\xi_{m,0.5})}Cov_{asy}\big(\boldsymbol{\Omega_\xi}'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right), \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} \le \xi_{m,0.5}) - 0.5)\big)$$
$$-\frac{2}{f(\xi_{m,\alpha})f(\xi_{m,0.5})}Cov_{asy}\big(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} \le \xi_{m,\alpha}) - \alpha), \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} \le \xi_{m,0.5}) - 0.5)\big)$$
$$=\big(\Delta\xi_m(\alpha)\big)^2\boldsymbol{\Omega_\xi}'\boldsymbol{\Psi}\boldsymbol{\Omega_\xi} + \frac{\alpha(1-\alpha)}{f^2(\xi_{m,\alpha})} + \frac{1}{4f^2(\xi_{m,0.5})}$$
$$+ \frac{2\Delta\xi_m(\alpha)}{f(\xi_{m,\alpha})}\boldsymbol{\Omega_\xi}'\boldsymbol{\Lambda}\boldsymbol{W}_\alpha - \frac{2\Delta\xi_m(\alpha)}{f(\xi_{m,0.5})}\boldsymbol{\Omega_\xi}'\boldsymbol{\Lambda}\boldsymbol{W}_{0.5} - \frac{\alpha}{f(\xi_{m,\alpha})f(\xi_{m,0.5})}.$$

Combining all previous results and defining the asymptotic variance of $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$ under Assumptions 1.1–1.7 as $\boldsymbol{\Psi}$, the asymptotic covariance in Theorem 1.1 is given by

$$\boldsymbol{\Sigma}_\alpha := \begin{pmatrix} \boldsymbol{\Psi} & \boldsymbol{\Xi}_{\theta\xi} \\ \boldsymbol{\Xi}_{\theta\xi}' & \Upsilon_\alpha \end{pmatrix}.$$

This concludes the proof. ■

C.2. Technical Details on the Donsker Property of the Check Function. We establish that the class of functions defined by the quantile regression check function in our case, i.e., $\rho_\alpha(y_{m,t}/\sigma_{m,t}(\boldsymbol{\theta}))$, forms a Donsker class under suitable regularity conditions.

Consider the MGARCH model

$$\boldsymbol{y}_t = \boldsymbol{H}_t^{1/2}(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t, \tag{1.43}$$

where $\boldsymbol{y}_t = (y_{m,t}, y_{i,t})'$, $\boldsymbol{\eta}_t = (\eta_{m,t}, \eta_{i,t})'$ is a sequence of i.i.d. $\mathbb{R}^2$-valued variables with $\mathbb{E}(\boldsymbol{\eta}_t) = \boldsymbol{0}$ and $\mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t') = \boldsymbol{I}_2$.

We focus on the function class for the marginal process $y_{m,t}$

$$\mathcal{F} = \left\{\rho_\alpha\left(\frac{y_{m,t}}{\sigma_{m,t}(\boldsymbol{\theta})}\right) : \boldsymbol{\theta} \in \Theta\right\},$$

where $\Theta \subset \mathbb{R}^d$ is a compact parameter space and $\sigma_{m,t} : \Theta \to (0, \infty)$ is the volatility function.

Theorem 1.3. *Under Assumptions 1.1–1.7, the function class*

$$\mathcal{F} = \left\{\rho_\alpha\left(\frac{y_{m,t}}{\sigma_{m,t}(\boldsymbol{\theta})}\right) : \boldsymbol{\theta} \in \Theta\right\}$$

is Donsker.

Proof of Theorem 1.3

We establish the Donsker property through the bracketing entropy approach following Van der Vaart (2000).

Lemma 1.1 (Uniform Volatility Bounds)**.** *Under Assumptions 1-8 and the specific structure of the MGARCH model, there exist constants* $0 < c \leq C < \infty$ *such that*

$$c \leq \sigma_{m,t}(\boldsymbol{\theta}) \leq C \quad \text{for all } \boldsymbol{\theta} \in \Theta \text{ and all } t.$$

Proof. The lower bound follows from the positive definiteness of $\boldsymbol{H}_t(\boldsymbol{\theta})$ and the parameter restrictions in Θ that ensure the intercept matrix is positive definite. The upper bound follows from the stationarity (Assumption 1.1) and the finite moment conditions (Assumption 1.5) which prevent volatility explosion. □

Step 1: Lipschitz Properties

Lemma 1.2 (Lipschitz continuity of check function)**.** *For any* $\alpha \in (0,1)$ *and* $u, v \in \mathbb{R}$ *we have*

$$|\rho_\alpha(u) - \rho_\alpha(v)| \leq \max(\alpha, 1-\alpha)|u - v|.$$

Proof. We use the convexity argument that ρ_α is convex with subgradients in $[\alpha - 1, \alpha]$. Hence, for any $u, v \in \mathbb{R}$, we have

$$|\rho_\alpha(u) - \rho_\alpha(v)| \leq \max(|\alpha - 1|, |\alpha|)|u - v| = \max(\alpha, 1-\alpha)|u - v|.$$

□

Lemma 1.3 (Lipschitz continuity of reciprocal scale)**.** *Under Lemma 1.1, the function* $1/\sigma_{m,t}(\boldsymbol{\theta})$ *is Lipschitz continuous on* Θ*.*

Proof. Since $\sigma_{m,t}$ is continuously differentiable on the compact set Θ (Assumption 1.6) , it is Lipschitz continuous. Let L_σ be its Lipschitz constant.

For any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ we have

$$\begin{aligned}\left|\frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_1)} - \frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_2)}\right| &= \frac{|\sigma_{m,t}(\boldsymbol{\theta}_2) - \sigma_{m,t}(\boldsymbol{\theta}_1)|}{\sigma_{m,t}(\boldsymbol{\theta}_1)\sigma_{m,t}(\boldsymbol{\theta}_2)} \\ &\leq \frac{L_\sigma \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|}{c^2}.\end{aligned}$$

Thus $1/\sigma_{m,t}(\boldsymbol{\theta})$ is Lipschitz with constant L_σ/c^2. □

Lemma 1.4 (Lipschitz continuity of function class)**.** *For any* $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ *we have*

$$\left|\rho_\alpha\left(\frac{y_{m,t}}{\sigma_{m,t}(\boldsymbol{\theta}_1)}\right) - \rho_\alpha\left(\frac{y_{m,t}}{\sigma_{m,t}(\boldsymbol{\theta}_2)}\right)\right| \leq \frac{L_\sigma}{c^2}|y_{m,t}|\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

Proof. Using Lemma 1.2 and Lemma 1.3 we have

$$\begin{aligned}
&\left|\rho_\alpha\left(\frac{y_{m,t}}{\sigma_{m,t}(\boldsymbol{\theta}_1)}\right) - \rho_\alpha\left(\frac{y_{m,t}}{\sigma_{m,t}(\boldsymbol{\theta}_2)}\right)\right| \\
&\leq \left|\frac{y_{m,t}}{\sigma_{m,t}(\boldsymbol{\theta}_1)} - \frac{y_{m,t}}{\sigma_{m,t}(\boldsymbol{\theta}_2)}\right| \\
&= |y_{m,t}|\left|\frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_1)} - \frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_2)}\right| \\
&\leq \frac{L_\sigma}{c^2}|y_{m,t}|\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.
\end{aligned}$$

□

Step 2: Envelope Function

Lemma 1.5 (Square-integrable envelope)**.** *The function* $F(y_{m,t}) = \frac{|y_{m,t}|}{c}$ *is a square-integrable envelope for* $\mathcal{F}$*.*

Proof. For any $\boldsymbol{\theta} \in \Theta$, we have

$$\left|\rho_\alpha\left(\frac{y_{m,t}}{\sigma_{m,t}(\boldsymbol{\theta})}\right)\right| \leq \left|\frac{y_{m,t}}{\sigma_{m,t}(\boldsymbol{\theta})}\right| \leq \frac{|y_{m,t}|}{c}.$$

Thus $F(y_{m,t}) = \frac{|y_{m,t}|}{c}$ dominates $\mathcal{F}$. Using the GARCH structure $y_{m,t} = \sigma_{m,t}(\boldsymbol{\theta}_0)\eta_{m,t}$ and the boundedness $\sigma_{m,t}(\boldsymbol{\theta}_0) \leq C$, we have

$$|y_{m,t}| \leq C|\eta_{m,t}|.$$

Therefore,

$$F(y_{m,t}) = \frac{|y_{m,t}|}{c} \leq \frac{C}{c}|\eta_{m,t}|.$$

By Assumption 1.7, $\mathbb{E}[\eta_{m,t}^4] < \infty$ implying $\mathbb{E}[\eta_{m,t}^2] < \infty$. Thus,

$$\mathbb{E}[F(y_{m,t})^2] \leq \frac{C^2}{c^2}\mathbb{E}[\eta_{m,t}^2] < \infty.$$

□

Step 3: Entropy Control

Since Θ is compact in $\mathbb{R}^d$, there exists a $K > 0$ such that the covering number satisfies

$$N(\epsilon, \Theta, \|\cdot\|) \leq K\left(\frac{1}{\epsilon}\right)^d.$$

By Lemma 1.4, the function class $\mathcal{F}$ is Lipschitz in parameter with random Lipschitz constant $M(y_{m,t}) = \frac{L_\sigma}{c^2}|y_{m,t}|$. Since $\mathbb{E}[|y_{m,t}|] \leq C\mathbb{E}[|\eta_{m,t}|] < \infty$, we can bound the covering number for $\mathcal{F}$, thus

$$N(\epsilon, \mathcal{F}, L_2(P)) \leq N\left(\frac{\epsilon}{M}, \Theta, \|\cdot\|\right) \leq K\left(\frac{M}{\epsilon}\right)^d,$$

where M is a constant related to the moment bound.

Step 4: Bracketing Entropy and Donsker Conclusion

The bracketing entropy satisfies

$$\log N_{[]}(\epsilon, \mathcal{F}, L_2(P)) \leq C \log\left(\frac{1}{\epsilon}\right)$$

for some constant $C > 0$. Therefore, the entropy integral converges, thus

$$\int_0^1 \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon < \infty.$$

By Theorem 19.14 of Van der Vaart (2000), this finite entropy integral condition implies that $\mathcal{F}$ is Donsker.

C.3. CoVaR Estimator and Asymptotic Properties. We have shown in Appendix B.2 (Eq. 1.30) that $CoVaR_{i,t}(\beta,\alpha) = \sigma_{i,t}\sqrt{1-\rho_{i,t}^2}\xi_{i,\beta} + \sigma_{i,t}\rho_{i,t}\xi_{m,\alpha}$. To estimate this expression semi-parametrically, it suffices to estimate $\xi_{i,\beta}$, the empirical β-quantile of the estimated innovation $\hat{\eta}_{i,t}$, similarly as we have done for $\xi_{m,\alpha}$. The vector of estimated innovations at time t is obtained as $\hat{\boldsymbol{\eta}}_t = \boldsymbol{H}_t^{-1/2}(\hat{\boldsymbol{\theta}}_n)\,\boldsymbol{y}_t$. We then extract $\hat{\eta}_{i,t}$ from $\hat{\boldsymbol{\eta}}_t$, for $t = 1, \ldots, n$, and compute the sample β-quantile of its empirical distribution $\hat{F}_{\eta_i}$, defined by $\hat{\xi}_{i,\beta} = \inf\{x : \hat{F}_{\eta_i}(x) \geq \beta\}$ and $\hat{F}_{\eta_i}(x) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}(\hat{\eta}_{i,t} \leq x)$. All other estimated quantities ($\hat{\sigma}_{i,t}$, $\hat{\rho}_{i,t}$, $\hat{\xi}_{m,\alpha}$, etc.) are defined as in Section 1.2.2, so that the estimate for the (β,α)-CoVaR at time t is obtained as $\widehat{CoVaR}_{i,t}(\beta,\alpha) = \hat{\sigma}_{i,t}\sqrt{1-\hat{\rho}_{i,t}^2}\hat{\xi}_{i,\beta} + \hat{\sigma}_{i,t}\hat{\rho}_{i,t}\hat{\xi}_{m,\alpha}$.

Theorem 1.4 establishes the asymptotic properties of the vector $(\hat{\boldsymbol{\theta}}_n', \hat{\xi}_{m,\alpha}, \hat{\xi}_{i,\beta})'$. Conditional confidence intervals for $CoVaR_{i,n+1}(\beta,\alpha)$ can be readily constructed by following the same mean value expansion procedure as for MES and ΔCoVaR in Appendix C.5.

Theorem 1.4 (Components of CoVaR)**.** *Suppose Assumptions 1.1–1.7 hold, $\eta_{m,t}$ and $\eta_{i,t}$ admit densities f_m and f_i that are continuous and strictly positive in neighborhoods of $\xi_{m,\alpha}$ and $\xi_{i,\beta}$, respectively, and the model $\boldsymbol{y}_t = \boldsymbol{H}_t^{1/2}(\boldsymbol{\theta}_0)\,\boldsymbol{\eta}_t$, with the Cholesky decomposition $\boldsymbol{H}_t(\boldsymbol{\theta}) = \boldsymbol{H}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{H}_t^{1/2}(\boldsymbol{\theta})'$ holds. Then, for fixed quantile levels $\alpha, \beta \in (0,1)$,*

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \hat{\xi}_{m,\alpha} - \xi_{m,\alpha} \\ \hat{\xi}_{i,\beta} - \xi_{i,\beta} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0},\ \boldsymbol{\Sigma}_{\alpha,\beta} := \begin{pmatrix} \boldsymbol{\Psi} & \boldsymbol{\Xi}_{\theta\xi_m} & \boldsymbol{\Xi}_{\theta\xi_i} \\ \boldsymbol{\Xi}_{\theta\xi_m}' & \Upsilon_{m,\alpha} & \Upsilon_{mi,\alpha\beta} \\ \boldsymbol{\Xi}_{\theta\xi_i}' & \Upsilon_{mi,\alpha\beta} & \Upsilon_{i,\beta} \end{pmatrix}\right),$$

where the components are defined as

$$
\begin{aligned}
\boldsymbol{\Xi}_{\theta\xi_m} &= -\frac{1}{f(\xi_{m,\alpha})}\boldsymbol{\Lambda}\boldsymbol{W}_{m,\alpha} - \xi_{m,\alpha}\boldsymbol{\Psi}\,\boldsymbol{\Omega}_{\xi_m},\\
\boldsymbol{\Xi}_{\theta\xi_i} &= -\frac{1}{f(\xi_{i,\beta})}\boldsymbol{\Lambda}\boldsymbol{W}_{i,\beta} - \xi_{i,\beta}\boldsymbol{\Psi}\boldsymbol{\Omega}_{\xi_i},\\
\Upsilon_{m,\alpha} &= \frac{\alpha(1-\alpha)}{f^2(\xi_{m,\alpha})} + \xi_{m,\alpha}^2\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Psi}\boldsymbol{\Omega}_{\xi_m} + 2\frac{\xi_{m,\alpha}}{f(\xi_{m,\alpha})}\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Lambda}\boldsymbol{W}_{m,\alpha},\\
\Upsilon_{i,\beta} &= \frac{\beta(1-\beta)}{f^2(\xi_{i,\beta})} + \xi_{i,\beta}^2\boldsymbol{\Omega}'_{\xi_i}\boldsymbol{\Psi}\boldsymbol{\Omega}_{\xi_i} + 2\frac{\xi_{i,\beta}}{f(\xi_{i,\beta})}\boldsymbol{\Omega}'_{\xi_i}\boldsymbol{\Lambda}\boldsymbol{W}_{i,\beta},\\
\Upsilon_{mi,\alpha\beta} &= \frac{\gamma_{mi,\alpha\beta}}{f(\xi_{m,\alpha})f(\xi_{i,\beta})} + \frac{\xi_{i,\beta}}{f(\xi_{m,\alpha})}\boldsymbol{\Omega}'_{\xi_i}\boldsymbol{\Lambda}\boldsymbol{W}_{m,\alpha}\\
&\quad + \frac{\xi_{m,\alpha}}{f(\xi_{i,\beta})}\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Lambda}\boldsymbol{W}_{i,\beta} + \xi_{m,\alpha}\xi_{i,\beta}\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Psi}\boldsymbol{\Omega}_{\xi_i},
\end{aligned}
$$

where $\boldsymbol{\Lambda} = \mathbb{E}(\boldsymbol{\Delta}_{t-1})$, $\boldsymbol{W}_{m,\alpha} = Cov(\boldsymbol{V}(\boldsymbol{\eta}_t), \mathbb{1}(\eta_{m,t} \le \xi_{m,\alpha}) - \alpha)$, $\boldsymbol{\Psi} = \mathbb{E}\left(\boldsymbol{\Delta}_t\, Var(\boldsymbol{V}(\boldsymbol{\eta}_t))\boldsymbol{\Delta}'_t\right)$, $\boldsymbol{W}_{i,\beta} = Cov(\boldsymbol{V}(\boldsymbol{\eta}_t), \mathbb{1}(\eta_{i,t} \le \xi_{i,\beta}) - \beta)$, $\gamma_{mi,\alpha\beta} = Cov(\mathbb{1}(\eta_{m,t} \le \xi_{m,\alpha}), \mathbb{1}(\eta_{i,t} \le \xi_{i,\beta}))$, $\boldsymbol{\Omega}_{\xi_m} = \mathbb{E}\left[\frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_0)}\frac{\partial\sigma_{m,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]$, $\boldsymbol{\Omega}_{\xi_i} = \mathbb{E}\left[\frac{1}{\sigma_{i,t}(\boldsymbol{\theta}_0)}\frac{\partial\sigma_{i,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]$.

Proof of Theorem 1.4

From Equation (1.42), we have the expansion

$$
\begin{aligned}
\sqrt{n}(\hat{\xi}_{m,\alpha}(\hat{\boldsymbol{\theta}}_n) - \xi_{m,\alpha}(\boldsymbol{\theta}_0)) = &-\frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t}(\boldsymbol{\theta}_0) < \xi_{m,\alpha}) - \alpha)\\
&- \xi_{m,\alpha}(\boldsymbol{\theta}_0)\mathbb{E}\left[\frac{1}{\sigma_{m,t}(\boldsymbol{\theta}_0)}\frac{\partial\sigma_{m,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]'\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1).
\end{aligned}
$$

Following the same line of reasoning, we can also obtain the corresponding expansion for a different firm:

$$
\begin{aligned}
\sqrt{n}(\hat{\xi}_{i,\alpha}(\hat{\boldsymbol{\theta}}_n) - \xi_{i,\alpha}(\boldsymbol{\theta}_0)) = &-\frac{1}{f(\xi_{i,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{i,t}(\boldsymbol{\theta}_0) < \xi_{i,\alpha}) - \alpha)\\
&- \xi_{i,\alpha}(\boldsymbol{\theta}_0)\mathbb{E}\left[\frac{1}{\sigma_{i,t}(\boldsymbol{\theta}_0)}\frac{\partial\sigma_{i,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]'\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1).
\end{aligned}
\tag{1.44}
$$

We derive the asymptotic covariance matrix for the joint distribution of the parameter

estimator and the two quantile estimators. For the market quantile, we use the expansion

$$\begin{aligned}\sqrt{n}(\hat{\xi}_{m,\alpha} - \xi_{m,\alpha}) &= A_m + B_m + o_p(1),\\ A_m &= -\frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha),\\ B_m &= -\xi_{m,\alpha}\boldsymbol{\Omega}'_{\xi_m}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).\end{aligned}$$

For the individual asset quantile, we use the expansion

$$\begin{aligned}\sqrt{n}(\hat{\xi}_{i,\beta} - \xi_{i,\beta}) &= A_i + B_i + o_p(1),\\ A_i &= -\frac{1}{f(\xi_{i,\beta})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{i,t} < \xi_{i,\beta}) - \beta),\\ B_i &= -\xi_{i,\beta}\boldsymbol{\Omega}'_{\xi_i}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).\end{aligned}$$

For the parameter estimator, we use

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Delta_{t-1}\boldsymbol{V}(\boldsymbol{\eta}_t) + o_p(1).$$

1. Covariance between $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ and $\sqrt{n}(\hat{\xi}_{m,\alpha} - \xi_{m,\alpha})$

The covariance can be written as

$$\begin{aligned}\boldsymbol{\Xi}_{\theta\xi_m} &= \mathrm{Cov}_{asy}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \sqrt{n}(\hat{\xi}_{m,\alpha} - \xi_{m,\alpha})\right)\\ &= \mathrm{Cov}_{asy}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), A_m + B_m\right)\\ &= \mathrm{Cov}_{asy}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), A_m\right) + \mathrm{Cov}_{asy}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), B_m\right),\end{aligned}$$

with first term

$$\begin{aligned}&\mathrm{Cov}_{asy}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), A_m\right)\\ &= \mathrm{Cov}_{asy}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Delta_{t-1}\boldsymbol{V}(\boldsymbol{\eta}_t), -\frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha)\right)\\ &= -\frac{1}{f(\xi_{m,\alpha})}\boldsymbol{\Lambda}\mathrm{Cov}\left(\boldsymbol{V}(\boldsymbol{\eta}_t), \mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha\right)\\ &= -\frac{1}{f(\xi_{m,\alpha})}\boldsymbol{\Lambda}\boldsymbol{W}_{m,\alpha},\end{aligned}$$

and second term

$$\begin{aligned}&\mathrm{Cov}_{asy}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), B_m\right)\\ &= \mathrm{Cov}_{asy}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), -\xi_{m,\alpha}\boldsymbol{\Omega}_{\boldsymbol{\xi_m}}{}'\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right)\\ &= -\xi_{m,\alpha}\mathrm{Var}_{asy}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))\boldsymbol{\Psi}\boldsymbol{\Omega}_{\xi_m}\\ &= -\xi_{m,\alpha}\boldsymbol{\Psi}\boldsymbol{\Omega}_{\xi_m}.\end{aligned}$$

Thus, we have

$$\boldsymbol{\Xi}_{\theta\xi_m} = -\frac{1}{f(\xi_{m,\alpha})}\boldsymbol{\Lambda}\boldsymbol{W}_{m,\alpha} - \xi_{m,\alpha}\boldsymbol{\Psi}\,\boldsymbol{\Omega}_{\boldsymbol{\xi_m}},$$

and similarly

$$\boldsymbol{\Xi}_{\theta\xi_i} = -\frac{1}{f(\xi_{i,\beta})}\boldsymbol{\Lambda}\boldsymbol{W}_{i,\beta} - \xi_{i,\beta}\boldsymbol{\Psi}\boldsymbol{\Omega}_{\xi_i}.$$

2. Variance of $\sqrt{n}(\hat{\xi}_{m,\alpha} - \xi_{m,\alpha})$

The variance can be written as

$$\begin{aligned}\Upsilon_{m,\alpha} &= \mathrm{Var}_{asy}(A_m + B_m)\\ &= \mathrm{Var}(A_m) + \mathrm{Var}(B_m) + 2\mathrm{Cov}(A_m, B_m),\end{aligned}$$

with first term

$$\begin{aligned}\mathrm{Var}(A_m) &= \mathrm{Var}\left(-\frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha)\right)\\ &= \frac{1}{f^2(\xi_{m,\alpha})}\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha)\right)\\ &= \frac{1}{f^2(\xi_{m,\alpha})}\alpha(1-\alpha),\end{aligned}$$

second term

$$\begin{aligned}\mathrm{Var}(B_m) &= \mathrm{Var}\left(-\xi_{m,\alpha}\boldsymbol{\Omega}'_{\xi_m}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right)\\ &= \xi^2_{m,\alpha}\boldsymbol{\Omega}'_{\xi_m}\mathrm{Var}_{asy}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))\boldsymbol{\Omega}_{\xi_m}\\ &= \xi^2_{m,\alpha}\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Psi}\boldsymbol{\Omega}_{\xi_m}\end{aligned}$$

and third term

$$\begin{aligned}\mathrm{Cov}(A_m, B_m) &= \mathbb{E}[A_m B_m] \quad (\text{since } \mathbb{E}[A_m] = \mathbb{E}[B_m] = 0)\\ &= \mathbb{E}\left[\left(-\frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha)\right)\left(-\xi_{m,\alpha}\boldsymbol{\Omega}'_{\xi_m}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right)\right]\\ &= \frac{\xi_{m,\alpha}}{f(\xi_{m,\alpha})}\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha)\right)\left(\boldsymbol{\Omega}'_{\xi_m}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right)\right].\end{aligned}$$

Substituting the expansion for $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, the last term becomes

$$\frac{\xi_{m,\alpha}}{f(\xi_{m,\alpha})}\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Lambda}\mathbb{E}\left[(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha)\boldsymbol{V}(\boldsymbol{\eta}_t)\right] = \frac{\xi_{m,\alpha}}{f(\xi_{m,\alpha})}\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Lambda}\boldsymbol{W}_{m,\alpha}.$$

Thus,

$$\Upsilon_{m,\alpha} = \frac{\alpha(1-\alpha)}{f^2(\xi_{m,\alpha})} + \xi_{m,\alpha}^2 \boldsymbol{\Omega}'_{\xi_m} \boldsymbol{\Psi} \boldsymbol{\Omega}_{\xi_m} + 2\frac{\xi_{m,\alpha}}{f(\xi_{m,\alpha})} \boldsymbol{\Omega}'_{\xi_m} \boldsymbol{\Lambda} \boldsymbol{W}_{m,\alpha}$$

and similarly

$$\Upsilon_{i,\beta} = \frac{\beta(1-\beta)}{f^2(\xi_{i,\beta})} + \xi_{i,\beta}^2 \boldsymbol{\Omega}'_{\xi_i} \boldsymbol{\Psi} \boldsymbol{\Omega}_{\xi_i} + 2\frac{\xi_{i,\beta}}{f(\xi_{i,\beta})} \boldsymbol{\Omega}'_{\xi_i} \boldsymbol{\Lambda} \boldsymbol{W}_{i,\beta}.$$

3. Covariance between $\sqrt{n}(\hat{\xi}_{m,\alpha} - \xi_{m,\alpha})$ and $\sqrt{n}(\hat{\xi}_{i,\beta} - \xi_{i,\beta})$

The covariance can be written as

$$\begin{aligned}\Upsilon_{mi,\alpha\beta} &= \mathrm{Cov}_{asy}(A_m + B_m, A_i + B_i) \\ &= \mathrm{Cov}(A_m, A_i) + \mathrm{Cov}(A_m, B_i) + \mathrm{Cov}(B_m, A_i) + \mathrm{Cov}(B_m, B_i).\end{aligned}$$

Under the independence of $\eta_{m,t}$ and $\eta_{i,t}$, we have for the first term

$$\begin{aligned}&\mathrm{Cov}(A_m, A_i) = \mathbb{E}[A_m A_i] \\ &= \mathbb{E}\left[\left(-\frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha)\right) \times \left(-\frac{1}{f(\xi_{i,\beta})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}(\eta_{i,t} < \xi_{i,\beta}) - \beta)\right)\right] \\ &= \frac{1}{f(\xi_{m,\alpha})f(\xi_{i,\beta})}\mathbb{E}\left[(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha)(\mathbb{1}(\eta_{i,t} < \xi_{i,\beta}) - \beta)\right] \\ &= \frac{1}{f(\xi_{m,\alpha})f(\xi_{i,\beta})}\mathbb{E}\left[(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha)\right]\mathbb{E}\left[(\mathbb{1}(\eta_{i,t} < \xi_{i,\beta}) - \beta)\right] = 0,\end{aligned}$$

the second term

$$\begin{aligned}\mathrm{Cov}(A_m, B_i) &= \mathbb{E}[A_m B_i] \\ &= \mathbb{E}\left[\left(-\frac{1}{f(\xi_{m,\alpha})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha\big)\right)\left(-\xi_{i,\beta}\,\boldsymbol{\Omega}'_{\xi_i}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right)\right] \\ &= \frac{\xi_{i,\beta}}{f(\xi_{m,\alpha})}\,\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha\big)\right)\left(\boldsymbol{\Omega}'_{\xi_i}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right)\right] \\ &= \frac{\xi_{i,\beta}}{f(\xi_{m,\alpha})}\,\boldsymbol{\Omega}'_{\xi_i}\boldsymbol{\Lambda}\,\mathbb{E}\left[\big(\mathbb{1}(\eta_{m,t} < \xi_{m,\alpha}) - \alpha\big)\boldsymbol{V}(\boldsymbol{\eta}_t)\right] \\ &= \frac{\xi_{i,\beta}}{f(\xi_{m,\alpha})}\,\boldsymbol{\Omega}'_{\xi_i}\boldsymbol{\Lambda}\boldsymbol{W}_{m,\alpha},\end{aligned}$$

the third term

$$\begin{aligned}\mathrm{Cov}(B_m, A_i) &= \mathbb{E}[B_m A_i] \\ &= \mathbb{E}\left[\left(-\xi_{m,\alpha}\,\boldsymbol{\Omega}'_{\xi_m}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right)\left(-\frac{1}{f(\xi_{i,\beta})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\mathbb{1}(\eta_{i,t} < \xi_{i,\beta}) - \beta\big)\right)\right] \\ &= \frac{\xi_{m,\alpha}}{f(\xi_{i,\beta})}\,\mathbb{E}\left[\left(\boldsymbol{\Omega}'_{\xi_m}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\mathbb{1}(\eta_{i,t} < \xi_{i,\beta}) - \beta\big)\right)\right] \\ &= \frac{\xi_{m,\alpha}}{f(\xi_{i,\beta})}\,\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Lambda}\,\mathbb{E}\left[\big(\mathbb{1}(\eta_{i,t} < \xi_{i,\beta}) - \beta\big)\boldsymbol{V}(\boldsymbol{\eta}_t)\right] \\ &= \frac{\xi_{m,\alpha}}{f(\xi_{i,\beta})}\,\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Lambda}\boldsymbol{W}_{i,\beta},\end{aligned}$$

and the fourth term

$$\begin{aligned}\mathrm{Cov}(B_m, B_i) &= \mathbb{E}[B_m B_i] \\ &= \mathbb{E}\left[\left(-\xi_{m,\alpha}\,\boldsymbol{\Omega}'_{\xi_m}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right)\left(-\xi_{i,\beta}\,\boldsymbol{\Omega}'_{\xi_i}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right)\right] \\ &= \xi_{m,\alpha}\xi_{i,\beta}\,\boldsymbol{\Omega}'_{\xi_m}\mathbb{E}\Big[\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)'\Big]\,\boldsymbol{\Omega}_{\xi_i} \\ &= \xi_{m,\alpha}\xi_{i,\beta}\,\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Psi}\boldsymbol{\Omega}_{\xi_i}.\end{aligned}$$

Thus:

$$\Upsilon_{mi,\alpha\beta} = \frac{\xi_{i,\beta}}{f(\xi_{m,\alpha})}\boldsymbol{\Omega}'_{\xi_i}\boldsymbol{\Lambda}\boldsymbol{W}_{m,\alpha} + \frac{\xi_{m,\alpha}}{f(\xi_{i,\beta})}\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Lambda}\boldsymbol{W}_{i,\beta} + \xi_{m,\alpha}\xi_{i,\beta}\boldsymbol{\Omega}'_{\xi_m}\boldsymbol{\Psi}\boldsymbol{\Omega}_{\xi_i}.$$

This completes the derivation of all components of the asymptotic covariance matrix for CoVaR. ■

C.4. Proof of Theorem 1.2. The proof follows a similar argument as before. First, we derive the expansion of $\sqrt{n}(\hat{\mu}_{m,\alpha} - \mu_{m,\alpha})$ into two parts: one that remains unaffected by the estimation of $\boldsymbol{\theta}_0$ and another depends on parameter estimation from the chosen model. We then construct the covariance matrix between these two parts. For the expansion of $\sqrt{n}(\hat{\mu}_{m,\alpha} - \mu_{m,\alpha})$ we have

$$\sqrt{n}(\hat{\mu}_{m,\alpha} - \mu_{m,\alpha}) = \frac{\sqrt{n}}{\alpha}\left(\int_{-\infty}^{\hat{\xi}_{m,\alpha}} x d\hat{F}_n(x) - \int_{-\infty}^{\xi_{m,\alpha}} x dF(x)\right) \tag{1.45}$$

$$= \frac{\sqrt{n}}{\alpha}\left(\hat{\xi}_{m,\alpha}\alpha - \int_{-\infty}^{\hat{\xi}_{m,\alpha}} \hat{F}_n(x)dx - \xi_{m,\alpha}\alpha + \int_{-\infty}^{\xi_{m,\alpha}} F(x)dx\right) \tag{1.46}$$

$$= \frac{\sqrt{n}}{\alpha}\left(\underbrace{(\hat{\xi}_{m,\alpha} - \xi_{m,\alpha})\alpha}_{A} - \underbrace{\int_{-\infty}^{\xi_{m,\alpha}} (\hat{F}_n(x) - F_n(x))dx}_{B} + \underbrace{\int_{-\infty}^{\xi_{m,\alpha}} (F(x) - F_n(x))dx}_{C} + \underbrace{\int_{\xi_{m,\alpha}}^{\hat{\xi}_{m,\alpha}} \hat{F}_n(x)dx}_{D}\right). \tag{1.47}$$

Using integration by parts from (1.45) - (1.46), and subtracting and adding $\int_{-\infty}^{\xi_{m,\alpha}} F_n(x)dx$ to (1.46), we obtain (1.47). Term A and D will vanish by the previous Theorem 1.1 as $n \to \infty$ under the same assumptions. Applying expansion (1.38) to Term B, we obtain the expansion

$$\sqrt{n}(\hat{\mu}_{m,\alpha} - \mu_{m,\alpha}) = \sum_{t=1}^{n} \frac{(\eta_{m,t} - \xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t} \leq \xi_{m,\alpha}) + \int_{-\infty}^{\xi_{m,\alpha}} F(x)dx}{\sqrt{n}\alpha} - \frac{1}{2}\mu_{m,\alpha}\boldsymbol{\Omega_{\mu}}'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) + o_p(1).$$

The convergence in distribution from Theorem 1.2 follows from the Central Limit Theorem of Billingsley (1961) for ergodic, stationary (Assumption 1.1), and square-integrable martingale differences (Assumption 1.5), applied to the sequence in the second bracket as follows:

$$\begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \hat{\mu}_{m,\alpha} - \mu_{m,\alpha} \end{pmatrix} = \frac{1}{n}\sum_{t=1}^{n} \begin{pmatrix} \Delta_{t-1}\boldsymbol{V}(\boldsymbol{\eta}_t) \\ \frac{(\eta_{m,t} - \xi_{m,\alpha})\mathbb{1}(\eta_{m,t} \leq \xi_{m,\alpha}) + \int_{-\infty}^{\xi_{m,\alpha}} F(x)\,x}{\alpha} - \frac{1}{2}\mu_{m,\alpha}\boldsymbol{\Omega}_{\mu}'\Delta_{t-1}\boldsymbol{V}(\boldsymbol{\eta}_t) \end{pmatrix} + o_p(1),$$

where

$$\boldsymbol{\Omega}_{\mu}' = \mathbb{E}\left[\frac{1}{\sigma_{m,t}^2(\boldsymbol{\theta}_0)}\frac{\partial\sigma_{m,t}^2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]' = \left(\mathbb{E}\left[\frac{1}{\sigma_{m,t}^2(\boldsymbol{\theta}_0)}\frac{\partial}{\partial\boldsymbol{\theta}_{\sigma_m}}\sigma_{m,t}^2\right] \quad \mathbb{E}\left[\frac{1}{\sigma_{m,t}^2(\boldsymbol{\theta}_0)}\frac{\partial}{\partial\boldsymbol{\theta}_{\sigma_i}}\sigma_{m,t}^2\right] \quad \mathbb{E}\left[\frac{1}{\sigma_{m,t}^2(\boldsymbol{\theta}_0)}\frac{\partial}{\partial\boldsymbol{\theta}_{\rho_i}}\sigma_{m,t}^2\right]\right).$$

We have

$$\begin{aligned}
&V_{asy}\left(\sqrt{n}(\hat{\mu}_{m,\alpha} - \mu_{m,\alpha})\right) \\
&\quad = V_{asy}\left(\sum_{t=1}^{n} \frac{(\eta_{m,t} - \xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t} \leq \xi_{m,\alpha}) + \int_{-\infty}^{\xi_{m,\alpha}} F(x)\,x}{\sqrt{n}\,\alpha} - \tfrac{1}{2}\mu_{m,\alpha}\boldsymbol{\Omega}_{\mu}'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)\right) \\
&\quad = V_{asy}\left(\sum_{t=1}^{n} \frac{(\eta_{m,t} - \xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t} \leq \xi_{m,\alpha}) + \int_{-\infty}^{\xi_{m,\alpha}} F(x)\,x}{\sqrt{n}\,\alpha}\right) + V_{asy}\left(\tfrac{1}{2}\mu_{m,\alpha}\boldsymbol{\Omega}_{\mu}'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)\right) \\
&\qquad - 2\,\mathrm{Cov}_{asy}\left(\sum_{t=1}^{n} \frac{(\eta_{m,t} - \xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t} \leq \xi_{m,\alpha}) + \int_{-\infty}^{\xi_{m,\alpha}} F(x)\,x}{\sqrt{n}\,\alpha},\ \tfrac{1}{2}\mu_{m,\alpha}\boldsymbol{\Omega}_{\mu}'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)\right).
\end{aligned}$$

Next, we build the summands of $V_{asy}\left(\sqrt{n}(\hat{\mu}_{m,\alpha} - \mu_{m,\alpha})\right)$. For the first term, we obtain

$$\begin{aligned}
&V_{asy}\left(\sum_{t=1}^{n} \frac{(\eta_{m,t} - \xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t} \leq \xi_{m,\alpha}) + \int_{-\infty}^{\xi_{m,\alpha}} F(x)dx}{\sqrt{n}\alpha}\right) \\
&= \frac{1}{\alpha^2}V\left[(\eta_{m,t} - \xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t} \leq \xi_{m,\alpha})\right] \\
&= \frac{1}{\alpha^2}\left(\mathbb{E}\left[(\eta_{m,t} - \xi_{m,\alpha})^2\,\mathbb{1}\,(\eta_{m,t} \leq \xi_{m,\alpha})\right] - \mathbb{E}\left[(\eta_{m,t} - \xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t} \leq \xi_{m,\alpha})\right]^2\right) \\
&= \frac{1}{\alpha^2}\left(\int_{-\infty}^{\xi_{m,\alpha}} x^2 f(x)dx - 2\alpha\xi_{m,\alpha}\mu_{m,\alpha} + \alpha\xi_{m,\alpha}^2 - \mu_{m,\alpha}^2\alpha^2 + 2\alpha^2\mu_{m,\alpha}\xi_{m,\alpha} - \alpha^2\xi_{m,\alpha}^2\right) \\
&= \frac{1}{\alpha^2}\left(\int_{-\infty}^{\xi_{m,\alpha}} x^2 f(x)dx + \alpha(1-\alpha)\xi_{m,\alpha}^2 - \alpha^2\mu_{m,\alpha}^2 - 2\alpha(1-\alpha)\mu_{m,\alpha}\xi_{m,\alpha}\right).
\end{aligned}$$

For the second summand, we have

$$V_{asy}\left(\frac{1}{2}\mu_{m,\alpha}\boldsymbol{\Omega_\mu}'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)\right)=\frac{\mu_{m,\alpha}^2}{4}\boldsymbol{\Omega_\mu}'\Psi\boldsymbol{\Omega_\mu}.$$

For the third summand, denoting $\boldsymbol{M}_\alpha = Cov(\boldsymbol{V}(\eta_t),(\eta_{m,t}-\xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t}\le\xi_{m,\alpha}))$, we have,

$$\begin{aligned}
&2Cov_{asy}\left(\frac{1}{2}\mu_{m,\alpha}\boldsymbol{\Omega_\mu}'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right),\sum_{t=1}^{n}\frac{(\eta_{m,t}-\xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t}\le\xi_{m,\alpha})+\int_{-\infty}^{\xi_{m,\alpha}}F(x)dx}{\sqrt{n}\alpha}\right)\\
&=2Cov_{asy}\left(\frac{1}{2}\mu_{m,\alpha}\boldsymbol{\Omega_\mu}'\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right),\frac{\sum_{t=1}^{n}(\eta_{m,t}-\xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t}\le\xi_{m,\alpha})}{\sqrt{n}\alpha}\right)\\
&=\frac{1}{\alpha}\mu_{m,\alpha}\boldsymbol{\Omega_\mu}'Cov_{asy}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0),\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\eta_{m,t}-\xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t}\le\xi_{m,\alpha})\right)\\
&=\frac{\mu_{m,\alpha}}{\alpha}\boldsymbol{\Omega_\mu}'\boldsymbol{\Lambda}\boldsymbol{M}_\alpha.
\end{aligned}$$

We thus end with

$$\begin{aligned}
V_{asy}\big(\sqrt{n}(\hat{\mu}_{m,\alpha}-\mu_{m,\alpha})\big)=\frac{1}{\alpha^2}\left(\int_{-\infty}^{\xi_{m,\alpha}}x^2f(x)\,x+\alpha(1-\alpha)\xi_{m,\alpha}^2-\alpha^2\mu_{m,\alpha}^2-2\alpha(1-\alpha)\mu_{m,\alpha}\xi_{m,\alpha}\right)\\
+\frac{\mu_{m,\alpha}^2}{4}\boldsymbol{\Omega}_\mu'\boldsymbol{\Psi}\,\boldsymbol{\Omega}_\mu-\frac{\mu_{m,\alpha}}{\alpha}\boldsymbol{\Omega}_\mu'\boldsymbol{\Lambda}\,\boldsymbol{M}_\alpha:=\zeta_\alpha.
\end{aligned}$$

Next, we build $Cov_{asy}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0),\sqrt{n}(\hat{\mu}_{m,\alpha}-\mu_{m,\alpha})\right)=:\Phi_{\theta\mu}$, and have

$$\begin{aligned}
&Cov_{asy}(\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right),\sqrt{n}(\hat{\mu}_{m,\alpha}-\mu_{m,\alpha}))\\
=&\frac{1}{\alpha}Cov_{asy}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0),\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\eta_{m,t}-\xi_{m,\alpha})\,\mathbb{1}\,(\eta_{m,t}\le\xi_{m,\alpha}))\\
&-\frac{\mu_{m,\alpha}}{2}Cov_{asy}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0),\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0))\Omega_\mu\\
=&\frac{1}{\alpha}\boldsymbol{\Lambda}\boldsymbol{M}_\alpha-\frac{\mu_{m,\alpha}}{2}\Psi\Omega_\mu=:\Phi_{\theta\mu}.
\end{aligned}$$

Combining all the results obtained thus far and under Assumptions 1.1–1.7 , we define the asymptotic variance of $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)$ as $\boldsymbol{\Psi}$. Consequently, the asymptotic covariance in Theorem 1.2 is given by

$$\overline{\boldsymbol{\Sigma}}_\alpha:=\begin{pmatrix}\boldsymbol{\Psi} & \boldsymbol{\Phi}_{\theta\mu}\\ \boldsymbol{\Phi}_{\theta\mu}' & \zeta_\alpha\end{pmatrix}.$$

This concludes the proof. ■

C.5. Proof of Corollary 1.1

ΔCoVaR. We have $\Delta CoVaR_{i,n+1}\left(\alpha;\boldsymbol{\theta}_0,\Delta\xi_{m,\alpha}\right)=\sigma_{i,n+1}\left(\boldsymbol{\theta}_0\right)\rho_{i,n+1}\left(\boldsymbol{\theta}_0\right)\left(\xi_{m,\alpha}-\xi_{m,0.5}\right)$

from Proposition 1.2. Under the assumptions of Theorem 1.1, for the true parameter $\Delta CoVaR_{i,n+1}(\alpha;\boldsymbol{\theta}_0,\Delta\xi_{m,\alpha})$ and its estimator $\Delta CoVaR_{i,n+1}(\alpha;\hat{\boldsymbol{\theta}}_n,\Delta\hat{\xi}_{m,\alpha})$, and assuming that $\Delta CoVaR_{i,n+1}(\alpha;\boldsymbol{\theta},\Delta\xi)$ is continuously differentiable in $(\boldsymbol{\theta}',\Delta\xi)$ in a neighborhood of $(\boldsymbol{\theta}_0',\Delta\xi_{m,\alpha})$, a first-order mean value expansion gives

$$\sqrt{n}\left(\Delta CoVaR_{i,n+1}(\alpha;\hat{\boldsymbol{\theta}}_n,\Delta\hat{\xi}_{m,\alpha})-\Delta CoVaR_{i,n+1}(\alpha;\boldsymbol{\theta}_0,\Delta\xi_{m,\alpha})\right)=$$
$$\left[\frac{\partial\Delta CoVaR_{i,n+1}(\alpha;\boldsymbol{\theta},\Delta\xi)}{\partial(\boldsymbol{\theta}',\Delta\xi)}\right]_{(\boldsymbol{\theta}_0,\Delta\xi_{m,\alpha})}\begin{pmatrix}\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\\ \sqrt{n}(\Delta\hat{\xi}_{m,\alpha}-\Delta\xi_{m,\alpha})\end{pmatrix}+o_p(1).$$

Consequently, by the asymptotic normality result in Theorem 1.1, we have,

$$\sqrt{n}\left(\Delta CoVaR_{i,n+1}(\alpha;\hat{\boldsymbol{\theta}}_n,\Delta\hat{\xi}_{m,\alpha})-\Delta CoVaR_{i,n+1}(\alpha;\boldsymbol{\theta}_0,\Delta\xi_{m,\alpha})\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\gamma^{\xi}_{\alpha,n+1}),$$

where $\gamma^{\xi}_{\alpha,n+1}=\left[\frac{\partial\Delta CoVaR_{i,n+1}(\alpha;\boldsymbol{\theta},\Delta\xi)}{\partial(\boldsymbol{\theta}',\Delta\xi)}\Big|_{(\boldsymbol{\theta}_0,\Delta\xi_{m,\alpha})}\right]\boldsymbol{\Sigma}_{\alpha}\left[\frac{\partial\Delta CoVaR_{i,n+1}(\alpha;\boldsymbol{\theta},\Delta\xi))}{\partial(\boldsymbol{\theta}',\Delta\xi)}\Big|_{(\boldsymbol{\theta}_0,\Delta\xi_{m,\alpha})}\right]'$ with $\boldsymbol{\Sigma}_{\alpha}$ from Theorem 1.1. ■

MES. We have $MES_{i,n+1}(\alpha;\boldsymbol{\theta}_0,\mu_{m,\alpha})=\sigma_{i,n+1}(\boldsymbol{\theta}_0)\,\rho_{i,n+1}(\boldsymbol{\theta}_0)\,\mu_{m,\alpha}$ from Proposition 1.1. Under the assumptions of Theorem 1.2, for the true parameter $MES_{i,n+1}(\alpha;\boldsymbol{\theta}_0,\mu_{m,\alpha})$ and its estimator $MES_{i,n+1}(\alpha;\hat{\boldsymbol{\theta}}_n,\hat{\mu}_{m,\alpha})$, and assuming that $MES_{i,n+1}(\alpha;\boldsymbol{\theta},\mu)$ is continuously differentiable in $(\boldsymbol{\theta}',\mu)$ in a neighborhood of $(\boldsymbol{\theta}_0',\mu_{m,\alpha})$, a first-order mean value expansion gives

$$\sqrt{n}\left(MES_{i,n+1}(\alpha;\hat{\boldsymbol{\theta}}_n,\hat{\mu}_{m,\alpha})-MES_{i,n+1}(\alpha;\boldsymbol{\theta}_0,\mu_{m,\alpha})\right)=$$
$$\left[\frac{\partial MES_{i,n+1}(\alpha;\boldsymbol{\theta},\mu)}{\partial(\boldsymbol{\theta}',\mu)}\right]_{(\boldsymbol{\theta}_0,\mu_{m,\alpha})}\begin{pmatrix}\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\\ \sqrt{n}(\hat{\mu}_{m,\alpha}-\mu_{m,\alpha})\end{pmatrix}+o_p(1).$$

Consequently, by the asymptotic normality result in Theorem 1.2, we have,

$$\sqrt{n}\left(MES_{i,n+1}(\alpha;\hat{\boldsymbol{\theta}}_n,\hat{\mu}_{m,\alpha})-MES_{i,n+1}(\alpha;\boldsymbol{\theta}_0,\mu_{m,\alpha})\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\gamma^{\mu}_{\alpha,n+1}),$$

where $\gamma^{\mu}_{\alpha,n+1}=\left[\frac{\partial MES_{i,n+1}(\alpha;\boldsymbol{\theta},\mu)}{\partial(\boldsymbol{\theta}',\mu)}\Big|_{(\boldsymbol{\theta}_0,\mu_{m,\alpha})}\right]\overline{\boldsymbol{\Sigma}}_{\alpha}\left[\frac{\partial MES_{i,n+1}(\alpha;\boldsymbol{\theta},\mu))}{\partial(\boldsymbol{\theta}',\mu)}\Big|_{(\boldsymbol{\theta}_0,\mu_{m,\alpha})}\right]'$ with $\overline{\boldsymbol{\Sigma}}_{\alpha}$ from Theorem 1.2. ■

C.6. Feasible Asymptotic Variance-Covariance Matrices in Theorems 1.1 and 1.2. We show how to consistently estimate the elements of the variance-covariance matrices in Theorems 1.1 and 1.2. The quantities entering the asymptotic variance-covariance formulas are obtained by replacing theoretical expectations or covariances with empirical counterparts, or consistent plug-in estimates. Note that we let the summations for empirical counterparts start at $t=1$, assuming the availability of a presample. If not the case, the sums can start at a higher index without affecting the asymptotic results. The estimated quantities are detailed below.

Elements for $\boldsymbol{\Psi}$ in Theorems 1.1 and 1.2. We have $\hat{\boldsymbol{J}}^{*}=\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\left(\boldsymbol{y}_t'\boldsymbol{H}_t^{-1}(\hat{\boldsymbol{\theta}}_n)\boldsymbol{y}_t+\log|\boldsymbol{H}_t(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})|\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$. Let $\hat{\boldsymbol{\Delta}}_{t-1}=-\hat{\boldsymbol{J}}^{*-1}\frac{\partial\,\mathrm{vec}'\boldsymbol{H}_t(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})}{\partial\boldsymbol{\theta}}\{\boldsymbol{H}_t^{-1/2}(\hat{\boldsymbol{\theta}}_n)\otimes\boldsymbol{H}_t^{-1/2}(\hat{\boldsymbol{\theta}}_n)\}'$. We compute $\hat{\boldsymbol{\Lambda}}=\frac{1}{n}\sum_{t=1}^{n}\hat{\boldsymbol{\Delta}}_{t-1}$

and we get $\hat{\boldsymbol{\Psi}} = \frac{1}{n}\sum_{t=1}^{n}\hat{\boldsymbol{\Delta}}_t\left[\frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{V}(\hat{\boldsymbol{\eta}}_t)-\overline{\boldsymbol{V}}\right)\left(\boldsymbol{V}(\hat{\boldsymbol{\eta}}_t)-\overline{\boldsymbol{V}}\right)'\right]\hat{\boldsymbol{\Delta}}_t'$, with $\boldsymbol{V}(\hat{\boldsymbol{\eta}}_t) = \text{vec}\{\boldsymbol{I}_2-\hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t'\}$ and $\overline{\boldsymbol{V}} = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{V}(\hat{\boldsymbol{\eta}}_t)$.

Elements for $\boldsymbol{\Xi}_{\theta_\xi}$ and Υ_α in Theorem 1.1. We estimate $\boldsymbol{\Xi}_{\theta_\xi}$ as $\hat{\boldsymbol{\Xi}}_{\theta_\xi} = -\Delta\hat{\xi}_{m,\alpha}\hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Omega}}_\xi - \frac{1}{\hat{f}_n(\hat{\xi}_{m,\alpha})}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{W}}_\alpha + \frac{1}{\hat{f}_n(\hat{\xi}_{m,0.5})}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{W}}_{0.5}$, with $\hat{\boldsymbol{\Omega}}_\xi = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sigma_{m,t}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})}\frac{\partial\sigma_{m,t}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})}{\partial\boldsymbol{\theta}}$, and $\hat{\boldsymbol{W}}_u = \frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{V}(\hat{\boldsymbol{\eta}}_t)-\overline{\boldsymbol{V}}\right)\left(\mathbf{1}(\hat{\eta}_{m,t}\le\hat{\xi}_{m,u})-u\right)$, where u stands for α or 0.5, and we estimate Υ_α as $\hat{\Upsilon}_\alpha = \Delta\hat{\xi}_{m,\alpha}^2\hat{\boldsymbol{\Omega}}_\xi'\hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Omega}}_\xi + \frac{2\Delta\hat{\xi}_{m,\alpha}}{\hat{f}_n(\hat{\xi}_{m,\alpha})}\hat{\boldsymbol{\Omega}}_\xi'\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{W}}_\alpha - \frac{2\Delta\hat{\xi}_{m,\alpha}}{\hat{f}_n(\hat{\xi}_{m,0.5})}\hat{\boldsymbol{\Omega}}_\xi'\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{W}}_{0.5} + \frac{\alpha(1-\alpha)}{\hat{f}_n^2(\hat{\xi}_{m,\alpha})} + \frac{1}{4\hat{f}_n^2(\hat{\xi}_{m,0.5})} - \frac{\alpha}{\hat{f}_n(\hat{\xi}_{m,\alpha})\hat{f}_n(\hat{\xi}_{m,0.5})}$. The density f is estimated consistently with a standard kernel density estimator applied to the residuals $\hat{\eta}_{m,t}$ as $\hat{f}_n(x) = \frac{1}{nh_n}\sum_{t=1}^{n}K\left(\frac{x-\hat{\eta}_{m,t}}{h_n}\right)$. Consistency holds under the usual conditions $h_n>0$, $h_n\to 0$, $\sqrt{n}h_n^2\to\infty$, and K being a unimodal density, symmetric around zero, and Lipschitz (see Kulperger and Yu, 2005; Gao and Song, 2008, for consistent kernel density estimation of the innovation distribution in GARCH-type models). In the simulation study and empirical application, we use the default values of the `Matlab` function `ksdensity()`, i.e., Gaussian K and a Silverman-type rule of thumb for h_n.

Elements for $\boldsymbol{\Phi}_{\theta\mu}$ and ζ_α in Theorem 1.2. We estimate $\boldsymbol{\Phi}_{\theta\mu}$ as $\hat{\boldsymbol{\Phi}}_{\theta\mu} = \frac{1}{\alpha}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{M}}_\alpha - \frac{\hat{\mu}_{m,\alpha}}{2}\hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Omega}}_\mu$, where we have $\hat{\boldsymbol{\Omega}}_\mu = \frac{1}{n}\sum_{t=1}^{n}\frac{2}{\sigma_{m,t}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})}\frac{\partial\sigma_{m,t}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})}{\partial\boldsymbol{\theta}}$, and $\hat{\boldsymbol{M}}_\alpha = \frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{V}(\hat{\boldsymbol{\eta}}_t)-\overline{\boldsymbol{V}}\right)\left((\hat{\eta}_{m,t}-\hat{\xi}_{m,\alpha})\,\mathbf{1}(\hat{\eta}_{m,t}\le\hat{\xi}_{m,\alpha})-\overline{\nu}\right)$, with $\overline{\nu} = \frac{1}{n}\sum_{t=1}^{n}(\hat{\eta}_{m,t}-\hat{\xi}_{m,\alpha})\,\mathbf{1}(\hat{\eta}_{m,t}\le\hat{\xi}_{m,\alpha})$. Finally, we estimate ζ_α using the following formula $\hat{\zeta}_\alpha = \frac{\hat{\mu}_{m,\alpha}^2}{4}\hat{\boldsymbol{\Omega}}_\mu'\hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Omega}}_\mu - \frac{\hat{\mu}_{m,\alpha}}{\alpha}\hat{\boldsymbol{\Omega}}_\mu'\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{M}}_\alpha + \frac{1}{\alpha^2}\left(\hat{p}_\alpha+\alpha(1-\alpha)\hat{\xi}_{m,\alpha}^2-\alpha^2\hat{\mu}_{m,\alpha}^2-2\alpha(1-\alpha)\hat{\mu}_{m,\alpha}\hat{\xi}_{m,\alpha}\right)$, where $\hat{p}_\alpha = \frac{1}{n}\sum_{t=1}^{n}\hat{\eta}_{m,t}^2\,\mathbf{1}\left(\hat{\eta}_{m,t}\le\hat{\xi}_{m,\alpha}\right)$.

D - Technical Details for ΔCoVaR$^{\leq}$ Benchmark Estimator

D.1. Semi-parametric Estimator for ΔCoVaR$^{\leq}$

The semi-parametric procedure for ΔCoVaR$^{\leq}$ estimation proceeds as follows:

1. Estimate the parameters $\boldsymbol{\theta}_0$ by minimizing Equation (1.12) and compute past variances and correlations ($t\le n$) as well as predicted variances and correlations ($t=n+1$) with $\hat{\sigma}_{m,t}^2\equiv\sigma_{m,t}^2(\hat{\boldsymbol{\theta}}_n)$, $\hat{\sigma}_{i,t}^2\equiv\sigma_{i,t}^2(\hat{\boldsymbol{\theta}}_n)$, and $\hat{\rho}_{i,t}\equiv\rho_{i,t}(\hat{\boldsymbol{\theta}}_n)$. This step is identical to the estimation step 1 in Section 1.2.2 of the main paper.

2. Using the estimated volatilities $\hat{\sigma}_{m,t}^2$, $t=1,\dots,n$, compute the innovations of the market return as $\hat{\eta}_{m,t} = y_{m,t}/\hat{\sigma}_{m,t}$, and compute the α, α_{inf}, and α_{sup}-quantiles of the empirical distribution $\hat{F}_{\eta_m}$ of $\{\hat{\eta}_{m,t}\}_{t=1}^n$ as $\hat{\xi}_{m,\alpha} = \inf\{x:\hat{F}_{\eta_m}(x)\ge\alpha\}$, $\hat{\xi}_{m,\alpha_{\text{inf}}} = \inf\{x:\hat{F}_{\eta_m}(x)\ge\alpha_{\text{inf}}\}$, $\hat{\xi}_{m,\alpha_{\text{sup}}} = \inf\{x:\hat{F}_{\eta_m}(x)\ge\alpha_{\text{sup}}\}$ where $\hat{F}_{\eta_m}(x) = \frac{1}{n}\sum_{t=1}^{n}\mathbb{1}\left(\hat{\eta}_{m,t}\le x\right)$.

3. Consider the two event-type conditions $\hat{\eta}_{m,t}\le\hat{\xi}_{m,\alpha}$ and $\hat{\xi}_{m,\alpha_{\text{inf}}}\le\hat{\eta}_{m,t}\le\hat{\xi}_{m,\alpha_{\text{sup}}}$, and denote by $\mathcal{S}_\alpha = \{\,s:\hat{\eta}_{m,s}\le\hat{\xi}_{m,\alpha}\,\}$ and $\mathcal{S}_{\alpha_{\text{inf}},\alpha_{\text{sup}}} = \{\,s:\hat{\xi}_{m,\alpha_{\text{inf}}}\le\hat{\eta}_{m,s}\le\hat{\xi}_{m,\alpha_{\text{sup}}}\,\}$

the corresponding sets of dates where these conditions hold, respectively. For each observation s belonging to either $\mathcal{S}_\alpha$ or $\mathcal{S}_{\alpha_{\text{inf}},\alpha_{\text{sup}}}$, construct the (devolatilized) firm's innovation as $\hat{\tilde{\eta}}_{i,t}^{(s)} = \hat{\rho}_{i,t}\,\hat{\eta}_{m,s} + \sqrt{1-\hat{\rho}_{i,t}^2}\,\hat{\eta}_{i,s}$. For each date $t = 1, \ldots, n$, this yields a truncated pseudo-sample $\{\,\hat{\tilde{\eta}}_{i,t}^{(s)} : s \in \mathcal{S}\,\}$, where $\mathcal{S}$ stands for $\mathcal{S}_\alpha$ or $\mathcal{S}_{\alpha_{\text{inf}},\alpha_{\text{sup}}}$, whose sizes are approximately $n \times \alpha$ and $n \times (\alpha_{\text{sup}} - \alpha_{\text{inf}})$, respectively.

4. Compute the β-quantiles of the empirical distribution $\hat{F}^{\mathcal{S}}_{\tilde{\eta}_{i,t}}$ of $\{\,\hat{\tilde{\eta}}_{i,t}^{(s)} : s \in \mathcal{S}\,\}$, for $\mathcal{S} \in \{\mathcal{S}_\alpha, \mathcal{S}_{\alpha_{\text{inf}},\alpha_{\text{sup}}}\}$, as $\hat{q}_{\beta,\alpha,t} = \inf\{x : \hat{F}^{\mathcal{S}_\alpha}_{\tilde{\eta}_{i,t}}(x) \geq \beta\}$ and $\hat{q}_{\beta,\alpha_{\text{inf}},\alpha_{\text{sup}},t} = \inf\{x : \hat{F}^{\mathcal{S}_{\alpha_{\text{inf}},\alpha_{\text{sup}}}}_{\tilde{\eta}_{i,t}}(x) \geq \beta\}$ with $\hat{F}^{\mathcal{S}}_{\tilde{\eta}_{i,t}}(x) = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} \mathbb{1}\Big(\hat{\tilde{\eta}}_{i,t}^{(s)} \leq x\Big)$.

5. The corresponding semi-parametric estimator of $\Delta\text{CoVaR}^{\leq}$ is given by

$$\Delta CoVaR^{\leq}_{i,t}(\beta, \hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\xi}}_{m,\alpha}) = \hat{\sigma}_{i,t}\left(\hat{q}_{\beta,\alpha,t} - \hat{q}_{\beta,\alpha_{\text{inf}},\alpha_{\text{sup}},t}\right),$$

with $\hat{\boldsymbol{\xi}}_{m,\alpha} = \left[\hat{\xi}_{m,\alpha}, \hat{\xi}_{m,\alpha_{\text{inf}}}, \hat{\xi}_{m,\alpha_{\text{sup}}}\right]'$.

Remark (Static versus dynamic dependence). When the correlation is constant, i.e., $\rho_{i,t} = \rho_i$ for all t, our framework encompasses the static setting considered by Francq and Zakoïan (2025). In this case, the devolatilized innovation $\tilde{\eta}_{i,t} = \rho_i\,\eta_{m,t} + \sqrt{1-\rho_i^2}\,\eta_{i,t}$ no longer depends on past information $\mathcal{F}_{t-1}$. Consequently, the conditional distributions $F_{\tilde{\eta}_{i,t}|\eta_{m,t}\leq\xi_{m,\alpha}}$ and $F_{\tilde{\eta}_{i,t}|\xi_{m,\alpha_{\text{inf}}}\leq\eta_{m,t}\leq\xi_{m,\alpha_{\text{sup}}}}$ are time-invariant, and the truncated quantiles $q_{\beta,\alpha,t}$ and $q_{\beta,\alpha_{\text{inf}},\alpha_{\text{sup}},t}$ are constant across t. Applying Step 3 of our procedure to the observed innovations $\eta_{m,t}$ and $\eta_{i,t}$ directly yields the realized values of $\tilde{\eta}_{i,t}$, so that no pseudo-sample is generated. While this static CCC structure aligns with the construction of Francq and Zakoïan (2025), exact equivalence of the resulting ΔCoVaR estimates additionally requires that the MGARCH parameters are estimated using an equation-by-equation GARCH approach, with the static correlation ρ_i obtained from the empirical correlation of the estimated innovations. When instead our estimator is based on joint MGARCH estimation, small differences may arise, as illustrated in Panel B of Table 1.2. Finally, when $\rho_{i,t}$ varies over time, our estimator adapts to the evolving dependence structure. At each prediction date t, Step 3 fixes the estimated correlation $\hat{\rho}_{i,t}$ and reconstructs a pseudo-sample of transformed innovations as $\hat{\tilde{\eta}}_{i,t}^{(s)} = \hat{\rho}_{i,t}\hat{\eta}_{m,s} + \sqrt{1-\hat{\rho}_{i,t}^2}\,\hat{\eta}_{i,s}$. These $\hat{\tilde{\eta}}_{i,t}^{(s)}$ are no longer the observed residuals but resampled conditional counterparts consistent with the dependence structure at time t.

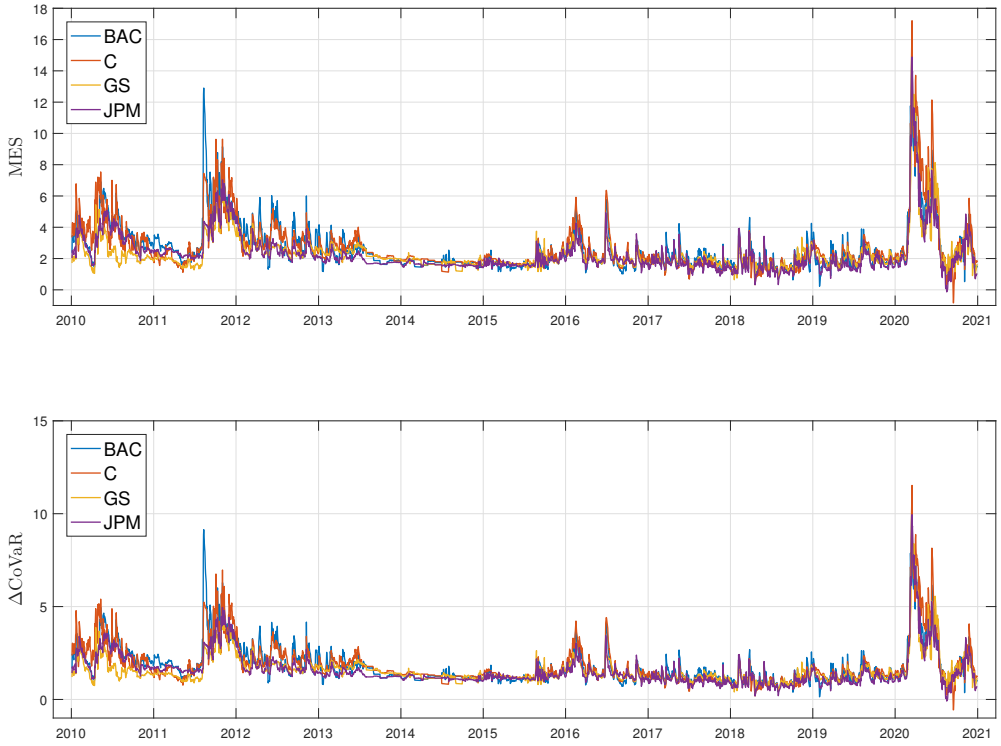
E - List of Tickers

Tickers and company names.

AFL	Aflac	LNC	Lincoln National
AIG	American International Group	MBI	MBIA
ALL	Allstate Corp	MMC	Marsh & McLennan
AON	Aon Corp	MS	Morgan Stanley
AXP	American Express	MTB	M & T Bank Corp
BAC	Bank of America	NTRS	Northern Trust
BEN	Franklin Resources	NYCB	New York Community Bancorp
BK	Bank of New York Mellon	PBCT	Peoples United Financial
BLK	Blackrock	PGR	Progressive
BRK	Berkshire Hathaway	PNC	PNC Financial Services
C	Citigroup	RF	Regions Financial
CI	CIGNA Corp	SCHW	Schwab Charles
CINF	Cincinnati Financial Corp	SEIC	SEI Investments Company
CMA	Comerica	SLM	SLM Corp
CNA	CNA Financial Corp	SNV	Synovus Financial
COF	Capital One Financial	STT	State Street
FITB	Fifth Third Bancorp	TFC	Truist Financial Corp
GL	Globe Life	TROW	T. Rowe Price
GS	Goldman Sachs	TRV	Travelers
HBAN	Huntington Bancshares	UNH	Unitedhealth Group
HIG	Hartford Financial Group	UNM	Unum Group
HUM	Humana	USB	US Bancorp
JPM	JP Morgan Chase	WFC	Wells Fargo & Co
KEY	Keycorp	WRB	W.R. Berkley Corp
L	Loews	ZION	Zion

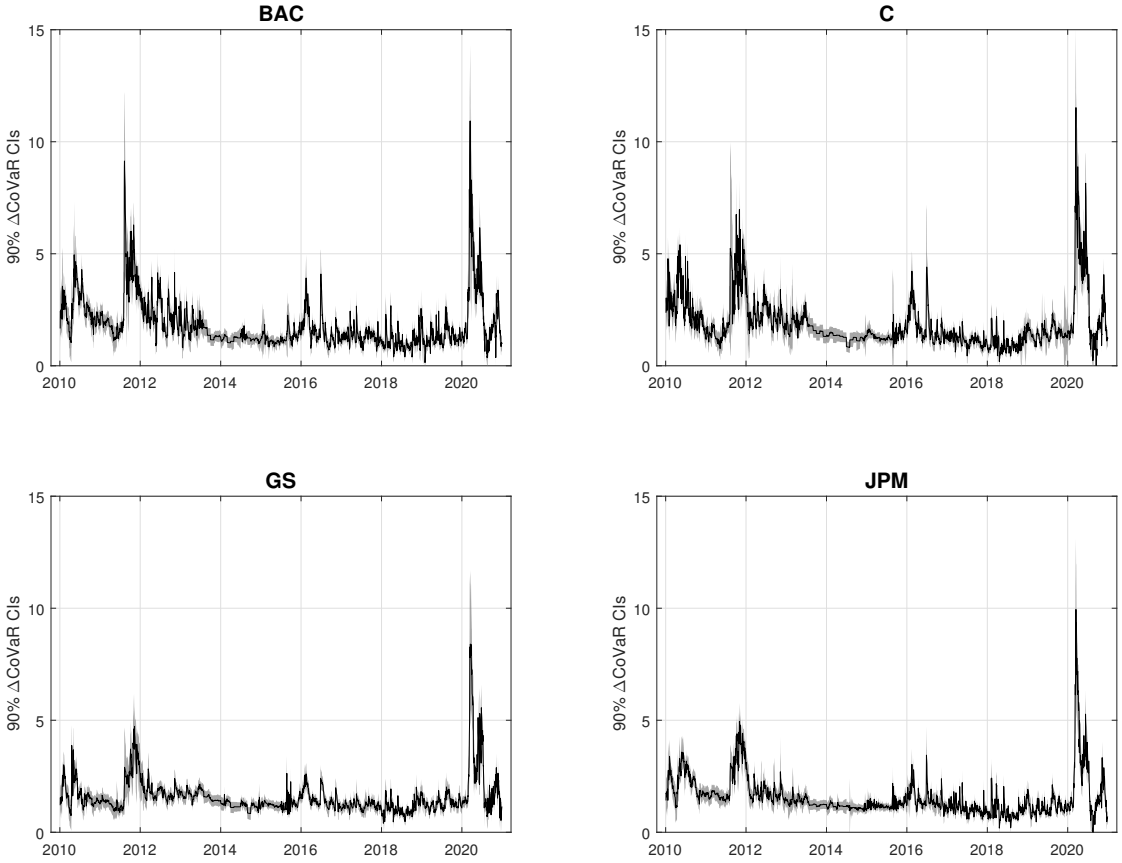
F - Additional Empirical Results

Figure 1.9: One-day ahead MES (top) and ΔCoVaR (bottom) forecasts.



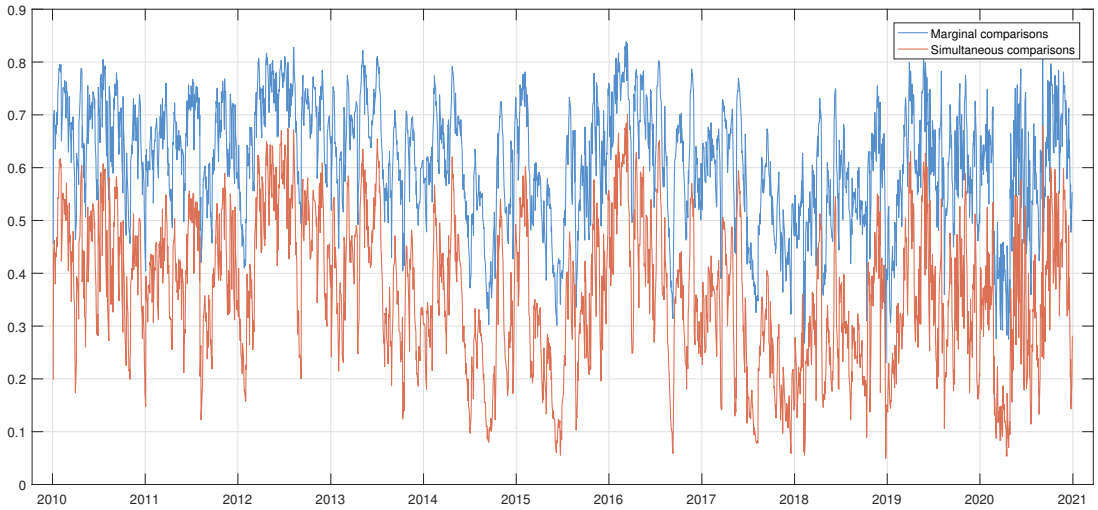
Daily forecasts of MES (top panel) and ΔCoVaR (bottom panel) at the α =10% probability level for Bank of America (BAC), Citigroup (C), Goldman Sachs (GS), and JP Morgan Chase (JPM). Parameters are estimated using a rolling window of size $n = 500$ and are updated on a monthly basis.

Figure 1.10: ΔCoVaR forecasts and 90% confidence intervals.



Daily ΔCoVaR forecasts (black line) and the corresponding 90% CIs (grey shaded area) for BAC, C, GS, and JPM. Parameters are estimated using a rolling window of size 500 and are updated on a monthly basis. The estimated CIs are obtained from (1.19) and feasible quantities are shown in Appendix C.6.

Figure 1.11: Rejection rates under marginal and joint inference ($n = 2000$).



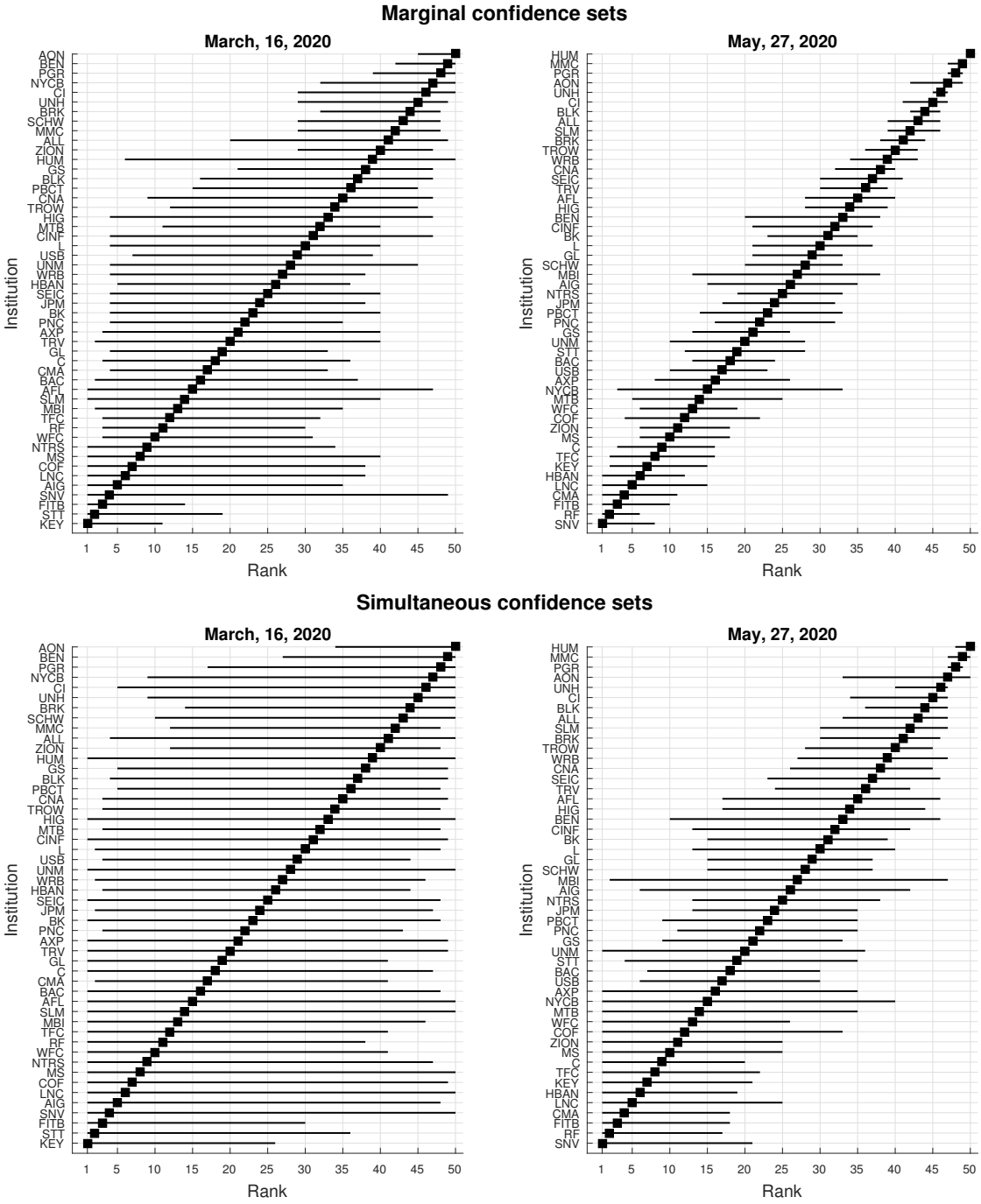
Rejection frequency of the null hypothesis $H_0 : \lambda_{j,t} = \lambda_{k,t}$ across all bank-market pairs based on marginal (blue) and joint (red) two-sided tests on the 95% confidence level. BEKK parameters are estimated with a rolling window of size $n = 2000$, updated monthly.

Table 1.4: Systemic risk rankings of financial institutions.

	Period Jan. 2010 to Dec. 2020						
Tickers	$\overline{\Delta C}_i$ (r)	$\overline{M}_i$ (r)	$\overline{\rho}_i$; $\overline{\sigma}_i$	$\overline{\Delta C}_i^{\leq}$(FZ) (r)	$\overline{\Delta C}_i$(CCC) (r)	$\tilde{r}^{\Delta C_i}$	$\tilde{r}^{M_i}$
LNC	2.224 (1)	3.262 (1)	0.72 ; 2.51	4.251 (1)	2.178 (1)	1	1
MBI	1.957 (2)	2.832 (2)	0.43 ; 3.46	3.724 (2)	2.095 (2)	3	3
MS	1.924 (3)	2.820 (3)	0.71 ; 2.16	3.426 (6)	1.907 (3)	2	2
RF	1.872 (4)	2.740 (4)	0.62 ; 2.44	3.681 (3)	1.863 (4)	6	6
C	1.842 (5)	2.704 (5)	0.69 ; 2.19	3.514 (5)	1.850 (5)	5	5
BAC	1.810 (6)	2.662 (6)	0.67 ; 2.16	3.562 (4)	1.816 (6)	8	8
SNV	1.795 (7)	2.635 (7)	0.60 ; 2.67	3.115 (15)	1.792 (7)	10	9
ZION	1.704 (8)	2.491 (8)	0.62 ; 2.21	3.176 (11)	1.732 (9)	9	9
KEY	1.693 (9)	2.483 (9)	0.64 ; 2.23	3.104 (17)	1.723 (11)	14	14
HIG	1.687 (10)	2.471 (10)	0.65 ; 2.09	2.706 (28)	1.646 (18)	15	15
BEN	1.687 (11)	2.468 (11)	0.74 ; 1.80	3.287 (8)	1.742 (8)	6	6
SCHW	1.671 (12)	2.453 (12)	0.66 ; 1.96	3.236 (10)	1.728 (10)	4	4
UNM	1.663 (13)	2.446 (13)	0.68 ; 2.07	3.332 (7)	1.626 (19)	15	15
FITB	1.661 (14)	2.430 (14)	0.63 ; 2.19	2.964 (20)	1.656 (16)	21	21
HBAN	1.656 (15)	2.428 (15)	0.63 ; 2.19	3.121 (14)	1.662 (13)	15	15
CMA	1.619 (16)	2.375 (16)	0.63 ; 2.11	3.159 (12)	1.668 (12)	15	19
BLK	1.617 (17)	2.373 (17)	0.75 ; 1.72	3.135 (13)	1.650 (17)	10	9
COF	1.608 (18)	2.369 (18)	0.64 ; 2.04	2.940 (22)	1.660 (14)	20	19
TROW	1.602 (19)	2.350 (19)	0.75 ; 1.67	2.985 (19)	1.657 (15)	12	12
AIG	1.591 (20)	2.332 (20)	0.58 ; 2.39	2.869 (23)	1.549 (23)	24	24
STT	1.570 (21)	2.314 (21)	0.66 ; 1.92	3.266 (9)	1.597 (20)	15	15
SLM	1.566 (22)	2.291 (23)	0.53 ; 2.37	3.106 (16)	1.523 (24)	24	24
GS	1.563 (23)	2.304 (22)	0.69 ; 1.83	3.005 (18)	1.588 (21)	12	12
JPM	1.551 (24)	2.278 (24)	0.71 ; 1.74	2.746 (25)	1.560 (22)	22	21
PNC	1.465 (25)	2.148 (25)	0.67 ; 1.79	2.692 (30)	1.459 (28)	27	27
WFC	1.465 (26)	2.145 (26)	0.68 ; 1.77	2.693 (29)	1.482 (26)	29	29
SEIC	1.431 (27)	2.098 (27)	0.69 ; 1.65	2.942 (21)	1.489 (25)	23	23
AXP	1.426 (28)	2.098 (28)	0.67 ; 1.78	2.797 (24)	1.480 (27)	29	29
BK	1.404 (29)	2.056 (29)	0.67 ; 1.63	2.557 (33)	1.419 (30)	24	24
TFC	1.400 (30)	2.052 (30)	0.65 ; 1.78	2.740 (26)	1.408 (32)	31	31
NTRS	1.384 (31)	2.035 (31)	0.68 ; 1.63	2.731 (27)	1.454 (29)	27	27
AFL	1.362 (32)	1.982 (32)	0.66 ; 1.69	2.561 (32)	1.409 (31)	34	33
MTB	1.299 (33)	1.902 (34)	0.61 ; 1.76	2.446 (36)	1.324 (33)	32	33
USB	1.296 (34)	1.902 (33)	0.68 ; 1.58	2.467 (35)	1.298 (34)	32	32
GL	1.259 (35)	1.855 (35)	0.71 ; 1.47	2.670 (31)	1.240 (35)	38	37
L	1.222 (36)	1.795 (36)	0.73 ; 1.44	2.317 (37)	1.228 (36)	35	37
CNA	1.197 (37)	1.757 (37)	0.61 ; 1.56	2.475 (34)	1.218 (37)	35	35
CI	1.168 (38)	1.716 (38)	0.50 ; 1.83	2.030 (40)	1.167 (38)	35	35
PBCT	1.103 (39)	1.630 (39)	0.59 ; 1.61	2.012 (41)	1.124 (40)	44	44
ALL	1.102 (40)	1.605 (41)	0.61 ; 1.39	2.003 (42)	1.090 (41)	39	39
CINF	1.092 (41)	1.606 (40)	0.64 ; 1.49	1.950 (46)	1.127 (39)	42	39
BRK	1.074 (42)	1.590 (42)	0.74 ; 1.20	2.079 (39)	1.043 (44)	39	39
UNH	1.050 (43)	1.539 (43)	0.51 ; 1.61	1.951 (44)	1.063 (42)	44	44
MMC	1.043 (44)	1.528 (44)	0.69 ; 1.21	1.957 (43)	1.026 (46)	42	43
AON	1.023 (45)	1.508 (45)	0.61 ; 1.36	2.287 (38)	1.025 (47)	39	39
NYCB	1.007 (46)	1.484 (46)	0.50 ; 1.66	1.950 (45)	1.044 (43)	46	46
PGR	0.994 (47)	1.458 (47)	0.57 ; 1.34	1.731 (49)	1.031 (45)	46	46
TRV	0.974 (48)	1.428 (48)	0.59 ; 1.36	1.918 (47)	0.956 (49)	49	49
HUM	0.96 (49)	1.408 (49)	0.39 ; 1.89	1.838 (48)	1.002 (48)	48	48
WRB	0.862 (50)	1.272 (50)	0.56 ; 1.23	1.515 (50)	0.866 (50)	50	50

Institutions are ordered according to $\overline{\Delta C}_i$. The metrics $\overline{M}_i$, $\overline{\Delta C}_i$, $\overline{\Delta C}_i^{\leq}$(FZ), and $\overline{\Delta C}_i$(CCC), denote institution-specific time averages for MES, ΔCoVaR, ΔCoVaR$^{\leq}$(FZ), and ΔCoVaR(CCC), respectively. Numbers in parentheses indicate cross-sectional ranks for the corresponding metrics. The metrics $\overline{\rho}_i$ and $\overline{\sigma}_i$ denote institution-specific time averages for market correlation and volatility, respectively. The quantities $\tilde{r}^{\Delta C_i}$ and $\tilde{r}^{\Delta M_i}$ denote the time median of daily cross-sectional ranks for ΔCoVaR and MES, respectively. Metrics and rankings are reported for January 2010-December 2020.

Figure 1.12: 95% confidence sets based on marginal and joint inference ($n = 2000$).



Each square indicates the point estimate of the rank of a given institution on the corresponding date. Horizontal lines represent 95% confidence sets for the ranks for March 16, 2020, and May 27, 2020. The upper panels show marginal confidence sets and the lower panels show joint confidence sets. BEKK parameters are estimated with a rolling window of size $n = 2000$, updated monthly.

G - Pairwise Rank Inference

To provide the basis for Corollaries 1.2-1.4 , we write the bivariate return as $\boldsymbol{y}_t = (y_{m,t}, y_{i,t})'$ with innovation $\boldsymbol{\eta}_t = (\eta_{m,t}, \eta_{i,t})'$, without an institution index as the analysis focuses on a single firm. To extend to a set of N institutions, we introduce the index $i \in I \equiv \{1, \ldots, N\}$, so that $\boldsymbol{y}_{i,t} = (y_{m,t}, y_{i,t})'$ and $\boldsymbol{\eta}_{i,t} = (\eta_{m,t}, \eta_{i,t})'$. The DGP is $\boldsymbol{y}_{i,t} = \boldsymbol{H}_{i,t}^{1/2}(\boldsymbol{\theta}_{0,i})\,\boldsymbol{\eta}_{i,t}$, where $\{\boldsymbol{\eta}_{i,t}\}$ are i.i.d. $\mathbb{R}^2$-valued random vectors such that $\mathbb{E}[\boldsymbol{\eta}_{i,t}] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{\eta}_{i,t}\boldsymbol{\eta}_{i,t}'] = \boldsymbol{I}_2$. The parameter vector $\boldsymbol{\theta}_{0,i}$, $i \in I$, belongs to the compact parameter space $\Theta \subseteq \mathbb{R}^p$. The QMLE of $\boldsymbol{\theta}_{0,i}$ is

$$\hat{\boldsymbol{\theta}}_{n,i} = \arg\min_{\boldsymbol{\theta}_i \in \Theta} L_{n,i}(\boldsymbol{\theta}_i), \qquad L_{n,i}(\boldsymbol{\theta}_i) = \sum_{t=1}^{n} \Big(\boldsymbol{y}_{i,t}' \boldsymbol{H}_{i,t}^{-1}(\boldsymbol{\theta}_i) \boldsymbol{y}_{i,t} + \log |\boldsymbol{H}_{i,t}(\boldsymbol{\theta}_i)| \Big). \tag{1.48}$$

Let $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{0,1}', \ldots, \boldsymbol{\theta}_{0,N}')'$, and $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\theta}}_{n,1}', \ldots, \hat{\boldsymbol{\theta}}_{n,N}')' \in \boldsymbol{\Theta}^N \subset \mathbb{R}^{Np}$ denote the stacked vectors of true and estimated parameters associated with the N institutions. Eq. (1.48) defines the QMLE for $i \in I$. Under the granular conditions 1.1–1.6 we have the Bahadur expansion for each institution i according to (1.18) as

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{n,i} - \boldsymbol{\theta}_{0,i}\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Delta}_{i,t-1}\, \boldsymbol{V}(\boldsymbol{\eta}_{i,t}) + o_p(1),$$

where, for $i \in I$, $\boldsymbol{V}(\boldsymbol{\eta}_{i,t}) = \mathrm{vec}\{\boldsymbol{I}_2 - \boldsymbol{\eta}_{i,t}\boldsymbol{\eta}_{i,t}'\}$, $\boldsymbol{J}_i^* = \mathbb{E}\Big[\frac{\partial^2 L_{n,i}(\boldsymbol{\theta}_{0,i})}{\partial \boldsymbol{\theta}_i\, \partial \boldsymbol{\theta}_i'}\Big]$ and $\boldsymbol{\Delta}_{i,t-1} = -\boldsymbol{J}_{\boldsymbol{i}}^{*-1} \frac{\partial\, \mathrm{vec}'\, \boldsymbol{H}_{i,t}(\boldsymbol{\theta}_{0,i})}{\partial \boldsymbol{\theta}_i} \left\{ \boldsymbol{H}_{i,t}^{-1/2}(\boldsymbol{\theta}_{0,i}) \otimes \boldsymbol{H}_{i,t}^{-1/2}(\boldsymbol{\theta}_{0,i}) \right\}'$. Consequently, the joint asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$ follows by assuming a joint linear expansion as (1.17) for all N institutions jointly (implied by the analogues of conditions 1.1–1.6). It reads as $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_\theta)$, where $\boldsymbol{\Psi}_\theta$ is an $Np \times Np$ block covariance matrix,

$$\boldsymbol{\Psi}_\theta = \begin{bmatrix} \boldsymbol{\Psi}_{1,1} & \boldsymbol{\Psi}_{1,2} & \cdots & \boldsymbol{\Psi}_{1,N} \\ \boldsymbol{\Psi}_{2,1} & \boldsymbol{\Psi}_{2,2} & \cdots & \boldsymbol{\Psi}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{N,1} & \boldsymbol{\Psi}_{N,2} & \cdots & \boldsymbol{\Psi}_{N,N} \end{bmatrix}.$$

The diagonal blocks are given by $\boldsymbol{\Psi}_{i,i} = E\big(\boldsymbol{\Delta}_{i,t}\, \mathrm{Var}(\boldsymbol{V}(\boldsymbol{\eta}_{i,t}))\, \boldsymbol{\Delta}_{i,t}'\big)$, for $i \in I$. For the off-diagonal blocks, we have $\boldsymbol{\Psi}_{j,k} = \mathbb{E}\big[\boldsymbol{\Delta}_{j,t}\, \mathrm{Cov}\big(\boldsymbol{V}(\boldsymbol{\eta}_{j,t}), \boldsymbol{V}(\boldsymbol{\eta}_{k,t})\big)\, \boldsymbol{\Delta}_{k,t}'\big]$ for $(j,k) \in I$ with $j \neq k$. Each block is consistently estimated using empirical means, variances, and covariances. Collecting these estimates yields the joint covariance estimator $\hat{\boldsymbol{\Psi}}_\theta$.

Define the mapping $\boldsymbol{\lambda}_t(\boldsymbol{\theta}) = \big(\lambda_{1,t}(\boldsymbol{\theta}_1), \ldots, \lambda_{N,t}(\boldsymbol{\theta}_N)\big)'$, where $\lambda_{i,t}(\boldsymbol{\theta}_i) = \sigma_{i,t}(\boldsymbol{\theta}_i)\rho_{i,t}(\boldsymbol{\theta}_i)$ is a smooth function of $\boldsymbol{\theta}_i$. Define $\boldsymbol{G}_{i,t} := \nabla_{\boldsymbol{\theta}_i} \lambda_{i,t}(\boldsymbol{\theta}_{0,i}) = \rho_{i,t}(\boldsymbol{\theta}_{0,i}) \dfrac{\partial \sigma_{i,t}(\boldsymbol{\theta}_{0,i})}{\partial \boldsymbol{\theta}_i} + \sigma_{i,t}(\boldsymbol{\theta}_{0,i}) \dfrac{\partial \rho_{i,t}(\boldsymbol{\theta}_{0,i})}{\partial \boldsymbol{\theta}_i} \in$

$\mathbb{R}^p$. Since each $\lambda_{i,t}$ depends only on $\boldsymbol{\theta}_i$, the Jacobian of $\boldsymbol{\lambda}_t$ w.r.t. $\boldsymbol{\theta}$ is block-diagonal with

$$\boldsymbol{D}_t = \text{blkdiag}\big(\boldsymbol{G}'_{1,t}, \ldots, \boldsymbol{G}'_{N,t}\big) = \begin{bmatrix} \boldsymbol{G}'_{1,t} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G}'_{2,t} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{G}'_{N,t} \end{bmatrix},$$

where each block $\boldsymbol{G}'_{i,t}$ is of dimension $1 \times p$. Hence, $\boldsymbol{D}_t$ is an $N \times (Np)$ matrix. From the delta method, it follows,

$$\sqrt{n}\Big(\boldsymbol{\lambda}_t(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)\Big) \;\to\; \mathcal{N}\Big(\mathbf{0},\; \boldsymbol{D}_t\, \boldsymbol{\Psi}_\theta\, \boldsymbol{D}'_t\Big), \tag{1.49}$$

where

$$\boldsymbol{D}_t\, \boldsymbol{\Psi}_\theta\, \boldsymbol{D}'_t = \begin{bmatrix} \boldsymbol{G}'_{1,t}\, \boldsymbol{\Psi}_{11}\, \boldsymbol{G}_{1,t} & \boldsymbol{G}'_{1,t}\, \boldsymbol{\Psi}_{12}\, \boldsymbol{G}_{2,t} & \cdots & \boldsymbol{G}'_{1,t}\, \boldsymbol{\Psi}_{1N}\, \boldsymbol{G}_{N,t} \\ \boldsymbol{G}'_{2,t}\, \boldsymbol{\Psi}_{21}\, \boldsymbol{G}_{1,t} & \boldsymbol{G}'_{2,t}\, \boldsymbol{\Psi}_{22}\, \boldsymbol{G}_{2,t} & \cdots & \boldsymbol{G}'_{2,t}\, \boldsymbol{\Psi}_{2N}\, \boldsymbol{G}_{N,t} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{G}'_{N,t}\, \boldsymbol{\Psi}_{N1}\, \boldsymbol{G}_{1,t} & \boldsymbol{G}'_{N,t}\, \boldsymbol{\Psi}_{N2}\, \boldsymbol{G}_{2,t} & \cdots & \boldsymbol{G}'_{N,t}\, \boldsymbol{\Psi}_{NN}\, \boldsymbol{G}_{N,t} \end{bmatrix}.$$

Then, Corollaries 1.2 - 1.4 build on the asymptotic distribution of $\boldsymbol{\lambda}_t(\hat{\boldsymbol{\theta}}_n)$ and the corresponding results follow directly.

Computation of $c_{1-\tau,t}$. Consider the global null hypothesis, i.e., $H_0 : \lambda_{1,t}(\boldsymbol{\theta}_{0,1}) = \lambda_{2,t}(\boldsymbol{\theta}_{0,2}) = \ldots = \lambda_{N,t}(\boldsymbol{\theta}_{0,N}) \equiv \omega_t$, where ω_t is a constant. Under H_0, we can write $\hat{\boldsymbol{\lambda}}_t = \omega_t \mathbf{1}_N + \hat{\boldsymbol{\Gamma}}_t$ with $\hat{\boldsymbol{\Gamma}}_t \xrightarrow{p} \mathbf{0}$ and $\sqrt{n}\,\hat{\boldsymbol{\Gamma}}_t \xrightarrow{d} \mathcal{N}\big(\mathbf{0},\; \boldsymbol{D}_t\, \boldsymbol{\Psi}_\theta\, \boldsymbol{D}'_t\big)$. Since $\hat{\boldsymbol{A}}_t \mathbf{1}_N = \mathbf{0}$, the common level ω_t vanishes in the contrasts, and we obtain $\hat{\boldsymbol{Z}}_t = \hat{\boldsymbol{A}}_t\, \hat{\boldsymbol{\Gamma}}_t$. The critical value of the max statistic is recovered by simulation. First, compute the plug-in covariance matrix $\hat{\boldsymbol{D}}_t\, \hat{\boldsymbol{\Psi}}_\theta\, \hat{\boldsymbol{D}}'_t$. Second, generate S independent draws $\big\{\hat{\boldsymbol{\Gamma}}_t^{(s)}\big\}_{s=1}^S \sim \mathcal{N}\big(\mathbf{0},\; n^{-1} \hat{\boldsymbol{D}}_t\, \hat{\boldsymbol{\Psi}}_\theta\, \hat{\boldsymbol{D}}'_t\big)$. Third, construct $\big\{\hat{\boldsymbol{Z}}_t^{(s)}\big\}_{s=1}^S$ under the null as $\hat{\boldsymbol{A}}_t\, \hat{\boldsymbol{\Gamma}}_t^{(s)}$ for $s = 1, \ldots, S$. Fourth, for each replication s, compute the max statistic $\hat{M}_t^{(s)} = \max_{(j,k)\in\mathcal{P}} |\hat{\boldsymbol{Z}}_t^{(s)}|$. The empirical distribution of $\{\hat{M}_t^{(s)}\}_{s=1}^S$ provides an approximation of the law of $\hat{M}_t$ under H_0, and the critical value $c_{1-\tau,t}$ is set as its empirical $(1-\tau)$-quantile. This computation relies on the max-t method of Westfall and Young (1993), which controls the FWER under i.i.d. data via permutation resampling. We extend the method to the time-series setting simulating the statistic from the conditional joint asymptotic distribution, which yields asymptotically valid critical values.

H - Further Simulation Evidence

H.1. Extended Results for Inference on MES and ΔCoVaR

We consider a specification in which MES and ΔCoVaR are time-invariant in both variances and correlations, in line with Acharya et al. (2017) for MES. We consider daily homoscedastic de-meaned returns $\boldsymbol{y}_t = (y_{m,t}, y_{i,t})'$ with $\boldsymbol{y}_t = \boldsymbol{H}^{1/2}\boldsymbol{\eta}_t$, where $y_{m,t}$ and $y_{i,t}$ denote the market and company returns observed at time t and $\boldsymbol{\eta}_t$ is an i.i.d. vector error process with $\mathbb{E}(\boldsymbol{\eta}_t) = 0$ and $\mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t') = \boldsymbol{I}_2$. For the distribution of $\boldsymbol{\eta}_t$ we use three alternative choices, the standard normal distribution (labeled by design N) and the normalized Student t distribution with 5 and 10 degrees of freedom, respectively (labeled by design $S5$ and $S10$, respectively). The (time-invariant) variance-covariance matrix $\boldsymbol{H}$ of $\boldsymbol{y}_t$ is given by

$$\boldsymbol{H} = \begin{pmatrix} \sigma_m^2 & \rho_i\sigma_m\sigma_i \\ \rho_i\sigma_m\sigma_i & \sigma_i^2 \end{pmatrix},$$

where σ_m and σ_i denote the volatilities of market and firm returns, respectively, and ρ_i is the correlation between both return processes. We calibrate the unknown parameters $\boldsymbol{\theta}_0 = (\sigma_m^2, \sigma_i^2, \rho_i)'$ using the daily log-returns of Bank of America (as the firm return) and the CRSP market value-weighted index (as the market return) over the period from January 2, 2015, to December 31, 2020. The resulting parameters are $\boldsymbol{\theta}_0 = (1.3340, 4.4460, 0.7708)'$.

For each Monte Carlo replication, we simulate the time series $\{y_{m,t}, y_{i,t}\}_{t=1}^{n+1}$ from the DGP above. Using the first n simulated observations $\{y_{m,t}, y_{i,t}\}_{t=1}^{n}$, we estimate $\hat{\boldsymbol{\theta}}_n$ by minimizing (1.12). We then compute $\hat{\mu}_{m,\alpha}$ and $\Delta\hat{\xi}_{m,\alpha}$ as in (1.13) and (1.14) using the estimated standardized residuals $\{\hat{\eta}_{m,t}\}_{t=1}^{n}$. Finally, we compute MES and ΔCoVaR for period $t = n+1$ as of (1.15) and (1.16), i.e., $MES_{i,n+1}(\alpha, \hat{\boldsymbol{\theta}}_n, \hat{\mu}_{m,\alpha})$ and $\Delta CoVaR_{i,n+1}(\alpha, \hat{\boldsymbol{\theta}}_n, \Delta\hat{\xi}_{m,\alpha})$. Note that in this full unconditional design, MES and ΔCoVaR are time-invariant and simplify to $MES_{i,n+1} = MES_i$ and $\Delta CoVaR_{i,n+1} = \Delta CoVaR_i$. We consider sample sizes ($n = 500$ and 1000) and coverage levels ($\alpha = 0.01$ and 0.10) to illustrate the impact of n and α on the small-sample properties of the estimators. We utilize 10,000 Monte Carlo replications. Tables 1.5 present the estimation results for the marginal model. In each panel, the top row displays the true parameter values, which are MES and ΔCoVaR. The second row shows the median estimated parameters across simulations, followed by the average bias and the corresponding standard deviations across simulations. As expected, estimates are centered around the true values. Dispersion decreases with n and increases as α decreases. We assess the finite-sample performance of standard error estimates for one-day-ahead ΔCoVaR and MES by reporting 90% coverage probabilities based on (1.19) and (1.20), as in Section 1.4.1. The time-invariant design is a special case of the dynamic setting with $\boldsymbol{H}_t \equiv \boldsymbol{H}$ for all t, and the derivatives $\frac{\partial \Delta\text{CoVaR}_{i,t}(\alpha;\boldsymbol{\theta},\Delta\xi)}{\partial(\boldsymbol{\theta}',\Delta\xi)}$ and $\frac{\partial \text{MES}_{i,t}(\alpha;\boldsymbol{\theta},\mu)}{\partial(\boldsymbol{\theta}',\mu)}$ in Corollary 1.1 are constant. Consequently – and in contrast to the time-varying design – the standard errors $\gamma_{\alpha,t}^{\xi}$ and $\gamma_{\alpha,t}^{\mu}$ are constant across periods, including $t = n+1$. The simulation results mirror those in Section 1.4.1. Coverage rates approach the nominal level as n and α increase and exhibit similar small-sample distortions, with no material differences across designs.

Table 1.5: Simulation results for the marginal design.

		$n = 500$, $\alpha = 0.01$			$n = 1000$, $\alpha = 0.01$		
		N	$S10$	$S5$	N	$S10$	$S5$
$\Delta CoVaR_{i,n+1}$	True	-3.7809	-4.0176	-4.2361	-3.7809	-4.0176	-4.2361
	Median	-3.7646	-3.9946	-4.2082	-3.7691	-4.0009	-4.2157
	Avg Bias	-0.0069	0.0033	0.0234	-0.0051	-0.0024	0.0075
	Std Dev	0.3007	0.4022	0.5377	0.2170	0.2813	0.3764
$MES_{i,n+1}$	True	-4.3316	-4.8890	-5.6052	-4.3316	-4.8890	-5.6052
	Median	-4.2559	-4.7643	-5.3535	-4.2946	-4.8154	-5.4735
	Avg Bias	-0.0597	-0.0770	-0.1028	-0.0316	-0.0472	-0.0482
	Std Dev	0.3626	0.5813	0.9725	0.2572	0.4165	0.7119
Coverage probabilities	$\sigma_{m,n+1}$	0.8908	0.8848	0.8662	0.8989	0.8938	0.8724
	$\sigma_{i,n+1}$	0.8937	0.8938	0.8768	0.9012	0.8963	0.8853
	$\rho_{i,n+1}$	0.8909	0.8879	0.8776	0.8968	0.8949	0.8907
	$\Delta\xi_{m,\alpha}$	0.8595	0.8405	0.8372	0.8592	0.8552	0.8488
	$\mu_{m,\alpha}$	0.8044	0.7750	0.7546	0.8486	0.8221	0.8168
	$\Delta CoVaR_{i,n+1}$	0.8538	0.8396	0.8272	0.8573	0.8563	0.8485
	$MES_{i,n+1}$	0.8190	0.7873	0.7533	0.8542	0.8288	0.8091
		$n = 500$, $\alpha = 0.10$			$n = 1000$, $\alpha = 0.10$		
		N	$S10$	$S5$	N	$S10$	$S5$
$\Delta CoVaR_{i,n+1}$	True	-2.0828	-1.9947	-1.8580	-2.0828	-1.9947	-1.8580
	Median	-2.0776	-1.9886	-1.8541	-2.0803	-1.9912	-1.8564
	Avg Bias	-0.0017	-0.0015	0.0006	-0.0019	-0.0011	0.0000
	Std Dev	0.1492	0.1476	0.1442	0.1057	0.1057	0.1030
$MES_{i,n+1}$	True	-2.8523	-2.8916	-2.8983	-2.8523	-2.8916	-2.8983
	Median	-2.8409	-2.8772	-2.8794	-2.8467	-2.8841	-2.8900
	Avg Bias	-0.0088	-0.0088	-0.0084	-0.0050	-0.0060	-0.0034
	Std Dev	0.1749	0.2022	0.2419	0.1242	0.1435	0.1735
Coverage probabilities	$\sigma_{m,n+1}$	0.8908	0.8848	0.8662	0.8989	0.8938	0.8724
	$\sigma_{i,n+1}$	0.8937	0.8938	0.8768	0.9012	0.8963	0.8853
	$\rho_{i,n+1}$	0.8909	0.8879	0.8776	0.8968	0.8949	0.8907
	$\Delta\xi_{m,\alpha}$	0.8962	0.8968	0.8859	0.8984	0.8967	0.8870
	$\mu_{m,\alpha}$	0.8976	0.8912	0.8827	0.9001	0.8991	0.8921
	$\Delta CoVaR_{i,n+1}$	0.8936	0.8914	0.8889	0.8900	0.8920	0.8901
	$MES_{i,n+1}$	0.8892	0.8916	0.8841	0.8963	0.8963	0.8912

This table presents results from 10,000 replications of the estimation of MES and $\Delta CoVaR$ from the time-invariant DGP with normal and Student innovations with 5 and 10 degrees of freedom as reported in the columns N, $S5$ and $S10$, respectively. The top rows of each panel present the true values of ΔCoVaR and MES. The second, third, and fourth rows present the median estimated parameters, the average bias, and the standard deviation (across simulations) of the estimated parameters. The last rows of each panel present the coverage rates for 90% CIs constructed using estimated standard errors.

Table 1.6: Simulation results under BEKK and CCC-type MGARCH specifications ($\alpha = 1\%$).

Panel A: Data generating process is a BEKK

Estimated Model	Dist	$\sigma_{i,n+1}$	$\sigma_{m,n+1}$	$\rho_{\mathbf{i,n+1}}$	$\Delta\xi_{m,\alpha}$	$\mu_{m,\alpha}$	$\mathbf{MES_{i,n+1}}$	$\mathbf{\Delta CoVaR_{i,n+1}}$
1) $n = 500$								
BEKK	*N*	0.464(-0.012)[0.90]	0.238(-0.011)[0.90]	**0.037(0.000)[0.90]**	0.148(-0.002)[0.85]	0.183(-0.039)[0.80]	**0.317(-0.006)[0.87]**	**0.367(-0.054)[0.85]**
CCC	*N*	0.688(0.075)[0.79]	0.222(0.013)[0.81]	**0.212(-0.006)[0.21]**	0.148(0.003)[0.86]	0.183(-0.032)[0.80]	**1.139(-0.047)[0.37]**	**1.312(-0.098)[0.38]**
BEKK	*S10*	0.724(-0.007)[0.89]	0.216(-0.012)[0.90]	**0.040(-0.001)[0.89]**	0.186(-0.010)[0.85]	0.298(-0.077)[0.76]	**0.396(-0.017)[0.88]**	**0.577(-0.094)[0.83]**
CCC	*S10*	0.825(0.082)[0.79]	0.269(0.028)[0.82]	**0.205(-0.012)[0.21]**	0.185(-0.007)[0.84]	0.297(-0.070)[0.76]	**1.177(-0.050)[0.41]**	**1.454(-0.134)[0.43]**
BEKK	*S5*	2.578(-0.151)[0.89]	0.938(-0.044)[0.87]	**0.050(0.002)[0.89]**	0.270(0.030)[0.83]	0.463(-0.112)[0.73]	**0.547(0.018)[0.88]**	**0.932(-0.133)[0.76]**
CCC	*S5*	2.106(0.016)[0.83]	0.781(0.025)[0.81]	**0.199(0.007)[0.27]**	0.268(0.032)[0.84]	0.456(-0.112)[0.73]	**1.425(-0.088)[0.52]**	**1.961(-0.270)[0.51]**
2) $n = 1000$								
BEKK	*N*	0.300(0.007)[0.91]	0.110(-0.006)[0.90]	**0.026(0.000)[0.90]**	0.108(-0.005)[0.85]	0.136(-0.021)[0.84]	**0.219(-0.009)[0.88]**	**0.260(-0.030)[0.88]**
CCC	*N*	0.675(0.102)[0.73]	0.161(0.014)[0.75]	**0.210(-0.005)[0.14]**	0.108(-0.002)[0.85]	0.134(-0.015)[0.84]	**1.120(-0.065)[0.29]**	**1.285(-0.091)[0.30]**
BEKK	*S10*	0.457(-0.002)[0.90]	0.139(-0.009)[0.91]	**0.028(-0.001)[0.90]**	0.136(-0.003)[0.85]	0.215(-0.037)[0.83]	**0.281(-0.004)[0.88]**	**0.396(-0.039)[0.85]**
CCC	*S10*	0.706(0.098)[0.78]	0.176(0.020)[0.80]	**0.205(-0.011)[0.15]**	0.138(-0.000)[0.85]	0.214(-0.032)[0.83]	**1.164(-0.047)[0.34]**	**1.425(-0.093)[0.38]**
BEKK	*S5*	1.584(-0.096)[0.90]	0.323(-0.010)[0.89]	**0.034(0.000)[0.90]**	0.204(0.014)[0.84]	0.342(-0.070)[0.79]	**0.431(0.023)[0.84]**	**0.752(-0.062)[0.81]**
CCC	*S5*	1.488(0.021)[0.82]	0.507(0.057)[0.80]	**0.200(0.005)[0.20]**	0.204(0.018)[0.84]	0.336(-0.069)[0.79]	**1.402(-0.076)[0.43]**	**1.876(-0.199)[0.45]**

Panel B: Data generating process is a CCC

Estimated Model	Dist	$\sigma_{i,n+1}$	$\sigma_{m,n+1}$	$\rho_{\mathbf{i,n+1}}$	$\Delta\xi_{m,\alpha}$	$\mu_{m,\alpha}$	$\mathbf{MES_{i,n+1}}$	$\mathbf{\Delta CoVaR_{i,n+1}}$
1) $n = 500$								
CCC	*N*	0.482(-0.009)[0.90]	0.675(-0.026)[0.91]	**0.025(0.000)[0.90]**	0.146(-0.002)[0.86]	0.181(-0.038)[0.80]	**0.282(-0.000)[0.88]**	**0.332(-0.044)[0.85]**
BEKK	*N*	0.763(0.053)[0.73]	1.055(-0.002)[0.82]	**0.119(0.042)[0.48]**	0.148(0.005)[0.87]	0.184(-0.031)[0.81]	**0.559(-0.138)[0.68]**	**0.654(-0.200)[0.65]**
CCC	*S10*	0.607(-0.009)[0.89]	0.436(0.003)[0.89]	**0.028(0.000)[0.90]**	0.183(-0.010)[0.85]	0.296(-0.075)[0.77]	**0.352(-0.014)[0.88]**	**0.503(-0.093)[0.81]**
BEKK	*S10*	1.015(0.084)[0.75]	0.832(-0.007)[0.84]	**0.129(0.051)[0.48]**	0.184(-0.010)[0.85]	0.295(-0.076)[0.77]	**0.699(-0.222)[0.64]**	**0.908(-0.338)[0.64]**
CCC	*S5*	2.960(-0.091)[0.89]	1.977(-0.029)[0.88]	**0.034(0.002)[0.89]**	0.273(0.036)[0.84]	0.466(-0.103)[0.72]	**0.463(0.020)[0.85]**	**0.835(-0.128)[0.77]**
BEKK	*S5*	3.535(-0.024)[0.77]	2.044(-0.023)[0.84]	**0.120(0.036)[0.59]**	0.266(0.024)[0.84]	0.459(-0.120)[0.72]	**0.778(-0.130)[0.73]**	**1.167(-0.321)[0.67]**
2) $n = 1000$								
CCC	*N*	0.381(0.005)[0.91]	0.397(-0.023)[0.91]	**0.018(-0.000)[0.91]**	0.107(-0.006)[0.85]	0.134(-0.023)[0.84]	**0.200(-0.007)[0.87]**	**0.233(-0.027)[0.88]**
BEKK	*N*	0.744(0.070)[0.65]	0.455(0.013)[0.78]	**0.120(0.044)[0.33]**	0.108(-0.001)[0.85]	0.135(-0.015)[0.83]	**0.504(-0.151)[0.56]**	**0.584(-0.189)[0.57]**
CCC	*S10*	0.448(-0.008)[0.89]	0.707(-0.032)[0.92]	**0.020(-0.000)[0.90]**	0.135(-0.002)[0.87]	0.215(-0.037)[0.82]	**0.259(0.000)[0.86]**	**0.355(-0.038)[0.84]**
BEKK	*S10*	0.857(0.084)[0.66]	1.274(-0.021)[0.84]	**0.129(0.052)[0.37]**	0.137(-0.004)[0.85]	0.214(-0.039)[0.83]	**0.622(-0.210)[0.59]**	**0.779(-0.290)[0.60]**
CCC	*S5*	2.462(-0.100)[0.90]	2.022(-0.014)[0.89]	**0.025(0.001)[0.90]**	0.204(0.018)[0.84]	0.341(-0.066)[0.79]	**0.364(0.022)[0.85]**	**0.624(-0.066)[0.80]**
BEKK	*S5*	2.524(-0.017)[0.71]	2.007(-0.003)[0.84]	**0.119(0.036)[0.47]**	0.200(0.009)[0.84]	0.338(-0.079)[0.79]	**0.712(-0.126)[0.66]**	**1.037(-0.259)[0.68]**

Simulation results for MES and ΔCoVaR estimators at $\alpha = 1\%$. Panel A (Panel B) reports results when the DGP follows the BEKK (CCC) specification. Subpanel 1) corresponds to $n = 500$, and subpanel 2) to $n = 1000$. Entries are reported as RMSE(bias)[coverage], where coverage denotes the empirical 90% coverage probability of the estimated asymptotic CI. Correct specification occurs when the first-step MGARCH specification coincides with the DGP, otherwise, the setting is misspecified. Each design is based on 1,000 Monte Carlo replications.

H.2. Simulation Evidence for Pairwise Rank Inference

We run simulations to study the finite-sample behavior of the testing procedure in Corollary 1.2 under equal SRMs for two institutions. The goal of this exercise is to show that the procedure performs well for marginal pairwise comparisons and also serves as a small-scale validation of simultaneous inference, since the marginal case already relies on the joint distribution of $\boldsymbol{\lambda}_t(\hat{\boldsymbol{\theta}}_n)$ as for the simultaneous case. We consider the BEKK setting as of Section 1.4.1. Specifically, we consider two synthetic institutions j and k, whose returns are as follows,

$$\boldsymbol{y}_{i,t} = \boldsymbol{H}_{i,t}^{1/2}(\boldsymbol{\theta}_{0,i})\,\boldsymbol{\eta}_{i,t}, \quad i \in \{j,k\}, \tag{1.50}$$

with conditional variance-covariance matrix

$$\boldsymbol{H}_{i,t} = \boldsymbol{C}_{0,i}\boldsymbol{C}_{0,i}' + \boldsymbol{A}_{0,i}\boldsymbol{y}_{i,t-1}\boldsymbol{y}_{i,t-1}'\boldsymbol{A}_{0,i}' + \boldsymbol{B}_{0,i}\boldsymbol{H}_{i,t-1}\boldsymbol{B}_{0,i}'. \tag{1.51}$$

We consider the same parameter vector for the two systems, i.e., $\boldsymbol{\theta}_{0,j} = \boldsymbol{\theta}_{0,k}$, implying $\boldsymbol{C}_{0,j} = \boldsymbol{C}_{0,k}$, $\boldsymbol{A}_{0,j} = \boldsymbol{A}_{0,k}$, and $\boldsymbol{B}_{0,j} = \boldsymbol{B}_{0,k}$. The parameter values are identical to those reported in Section 1.4.1. For each replication, we draw the innovations $\{\eta_{m,t}\}_{t=1}^{n+1}$, $\{\eta_{j,t}\}_{t=1}^{n+1}$, and $\{\eta_{k,t}\}_{t=1}^{n+1}$ considering three distributional choices, namely the standard normal distribution (N), the standardized Student-t distribution with 5 degrees of freedom ($S5$), and with 10 degrees of freedom ($S10$). Applying Eqs. (1.50)-(1.51), we generate the return series $\{\boldsymbol{y}_{j,t}\}_{t=1}^{n+1}$ and $\{\boldsymbol{y}_{k,t}\}_{t=1}^{n+1}$. We estimate the parameters $\boldsymbol{\theta}_{0,j}$ and $\boldsymbol{\theta}_{0,k}$ by QMLE, using Eq. (1.48) from the samples $\{\boldsymbol{y}_{j,t}\}_{t=1}^{n}$ and $\{\boldsymbol{y}_{k,t}\}_{t=1}^{n}$, respectively, yielding the estimates $\hat{\boldsymbol{\theta}}_{n,j}$ and $\hat{\boldsymbol{\theta}}_{n,k}$. We then compute the statistic $\hat{Z}_{j,k,t}$ under H_0 at date $t = n+1$, i.e., out-of-sample.

Although we generate the returns considering $\boldsymbol{\theta}_{0,j} = \boldsymbol{\theta}_{0,k}$, this condition alone does not guarantee that H_0 holds, since $\lambda_{j,n+1}(\boldsymbol{\theta}_{0,j})$ and $\lambda_{k,n+1}(\boldsymbol{\theta}_{0,k})$ depend on their respective pasts. To enforce H_0, we construct a synthetic scenario by setting the past of institution k (i.e., $\{\boldsymbol{y}_{k,t}\}_{t=1}^{n}$) equal to that of institution j. This ensures that the two risk measures are evaluated conditionally on exactly the same information set, making the null hypothesis to hold by construction. The controlled version of the null hypothesis is given by

$$H_0^* : \lambda_{j,n+1}(\boldsymbol{\theta}_{0,j}) = \lambda_{j,n+1}(\boldsymbol{\theta}_{0,k}),$$

where $\lambda_{j,n+1}(\boldsymbol{\theta}_{0,k})$ is evaluated using the historical information of institution j, rather than that of institution k. We then compute the statistic as

$$Z_{j,k,n+1}^* = \frac{\lambda_{j,n+1}(\hat{\boldsymbol{\theta}}_{n,j}) - \lambda_{j,n+1}(\hat{\boldsymbol{\theta}}_{n,k})}{n^{-1/2}\sqrt{\hat{\iota}_{j,k,n+1}^*}},$$

where $\hat{\iota}_{j,k,n+1}^* = \boldsymbol{G}_{j,n+1}(\hat{\boldsymbol{\theta}}_n)'\,\hat{\boldsymbol{\Psi}}_{jj}\,\boldsymbol{G}_{j,n+1}(\hat{\boldsymbol{\theta}}_n) + \boldsymbol{G}_{k,n+1}^*(\hat{\boldsymbol{\theta}}_n)'\,\hat{\boldsymbol{\Psi}}_{kk}\,\boldsymbol{G}_{k,n+1}^*(\hat{\boldsymbol{\theta}}_n) - 2\,\boldsymbol{G}_{j,n+1}(\hat{\boldsymbol{\theta}}_n)'\,\hat{\boldsymbol{\Psi}}_{jk}\,\boldsymbol{G}_{k,n+1}^*(\hat{\boldsymbol{\theta}}_n)$, with $\boldsymbol{G}_{k,n+1}^*(\hat{\boldsymbol{\theta}}_n) = \nabla_{\boldsymbol{\theta}_k}\lambda_{j,n+1}(\boldsymbol{\theta}_{0,k})$. Note that we use the notation * in the superscript of H_0, $Z_{j,k,n+1}$ and $\iota_{j,k,n+1}$ to reflect that the statistic is constructed under the controlled

null hypothesis. The procedure is repeated $S = 10000$ times, yielding S realizations of the test statistic $\{Z^{*(s)}_{j,k,n+1}\}_{s=1}^{S}$ under H_0^*.

We then compute the empirical size at the 5% nominal level for three types of tests: (i) two-sided, with alternative $\lambda_{j,n+1}(\boldsymbol{\theta}_{0,j}) \neq \lambda_{j,n+1}(\boldsymbol{\theta}_{0,k})$, (ii) left-tailed, with alternative $\lambda_{j,n+1}(\boldsymbol{\theta}_{0,j}) < \lambda_{j,n+1}(\boldsymbol{\theta}_{0,k})$, and (iii) right-tailed, with alternative $\lambda_{j,n+1}(\boldsymbol{\theta}_{0,j}) > \lambda_{j,n+1}(\boldsymbol{\theta}_{0,k})$. The corresponding empirical sizes are computed as $\frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\left(|Z^{*(s)}_{j,k,n+1}| > \Phi^{-1}(0.975)\right)$, $\frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\left(Z^{*(s)}_{j,k,n+1} < \Phi^{-1}(0.05)\right)$, and $\frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\left(Z^{*(s)}_{j,k,n+1} > \Phi^{-1}(0.95)\right)$, respectively, where $\Phi^{-1}(.)$ is the inverse cdf of the standard normal distribution.

Table 1.7 shows that the empirical rejection rates are close to the 5% nominal level across all settings, with only mild deviations for smaller sample sizes ($n = 500$), indicating that the test maintains good size properties in finite samples. As n increases, the rejection frequencies stabilize around the nominal size for both normal and Student innovations.

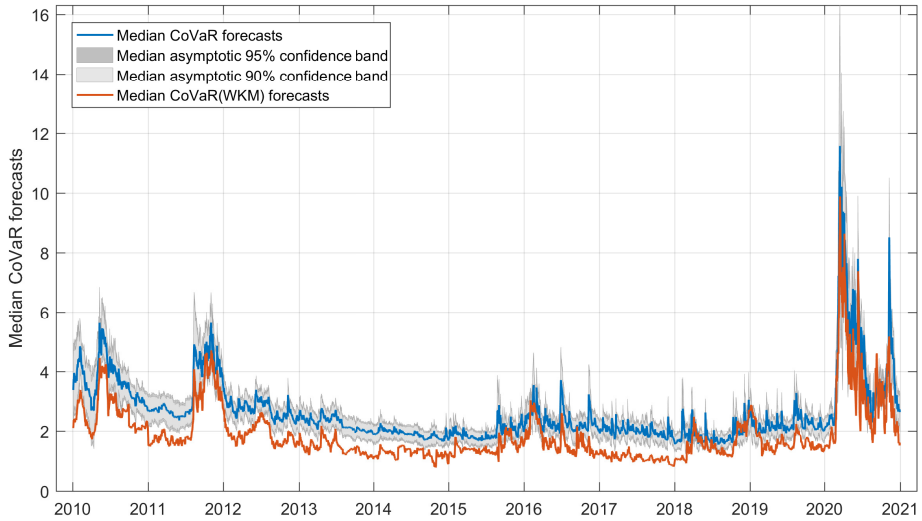
Table 1.7: Empirical sizes at 5% nominal level for two-sided, right-tailed, and left-tailed tests.

$n = 500$			
	N	$S10$	$S5$
Two-sided	0.0623	0.0589	0.0515
Left-tailed	0.0636	0.0543	0.0511
Right-tailed	0.0583	0.0541	0.0497
$n = 1000$			
	N	$S10$	$S5$
Two-sided	0.0568	0.0521	0.0476
Left-tailed	0.0513	0.0554	0.0484
Right-tailed	0.0548	0.0511	0.0406
$n = 5000$			
	N	$S10$	$S5$
Two-sided	0.0441	0.0507	0.0486
Left-tailed	0.0464	0.0449	0.0445
Right-tailed	0.0469	0.0472	0.0490

Empirical size of the two-sided, left-tailed, and right-tailed tests for the pairwise statistic $Z^*_{j,k,n+1}$ at the 5% nominal level, based on 10,000 Monte Carlo replications. Time series $\{\boldsymbol{y}_{j,t}\}_{t=1}^{n+1}$ and $\{\boldsymbol{y}_{k,t}\}_{t=1}^{n+1}$ are simulated from BEKK models with innovations drawn from the normal distribution (N) and standardized Student-t distributions with 5 ($S5$) and 10 ($S10$) degrees of freedom. For each replication, BEKK parameters are estimated using the first n observations with $n = 500$, 1000, and 5000.

I - CoVaR Estimation: A Comparative Analysis

We compare the CoVaR estimator of White et al. (2015), denoted by CoVaR(WKM), with our CoVaR estimator, whose semi-parametric estimation procedure is presented in Appendix C.3. For our CoVaR, we use $\alpha = 0.1$ and $\beta = 0.1$. The CoVaR of White et al. (2015) is based on multivariate quantile models. We follow their construction as defined in their footnote 7 and adopt the model specification in their Section 4. We set the quantile probability level for CoVaR(WKM) to $\theta = 0.1$ (their notation), which corresponds to $\alpha = \beta = 0.1$ in our setup.[19] The remainder of the empirical setup is identical to that in Section 5 of the main text. Our CoVaR estimator relies on a diagonal BEKK specification for $\boldsymbol{H}_t$. For both CoVaR estimators, model parameters are estimated using a rolling-window procedure with window size $n = 500$, and the estimators are computed out of sample using one-step-ahead forecasts over the subsequent month. The rolling window is updated on a monthly basis. CoVaRs are reported with opposite signs: higher values indicate larger systemic risk contributions.



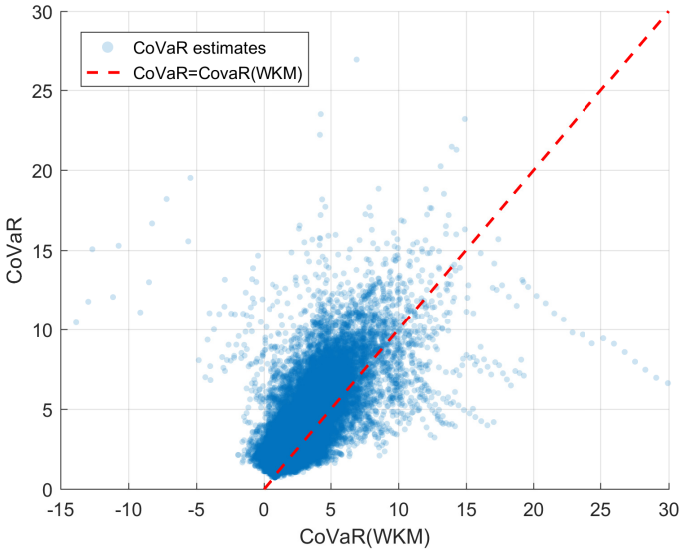
CoVaR (in blue) and CoVaR(WKM) (in red), computed at each period t as the median across the 50 banks. The grey shaded areas shows the cross-sectional median of the asymptotic confidence bands of our CoVaR estimator: the darker grey shows the 95% level, while the lighter grey shows the 90% level, using Theorem 4 (Appendix C.3). Model parameters for both CoVaR estimators are estimated using a rolling-window procedure with $n = 500$, updated on a monthly basis.

Figure 1.13 reports the median CoVaR forecasts across the 50 institutions over time. The grey shaded areas display the cross-sectional median of the asymptotic confidence bands for our CoVaR estimator (Theorem 4, Appendix D.3), with darker and lighter shades corresponding to the 95% and 90% confidence levels, respectively. The blue

[19]The estimation of CoVaR(WKM) is performed using Matlab code provided by the authors.

line depicts our corresponding CoVaR estimator, while the red line corresponds to the estimator of White et al. (2015). Overall, our CoVaR values tend to be larger than those obtained using the approach in White et al. (2015). This difference is unlikely to be driven by the empirical setup, as comparable probability levels are used in both cases. Instead, it may reflect differences in the definition of CoVaR. In White et al. (2015), CoVaR is defined as the β-quantile of returns conditional on the market return being equal to its α-VaR at time $t-1$, whereas in our case the conditioning occurs contemporaneously at time t. This difference in timing may naturally lead to a more severe reaction of our CoVaR estimates.

Figure 1.14: Scatter plot of CoVaR forecasts.



Scatter plot of CoVaR estimates against CoVaR(WKM). Each point corresponds to the CoVaR values for a given day and a given bank among the 90 institutions over the 2010-2020 period. Model parameters for both CoVaR estimators are estimated using a rolling-window procedure with $n = 500$, updated on a monthly basis.

Figure 1.14 compares the CoVaR point estimates obtained from the two approaches. The red line represents the 45-degree line, corresponding to equality between our CoVaR and CoVaR(WKM). First, while the two measures are correlated, the figure reveals substantial heterogeneity across the two methods. Second, consistent with Figure 1.13, our approach tends to produce more prudent CoVaR estimates than the WKM method. Finally, (the opposite of) our CoVaR is naturally bounded below by zero, a direct implication of its form in (1.30), since negative values would require a negative correlation between the market and the firm. By contrast, the CoVaR estimator of White et al. (2015) is not subject to such a bound.

2 Cardinality-Constrained Optimization for Large-Scale Portfolio

Abstract. We propose a portfolio optimization model that reconciles Keynes's advocacy for concentrated investments with Markowitz's emphasis on diversification. By incorporating cardinality constraints into the Markowitz mean-variance framework, we enable investors to focus on a small set of assets, fostering specialized expertise. Cardinality constraints allow investors to still use the sample covariance matrix in high-dimensional settings with limited data, balancing diversification needs while mitigating estimation errors inherent in such environments.

Keywords: Co-positive Optimization, Portfolio Optimization, Financial Econometrics.

2.1 Introduction

Since the seminal paper of Markowitz (1952a), the *mean-variance* framework has been the dominant model used in practice for asset allocation and active portfolio management. Markowitz highlighted diversification as a fundamental principle of portfolio construction, especially when investing in a broad universe of assets. Under this view, conventional wisdom suggests that an unconstrained rational investor should hold as many assets as possible to diversify idiosyncratic risk and achieve a more efficient risk–return trade-off. On the other hand, Keynes (Keynes, 1983) advocated concentrating wealth in a few stocks in which one has the highest conviction. Warren Buffett expressed the same view in his 1991 Chairman's Letter to Berkshire Hathaway shareholders. In practice, concentrated portfolios are commonly observed among individual investors (Campbell, 2006; Calvet et al., 2007; Goetzmann and Kumar, 2008).

This raises a natural question: Is holding more stocks always better due to improved diversification? A related question is: How many stocks are needed to form a well-diversified portfolio? Statman (1987) addressed this issue in a seminal paper. Besides the opposing views in the financial literature, an important practical reason for reducing portfolio size is estimation error. In theory, the optimal portfolio can only benefit from the increased investment opportunities associated with holding more assets. In practice, however, the number of parameters in the covariance matrix grows with n^2, which increases estimation risk and worsens out-of-sample performance. Thus, there is a clear trade-off to consider. In our study, we propose a framework that reconciles Markowitz's principle of broad diversification with Keynesian-style concentrated investing. Our approach allows investors to focus on a small subset of stocks while still capturing the benefits of diversification.

To improve portfolio performance, it is crucial to mitigate estimation errors, which are widely considered a primary source of poor out-of-sample results. Consequently, the focus is typically on improving covariance matrix estimations rather than addressing optimization problems with constraints for better portfolio management. Zhao et al. (2019) argue that even small, unbiased estimation errors can lead to significantly poor performance because the optimization step amplifies these errors in an asymmetric manner. Therefore, considering the optimization problem itself is also crucial for effective portfolio optimization.

This paper argues for combining advanced estimation methods with optimization approaches that include specific constraints, such as cardinality constraints, to enhance decision-making and investment strategies. Cardinality constraints, which limit the number of assets in a portfolio, can render simpler estimators, like the sample covariance matrix, sufficient for investment decisions—particularly in high-dimensional settings where significant estimation errors typically impact portfolio performance.

We impose cardinality constraints by using a zero-norm constraint to select a smaller portfolio size, which differs from the commonly used approach of imposing LASSO-type constraints (DeMiguel et al., 2009a; Fan et al., 2012; Ao et al., 2019). There are two main advantages to using the zero-norm constraint: First, by directly addressing the cardinality constraint, our algorithm provides meaningful explanations for investors. For example, if an investor needs to manage no more than M stocks, our approach identifies the specific set of stocks on which to focus. Investors are not required to fine-tune the LASSO penalty parameter to achieve their desired outcomes. Second, Green and Hollifield (1992) argue that the optimal portfolio can have sizeable asset weights. While norm constraints like those used in LASSO can help reduce portfolio size, they risk excluding the optimal solution by penalizing large portfolio weights. In contrast, our cardinality constraint does not alter the weights of the selected stocks; it focuses solely on dimensionality reduction. Norm-constrained approaches, on the other hand, often require ad hoc modifications to the objective function to maintain a low-dimensional portfolio.

Empirically, we find that smaller portfolios, constrained by cardinality to include only a subset of available assets, can achieve diversification similar to market portfolios, while reducing turnover and simplifying analysis. This suggests that focusing on smaller, strategically selected portfolios could offer investors efficient and cost-effective outcomes.

We contribute to the literature by providing a tool that enables investors to achieve sufficient diversification while satisfying their preference for small portfolios. Using formal statistical tests, we identify a range of portfolio sizes that achieve diversification benefits comparable to those of larger portfolios. In doing so, we offer a new perspective on a fundamental question in portfolio theory: how many stocks are sufficient for a well-diversified portfolio? Our framework provides investors with a practical tool to understand and navigate the trade-off between diversification and the complexity of estimation.

The remainder of the paper is organized as follows. Section 2.2 reviews the mathematical models considered in this study. Section 2.3 presents the tractable yet tight relaxations that are essential for developing our algorithm. Section 2.4 provides details of our SDP-based branch-and-bound algorithm, and Section 2.5 introduces the data and reports the

empirical results. Section 2.6 concludes.

2.2 Mathematical Models

The mean-variance model, introduced by Markowitz (1952b), is a cornerstone of modern portfolio theory and provides a quantitative framework for balancing risk and return. In this study, we address a cardinality-constrained variant that explicitly limits the number of active assets while permitting short-selling. This yields the following cardinality-constrained mean-variance (CCMV) model:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & \mathbf{w}^\top \Sigma \mathbf{w} \\ \text{s.t.} \quad & \mathbf{e}^\top \mathbf{w} = 1, \\ & \mu^\top \mathbf{w} = r, \\ & \mathbf{l} \leq \mathbf{w} \leq \mathbf{u}, \\ & \|\mathbf{w}\|_0 \leq s, \end{aligned} \tag{CCMV}$$

where $\Sigma \in \mathcal{S}^n_+$ denotes the symmetric positive semidefinite covariance matrix, $\mu \in \mathbb{R}^n$ is the vector of expected returns, and $\mathbf{e} \in \mathbb{R}^n$ is the vector of all ones. The vectors $\mathbf{l}, \mathbf{u} \in \mathbb{R}^n$ define lower and upper bounds on portfolio weights, satisfying $l_i \leq w_i \leq u_i$ for all $i \in [1:n]$. The term $\|\mathbf{w}\|_0 := |\{i : w_i \neq 0\}|$ represents the cardinality of the portfolio, restricting it to at most s assets. We define a portfolio $\mathbf{w}$ as *s-sparse* if $\|\mathbf{w}\|_0 \leq s$. Without loss of generality, we assume $0 \in [l_i, u_i]$ for all $i \in [1:n]$. This assumption implies that any asset can be inactive ($w_i = 0$). We distinguish between the general case where short-selling is permitted ($l_i < 0$ for some i) and the "no-short-selling" case ($l \geq 0$). Note that cases where $l_i > 0$ are excluded from the sparsity considerations, as such assets are forced to be active by the linear constraints.

Problem (CCMV) is NP-hard. This complexity arises directly from the cardinality constraint; indeed, Natarajan (1995) establishes that selecting a subset of variables to optimize a quadratic objective subject to linear constraints is NP-hard. While (CCMV) reduces to a tractable convex quadratic program (QP) if the support of $\mathbf{w}$ is fixed, the number of possible supports grows combinatorially with n, rendering approaches intractable.

For generality and notational convenience, we embed (CCMV) into a broader class of cardinality-constrained QP:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n_+} \quad & \mathbf{x}^\top \mathsf{Q} \mathbf{x} \\ \text{s.t.} \quad & \mathsf{A}\mathbf{x} = \mathbf{b}, \\ & \|\mathbf{x}\|_0 \leq s, \end{aligned} \tag{CCQP}$$

where $\mathsf{Q} \in \mathcal{S}^n$, $\mathsf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Clearly, (CCQP) encompasses (CCMV) as a special case. The formulation (CCMV) is a structured instance of (CCQP). This is evident by decomposing the weight vector into positive and negative parts, $\mathbf{w} = \mathbf{w}^+ - \mathbf{w}^-$

with $\mathbf{w}^+, \mathbf{w}^- \in \mathbb{R}^n_+$, and defining the nonnegative state vector $\mathbf{x} := (\mathbf{w}^+, \mathbf{w}^-)^\top \in \mathbb{R}^{2n}_+$. The sparsity structure is preserved as $\|\mathbf{w}\|_0 = \|\mathbf{w}^+\|_0 + \|\mathbf{w}^-\|_0 = \|\mathbf{x}\|_0$, and the linear constraints are encoded via A and $\mathbf{b}$ in the extended space. Consequently, we focus our analysis on the general model (CCQP).

A standard approach to modeling cardinality is the introduction of auxiliary binary variables to enforce complementarity. Consider the following mathematical program with complementarity constraints (QPCC):

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n_+,\, \mathbf{v}\in\{0,1\}^n} \quad & \mathbf{x}^\top \mathsf{Q}\mathbf{x} \\ \text{s.t.} \quad & \mathsf{A}\mathbf{x} = \mathbf{b}, \\ & \mathbf{e}^\top \mathbf{v} = n - s, \\ & \mathbf{x}^\top \mathbf{v} = 0. \end{aligned} \tag{QPCC}$$

Any feasible solution to (QPCC) is s-sparse. Since $\mathbf{x} \geq 0$ and $\mathbf{v} \geq 0$, the orthogonality condition $\mathbf{x}^\top \mathbf{v} = 0$ implies $x_i v_i = 0$ for all i. If $\mathbf{x}$ were s'-sparse with $s' > s$, at least s' components of $\mathbf{x}$ would be nonzero, forcing the corresponding v_i to zero. This would imply $\mathbf{e}^\top \mathbf{v} \leq n - s' < n - s$, contradicting the feasibility of $\mathbf{v}$. Thus, exact sparsity is guaranteed.

Remark 1. The binary constraint $\mathbf{v} \in \{0,1\}^n$ in (QPCC) can be relaxed to the box constraint $0 \leq \mathbf{v} \leq 1$. This yields a continuous complementarity-based formulation where v_i acts as a surrogate indicator for the inactivity of x_i. This continuous relaxation will be the basis for the conic models developed later.

One might alternatively model sparsity using a big-M formulation, which only involves linear constraints and binary variables. However, there are several drawbacks of this approach in our setting:

- If the feasible set $\{\mathbf{x} \in \mathbb{R}^n_+ : \mathsf{A}\mathbf{x} = \mathbf{b}\}$ is unbounded, selecting a valid Big-M constant is difficult. Overly large values of M weaken the continuous relaxation, impeding branch-and-bound performance, while overly small values may cut off the global optimum.
- Even with carefully tuned M-values, the convex relaxations obtained from big-M formulations are typically weaker than those derived from complementarity-based reformulations for cardinality-constrained quadratic problems Bomze et al. (2025a,b). In particular, they do not exploit the quadratic structure of the problem as effectively.

For these reasons, we work with the complementarity formulation (QPCC) as the starting point for our conic relaxations.

Recent advances in copositive optimization Burer (2009); Bomze and Peng (2023) show that (QPCC) admits an exact reformulation as a linear problem over the completely positive (CP) cone, defined by

$$\mathcal{CP}^n := \left\{ \mathsf{S} \in \mathcal{S}^n \;:\; \exists\, r \in \mathbb{N}_+,\ \mathsf{F} \in \mathbb{R}^{r\times n}_+ \text{ such that } \mathsf{S} = \mathsf{F}^\top \mathsf{F} \right\}.$$

This approach yields a convex extended formulation in which the quadratic term $\mathbf{x}\mathbf{x}^\top$ is lifted to a matrix variable constrained to lie in $\mathcal{CP}^n$. While optimization over $\mathcal{CP}^n$ is computationally intractable, it provides a powerful modeling paradigm. By replacing $\mathcal{CP}^n$ with tractable inner approximations, such as the doubly nonnegative (DNN) cone $\mathcal{D}^n := \mathcal{S}^n_+ \cap \mathbb{R}^{n\times n}_+$, we obtain strong conic relaxations. These can be embedded within a branch-and-bound scheme to ensure global optimality. In the remainder of this section, we exploit this framework to derive specialized DNN and SDP relaxations tailored to the CCMV structure.

2.3 Tractable Yet Tight Relaxations

This section develops a hierarchy of conic relaxations for the CCMV model, derived from the copositive reformulation discussed in Section 2.2. Our objective is to balance tightness with scalability: while semidefinite relaxations offer strong bounds, they scale poorly with the number of assets n. Conversely, second-order cone (SOC) representations are computationally lighter but potentially yield weaker bounds. We demonstrate how to exploit the specific structure of the CCMV model—namely box constraints, sign patterns, and covariance factorizations—to construct relaxations that are both tight and computationally efficient.

2.3.1 DNN Relaxations for (CCMV)

We begin with a generic DNN relaxation, obtained by replacing the intractable CP cone with the DNN cone in the lifted reformulation of (QPCC). Following the template provided by Bomze et al. (2025a), this yields:

$$
\begin{aligned}
\min_{\mathsf{X},\mathbf{v}} \quad & \langle \mathsf{Q}, \mathsf{X} \rangle \\
\text{s.t.} \quad & \mathsf{A}\mathbf{x} = \mathbf{b} \\
& (\mathsf{A}\mathsf{X}\mathsf{A}^\top) = \mathbf{b} \circ \mathbf{b} \\
& \mathbf{e}^\top \mathbf{v} = n - s \\
& \begin{pmatrix} 1 & x_i & v_i \\ x_i & X_{ii} & 0 \\ v_i & 0 & v_i \end{pmatrix} \in \mathcal{S}^3_+, \quad \forall i \in [1:n] \\
& \begin{pmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathsf{X} \end{pmatrix} \in \mathcal{D}^{n+1}.
\end{aligned}
\tag{DNN}
$$

While (DNN) serves as a standard template, we now derive specialized variants that exploit the specific structure of (CCMV).

2.3.2 Relaxations for (CCMV) without short-selling

In the no-short-selling regime, the non-negativity constraint $\mathbf{w} \geq 0$ eliminates the need for a sign decomposition of $\mathbf{w}$. We address the upper bound constraints $\mathbf{w} \leq \mathbf{u}$ via two

distinct approaches:

1. introducing slack variables and reformulating the problem as a QPCC to which the DNN relaxation (DNN) is applied; or
2. directly imposing suitable bounds on the lifted variables in the DNN relaxation.

We show that while both approaches yield the same lower bounds, the second is computationally superior due to the reduced dimension of the lifted cone.

The DNN relaxation with slack variables for handling the box constraints is given by

$$
\begin{aligned}
\min_{\mathsf{Y},\mathbf{v}} \quad & \langle \Sigma, \mathsf{Y}^{\mathbf{ww}} \rangle \\
\text{s.t.} \quad & \mathbf{e}^\top \mathbf{w} = 1, \quad \mu^\top \mathbf{w} = r \\
& \langle \mathsf{E}, \mathsf{Y}^{\mathbf{ww}} \rangle = 1, \quad \langle \mu\mu^\top, \mathsf{Y}^{\mathbf{ww}} \rangle = r^2 \\
& \mathbf{w} + \mathbf{z} = \mathbf{u} \\
& (\mathsf{Y}^{\mathbf{ww}} + \mathsf{Y}^{\mathbf{zz}} + 2\mathsf{Y}^{\mathbf{wz}}) = \mathbf{u} \circ \mathbf{u} \\
& \mathbf{e}^\top \mathbf{v} = n - s \\
& \begin{pmatrix} 1 & w_i & v_i \\ w_i & Y_{ii}^{\mathbf{ww}} & 0 \\ v_i & 0 & v_i \end{pmatrix} \in \mathcal{S}^3_+, \quad \forall i \in [1:n] \\
& \mathsf{Y} := \begin{pmatrix} 1 & \mathbf{w}^\top & \mathbf{z}^\top \\ \mathbf{w} & \mathsf{Y}^{\mathbf{ww}} & \mathsf{Y}^{\mathbf{wz}_1} \\ \mathbf{z} & (\mathsf{Y}^{\mathbf{wz}})^\top & \mathsf{Y}^{\mathbf{zz}} \end{pmatrix} \in \mathcal{D}^{2n+1}.
\end{aligned}
\tag{DNN-SlackBox}
$$

Alternatively, we propose a more compact relaxation that avoids slack variables by enforcing the box constraints directly on the lifted matrix W:

$$
\begin{aligned}
\min_{\mathsf{W},\mathbf{w},\mathbf{v}} \quad & \langle \Sigma, \mathsf{W} \rangle \\
\text{s.t.} \quad & \mathbf{e}^\top \mathbf{w} = 1, \quad \mu^\top \mathbf{w} = r \\
& \langle \mathsf{E}, \mathsf{W} \rangle = 1, \quad \langle \mu\mu^\top, \mathsf{W} \rangle = r^2 \\
& (\mathsf{W}) \leq \mathbf{u} \circ \mathbf{u} \\
& \mathbf{e}^\top \mathbf{v} = n - s \\
& \begin{pmatrix} 1 & w_i & v_i \\ w_i & W_{ii} & 0 \\ v_i & 0 & v_i \end{pmatrix} \in \mathcal{S}^3_+, \quad \forall i \in [1:n] \\
& \begin{pmatrix} 1 & \mathbf{w}^\top \\ \mathbf{w} & \mathsf{W} \end{pmatrix} \in \mathcal{D}^{n+1}, \quad \begin{pmatrix} 1 & (\mathbf{u}-\mathbf{w})^\top \\ \mathbf{u}-\mathbf{w} & \operatorname{Diag}(\mathbf{u} \circ \mathbf{u} - \operatorname{diag}(\mathsf{W})) \end{pmatrix} \in \mathcal{S}^{n+1}_+.
\end{aligned}
\tag{DNN-Box}
$$

Theorem 2.1. *The relaxations* (DNN-SlackBox) *and* (DNN-Box) *are equivalent. Specifically, the projection of the feasible set of* (DNN-SlackBox) *onto the subspace of variables* $(\mathsf{W}, \mathbf{w}, \mathbf{v})$ *coincides with the feasible set of* (DNN-Box)*. Consequently, both formulations yield the same optimal lower bound for the original problem* (CCMV) *for any covariance matrix* Σ*.*

Proof. We establish the equivalence by showing mutual inclusion of the feasible sets projected onto the common variables $(\mathsf{W}, \mathbf{w}, \mathbf{v})$. Since the objective function $\langle \Sigma, \mathsf{W} \rangle$ depends only on W, the optimal values must coincide.

First, let $(\mathsf{Y}, \mathbf{v})$ be feasible for (DNN-SlackBox) and define $\mathsf{W} := \mathsf{Y}^{\mathbf{ww}}$. Then $(\mathsf{W}, \mathbf{w}, \mathbf{v})$ satisfies all equality constraints of (DNN-Box) by construction. The diagonal inequality

$$(\mathsf{W}) \leq \mathbf{u} \circ \mathbf{u}$$

follows directly from

$$(\mathsf{Y}^{\mathbf{ww}} + \mathsf{Y}^{\mathbf{zz}} + 2\mathsf{Y}^{\mathbf{wz}}) = \mathbf{u} \circ \mathbf{u}$$

and the entrywise nonnegativity of $\mathsf{Y}^{\mathbf{zz}}$ and $\mathsf{Y}^{\mathbf{wz}}$.

It remains to verify the additional linear matrix inequality (LMI) in (DNN-Box). The principal block of Y corresponding to $\mathbf{z}$ satisfies

$$\begin{pmatrix} 1 & \mathbf{z}^\top \\ \mathbf{z} & \mathsf{Y}^{\mathbf{zz}} \end{pmatrix} \succeq \mathsf{O},$$

which, by Schur complementation, implies $\mathsf{Y}^{\mathbf{zz}} - \mathbf{z}\mathbf{z}^\top \succeq 0$. In addition,

$$(\mathsf{Y}^{\mathbf{zz}}) = \mathbf{u} \circ \mathbf{u} - \mathsf{W} - 2(\mathsf{Y}^{\mathbf{wz}}) \leq \mathbf{u} \circ \mathbf{u} - \mathsf{W},$$

so

$$(\mathbf{u} \circ \mathbf{u} - \mathsf{W}) - \mathbf{z}\mathbf{z}^\top \succeq ((\mathsf{Y}^{\mathbf{zz}})) - \mathbf{z}\mathbf{z}^\top \succeq 0,$$

which is exactly the added LMI in (DNN-Box) with $\mathbf{z} = \mathbf{u} - \mathbf{w}$. Hence $(\mathsf{W}, \mathbf{w}, \mathbf{v})$ is feasible for (DNN-Box).

Conversely, suppose that $(\mathsf{W}, \mathbf{w}, \mathbf{v})$ is feasible for (DNN-Box). Define

$$\mathsf{Y}^{\mathbf{ww}} := \mathsf{W}, \qquad \mathsf{Y}^{\mathbf{wz}} := \mathsf{O}, \qquad \mathsf{Y}^{\mathbf{zz}} := (\mathbf{u} \circ \mathbf{u} - \mathsf{W}), \qquad \mathbf{z} := \mathbf{u} - \mathbf{w},$$

and put

$$\mathsf{Y} = \begin{pmatrix} 1 & \mathbf{w}^\top & \mathbf{z}^\top \\ \mathbf{w} & \mathsf{Y}^{\mathbf{ww}} & \mathsf{Y}^{\mathbf{wz}} \\ \mathbf{z} & (\mathsf{Y}^{\mathbf{wz}})^\top & \mathsf{Y}^{\mathbf{zz}} \end{pmatrix}.$$

Then

$$(\mathsf{Y}^{\mathbf{ww}} + \mathsf{Y}^{\mathbf{zz}} + 2\mathsf{Y}^{\mathbf{wz}}) = (\mathsf{W}) + (\mathbf{u} \circ \mathbf{u} - \mathsf{W}) = \mathbf{u} \circ \mathbf{u},$$

so the diagonal identity in (DNN-SlackBox) holds. From the DNN constraint in (DNN-Box), we have $\mathbf{w} \geq \mathbf{o}$ and $\mathsf{W} \geq \mathsf{O}$ entrywise, and $\mathsf{W} \leq \mathbf{u} \circ \mathbf{u}$ implies $0 \leq w_i \leq u_i$, hence

$\mathbf{z} = \mathbf{u} - \mathbf{w} \geq 0$. Moreover, $\mathsf{Y}^{\mathbf{wz}} = \mathsf{O}$, and $\mathsf{Y}^{\mathbf{zz}}$ is diagonal with nonnegative entries, so $\mathsf{Y} \geq \mathsf{O}$ entrywise.

Finally, applying the Schur complement with respect to the upper-left scalar 1,

$$\mathsf{Y} \succeq \mathsf{O} \quad \Longleftrightarrow \quad \begin{pmatrix} \mathsf{W} - \mathbf{w}\mathbf{w}^\top & \mathbf{0} \\ \mathbf{0} & \mathsf{Y}^{\mathbf{zz}} - \mathbf{z}\mathbf{z}^\top \end{pmatrix} \succeq \mathsf{O}\,.$$

The first block is PSD because $\left(\begin{smallmatrix} 1 & \mathbf{w}^\top \\ \mathbf{w} & \mathsf{W} \end{smallmatrix}\right) \in \mathcal{D}^{n+1}$, and the second block is $(\mathbf{u}\circ\mathbf{u} - \mathsf{W}) - \mathbf{z}\mathbf{z}^\top \succeq \mathsf{O}$ by the LMI in (DNN-Box). Hence $\mathsf{Y} \succeq \mathsf{O}$, and $(\mathsf{Y}, \mathbf{v})$ is feasible for (DNN-SlackBox).

Therefore, the feasible sets of the two relaxations coincide when projected onto $(\mathsf{W}, \mathbf{w}, \mathbf{v})$, and since the objectives agree, the optimal values coincide. □

Remark 2. As established in Theorem 2.1, it is unnecessary to introduce slack variables and embed them into a high-dimensional DNN matrix in order to model the box constraints. Instead, one can impose a single low-dimensional LMI that guarantees the existence of a valid slack completion. This reformulation is computationally more attractive: the largest semidefinite block has dimension $n + 1$ rather than $2n + 1$, and no additional cross-terms involving slack variables appear in the lifted version. Consequently, both the number of decision variables and the size of the conic constraints are reduced, leading to smaller blocks and faster interior-point iterations in practice.

Remark 3. The additional LMI in (DNN-Box) can equivalently be expressed by the following system of rotated SOC constraints:

$$\begin{aligned} &(u_i^2 - W_{ii},\, t_i,\, u_i - w_i) \in \mathcal{Q}_r, \quad \forall i \in [1:n], \\ &\sum_{i=1}^{n} t_i \leq \tfrac{1}{2}, \qquad t_i \geq 0, \end{aligned}$$

where

$$\mathcal{Q}_r := \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \;:\; 2x_1x_2 \geq x_3^2,\; x_1 \geq 0,\; x_2 \geq 0\right\}$$

denotes the *rotated second-order cone*. This representation replaces a single $(n+1)\times(n+1)$ semidefinite constraint with n small three-dimensional cones and one linear coupling constraint. As a result, the relaxation can be handled by SOCP solvers, which typically scale more favorably with n, and are more numerically robust than general-purpose SDP solvers.

2.3.3 Relaxations for (CCMV) with short-selling

When short-selling is permitted, the weights $\mathbf{w}$ are no longer non-negative. A natural approach is to decompose the weights as $\mathbf{w} := \mathbf{w}^p - \mathbf{w}^m$ and apply the relaxation (DNN-Box) developed for the non-negative case. However, this leads to a substantially larger conic formulation: the dimension of the lifted matrix essentially doubles, and the resulting SDP cone, intersected with a polyhedral set involving $O(n^2)$ nonnegativity

constraints, quickly becomes computationally intractable for moderate values of n. To improve scalability, we drop the bound-enforcing LMIs, which yields a slightly weaker but significantly more efficient relaxation. As shown below, this relaxation retains its strength even after eliminating the explicit sign decomposition.

We first present the relaxation obtained by applying the sign decomposition to (DNN-Box), relaxing the DNN cone to an SDP cone, and discarding the bound-forcing LMIs:

$$
\begin{aligned}
\min_{\mathsf{Y}\,\mathbf{v}^p,\mathbf{v}^m} \quad & \langle \Sigma, \mathsf{Y}^{pp} + \mathsf{Y}^{mm} - 2\mathsf{Y}^{pm} \rangle \\
\text{s.t.} \quad & \mathbf{e}^\top(\mathbf{w}^p - \mathbf{w}^m) = 1, \\
& \langle \mathsf{E}, \mathsf{Y}^{pp} + \mathsf{Y}^{mm} - 2\mathsf{Y}^{pm} \rangle = 1, \\
& \mu^\top(\mathbf{w}^p - \mathbf{w}^m) = r, \\
& \langle \mu\mu^\top, \mathsf{Y}^{pp} + \mathsf{Y}^{mm} - 2\mathsf{Y}^{pm} \rangle = r^2, \\
& \mathbf{e}^\top \mathbf{v}^p + \mathbf{e}^\top \mathbf{v}^m = n - s, \\
& 0 \le \mathbf{w}^p \le \mathbf{u}, \qquad 0 \le \mathbf{w}^m \le -\ell, \\
& \begin{pmatrix} 1 & w_i^p & v_i^p \\ w_i^p & Y_{ii}^{pp} & 0 \\ v_i^p & 0 & v_i^p \end{pmatrix} \in \mathcal{S}^3_+, \qquad \begin{pmatrix} 1 & w_i^m & v_i^m \\ w_i^m & Y_{ii}^{mm} & 0 \\ v_i^m & 0 & v_i^m \end{pmatrix} \in \mathcal{S}^3_+, \quad \forall i \in [1:n], \\
& \mathsf{Y} = \begin{pmatrix} 1 & (\mathbf{w}^p)^\top & (\mathbf{w}^m)^\top \\ \mathbf{w}^p & \mathsf{Y}^{pp} & \mathsf{Y}^{pm} \\ \mathbf{w}^m & (\mathsf{Y}^{pm})^\top & \mathsf{Y}^{mm} \end{pmatrix} \in \mathcal{S}^{2n+1}_+.
\end{aligned}
\tag{SDP-sign-decomposition}
$$

A more compact alternative is obtained by working directly with the signed variable $\mathbf{w}$:

$$
\begin{aligned}
\min_{\mathsf{W}\in\mathcal{S}^n,\,\mathbf{w},\mathbf{v}\in\mathbb{R}^n} \quad & \langle \Sigma, \mathsf{W} \rangle \\
\text{s.t.} \quad & \mathbf{e}^\top \mathbf{w} = 1, \quad \mu^\top \mathbf{w} = r, \\
& \langle \mathsf{E}, \mathsf{W} \rangle = 1, \quad \langle \mu\mu^\top, \mathsf{W} \rangle = r^2, \\
& \ell \le \mathbf{w} \le \mathbf{u}, \\
& \mathbf{e}^\top \mathbf{v} = n - s, \\
& \begin{pmatrix} 1 & w_i & v_i \\ w_i & W_{ii} & 0 \\ v_i & 0 & v_i \end{pmatrix} \in \mathcal{S}^3_+, \qquad \forall i \in [1:n], \\
& \overline{\mathsf{W}} := \begin{pmatrix} 1 & \mathbf{w}^\top \\ \mathbf{w} & \mathsf{W} \end{pmatrix} \in \mathcal{S}^{n+1}_+ .
\end{aligned}
\tag{SDP-direct}
$$

Proposition 2.1. *The relaxations* (SDP-sign-decomposition) *and* (SDP-direct) *are equivalent; that is, they yield the same optimal objective value:*

$$\text{val}\,(\text{SDP-sign-decomposition}) = \text{val}\,(\text{SDP-direct}).$$

Proof. We first observe that (SDP-direct) is obtained from (SDP-sign-decomposition) by aggregating the variables

$$\mathbf{w} = \mathbf{w}^p - \mathbf{w}^m, \qquad \mathbf{v} = \mathbf{v}^p + \mathbf{v}^m, \qquad \mathsf{W} = \mathsf{Y}^{pp} + \mathsf{Y}^{mm} - 2\mathsf{Y}^{pm},$$

and then discarding the sign–decomposition structure. For any feasible point of (SDP-sign-decomposition), these aggregated quantities satisfy all constraints in (SDP-direct), and the objective values coincide. Thus

$$\text{val}\,(\text{SDP-direct}) \leq \text{val}\,(\text{SDP-sign-decomposition}).$$

It remains to prove the reverse inequality. Let $(\mathsf{W}, \mathbf{w}, \mathbf{v})$ be any feasible triple of (SDP-direct). Define the sign–decomposition

$$w_i^p := \max\{w_i, 0\}, \qquad w_i^m := \max\{-w_i, 0\}, \qquad \forall\, i \in [1:n],$$

so that $\mathbf{w}^p, \mathbf{w}^m \geq 0$ and $\mathbf{w}^p - \mathbf{w}^m = \mathbf{w}$. Since $\mathbf{l} \leq \mathbf{w} \leq \mathbf{u}$ and $l_i < 0 < u_i$, these automatically satisfy

$$0 \leq w_i^p \leq u_i, \qquad 0 \leq w_i^m \leq -l_i.$$

From the local PSD constraints in (SDP-direct), considering the principal submatrix associated $1, 3$ elements, we obtain in particular $v_i \in [0, 1]$. We therefore split $\mathbf{v}$ according to the sign of $\mathbf{w}$:

$$v_i^p := \big(1 - \text{sign}(w_i^p)\big)\, v_i, \qquad v_i^m := \text{sign}(w_i^p)\, v_i,$$

where $\text{sign}(t) = 1$ for $t > 0$ and 0 otherwise. Then $v_i^p, v_i^m \geq 0$, $v_i^p + v_i^m = v_i$, and hence

$$\mathbf{e}^\top v^p + \mathbf{e}^\top v^m = \mathbf{e}^\top v = n - s.$$

Moreover, the linear constraints

$$\mathbf{e}^\top(\mathbf{w}^p - \mathbf{w}^m) = \mathbf{e}^\top \mathbf{w} = 1, \qquad \mu^\top(\mathbf{w}^p - \mathbf{w}^m) = \mu^\top \mathbf{w} = r,$$

are satisfied by construction.

Define the residual matrix

$$\mathsf{M} := \mathsf{W} - \mathbf{w}\mathbf{w}^\top,$$

which is PSD due to the lifted matrix of variable $\mathbf{w}$ in (SDP-direct) is PSD. We choose $\alpha, \beta > 0$ with $\alpha + \beta = 1$ and construct:

$$\mathsf{Y}^{pp} := \mathbf{w}^p(\mathbf{w}^p)^\top + \alpha\mathsf{M}, \qquad \mathsf{Y}^{mm} := \mathbf{w}^m(\mathbf{w}^m)^\top + \beta\mathsf{M}, \qquad \mathsf{Y}^{pm} := \mathbf{w}^p(\mathbf{w}^m)^\top.$$

A direct calculation verifies

$$\begin{aligned}\mathsf{Y}^{pp} + \mathsf{Y}^{mm} - 2\mathsf{Y}^{pm} &= (\mathbf{w}^p - \mathbf{w}^m)(\mathbf{w}^p - \mathbf{w}^m)^\top + (\alpha + \beta)\mathsf{M} \\ &= \mathbf{w}\mathbf{w}^\top + \mathsf{M} = \mathsf{W},\end{aligned}$$

ensuring that the objective value and global moment constraints match.

It remains to verify the local PSD constraints. For each i, consider the matrices

$$M_i^p := \begin{pmatrix} 1 & w_i^p & v_i^p \\ w_i^p & Y_{ii}^{pp} & 0 \\ v_i^p & 0 & v_i^p \end{pmatrix}, \qquad M_i^m := \begin{pmatrix} 1 & w_i^m & v_i^m \\ w_i^m & Y_{ii}^{mm} & 0 \\ v_i^m & 0 & v_i^m \end{pmatrix}.$$

From $\mathsf{M} \succeq \mathsf{O}$ we have $M_{ii} \geq 0$, and hence

$$Y_{ii}^{pp} = (w_i^p)^2 + \alpha M_{ii} \geq (w_i^p)^2, \qquad Y_{ii}^{mm} = (w_i^m)^2 + \beta M_{ii} \geq (w_i^m)^2.$$

For each i, exactly one of w_i^p and w_i^m is positive (or both are zero). If $w_i^p > 0$, then $v_i^p = 0$, $v_i^m = v_i \in [0,1]$, and

$$M_i^p = \begin{pmatrix} 1 & w_i^p & 0 \\ w_i^p & Y_{ii}^{pp} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad M_i^m = \begin{pmatrix} 1 & 0 & v_i \\ 0 & Y_{ii}^{mm} & 0 \\ v_i & 0 & v_i \end{pmatrix}.$$

Since $Y_{ii}^{pp} \geq (w_i^p)^2$ and $v_i \in [0,1]$, both blocks are PSD. A symmetric argument holds if $w_i^m > 0$. If $w_i = 0$, then $w_i^p = w_i^m = 0$ and one of v_i^p, v_i^m equals $v_i \in [0,1]$, both matrices reduce to PSD. Thus both $M_i^p \succeq 0$ and $M_i^m \succeq 0$ for all $i \in [1:n]$, and all local 3×3 constraints of (SDP-sign-decomposition) are satisfied.

Finally, consider the lifted matrix

$$\mathsf{Y} := \begin{pmatrix} 1 & (\mathbf{w}^p)^\top & (\mathbf{w}^m)^\top \\ \mathbf{w}^p & \mathsf{Y}^{pp} & \mathsf{Y}^{pm} \\ \mathbf{w}^m & (\mathsf{Y}^{pm})^\top & \mathsf{Y}^{mm} \end{pmatrix}.$$

With the above choices one can write

$$\mathsf{Y} = \underbrace{\begin{pmatrix} 1 \\ \mathbf{w}^p \\ \mathbf{w}^m \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{w}^p \\ \mathbf{w}^m \end{pmatrix}^\top}_{\succeq 0} + \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & \alpha\mathsf{M} & 0 \\ 0 & 0 & \beta\mathsf{M} \end{pmatrix}}_{\succeq 0},$$

which is a sum of two positive semidefinite matrices. Hence $\mathsf{Y} \in \mathcal{S}_+^{2n+1}$.

We have therefore constructed a feasible point of (SDP-sign-decomposition) with the same objective value as $(\mathsf{W}, \mathbf{w}, \mathbf{v})$, proving

$$\text{val}\,(\text{SDP-sign-decomposition}) \leq \text{val}\,(\text{SDP-direct}).$$

Combined with the inequality in the opposite direction, this yields the claimed equality of optimal values. □

Alternatively, we may introduce the shifted variable $\mathbf{x} = \mathbf{w} - \mathbf{l}$ and obtain the following DNN relaxation, defining $R := r + \mu^\top \mathbf{l}$:

$$
\begin{aligned}
\min_{\mathsf{X}\in\mathcal{S}^{n+1},\mathbf{x},\mathbf{v}\in\mathbb{R}^n} \quad & \langle \Sigma, \mathsf{X} + 2\mathbf{l}\mathbf{x}^\top + \mathbf{l}\mathbf{l}^\top \rangle \\
\text{s.t.} \quad & \mathbf{e}^\top(\mathbf{x}+\mathbf{l}) = 1, \quad \langle \mathsf{E}, \mathsf{X} + 2\mathbf{l}\mathbf{x}^\top + \mathbf{l}\mathbf{l}^\top \rangle = 1, \\
& \mu^\top(\mathbf{x}+\mathbf{l}) = R, \quad \langle \mu\mu^\top, \mathsf{X} + 2\mathbf{l}\mathbf{x}^\top + \mathbf{l}\mathbf{l}^\top \rangle = R^2, \\
& \mathbf{x} \le \mathbf{u} - \mathbf{l}, \quad \mathbf{e}^\top \mathbf{v} = n - s, \\
& \begin{pmatrix} 1 & x_i & v_i \\ x_i & X_{ii} & -l_i v_i \\ v_i & -l_i v_i & v_i \end{pmatrix} \in \mathcal{S}^3_+, \quad \forall i \in [1:n], \\
& \overline{\mathsf{X}} := \begin{pmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathsf{X} \end{pmatrix} \in \mathcal{D}^{n+1}.
\end{aligned}
\tag{SDP-shifted}
$$

We compare (SDP-direct) with the shifted relaxation (**??**). The two formulations are linked by the affine change of variables

$$
\mathbf{x} := \mathbf{w} - \mathbf{l}, \qquad \mathsf{W} = \mathsf{X} + \mathbf{l}\mathbf{x}^\top + \mathbf{x}\mathbf{l}^\top + \mathbf{l}\mathbf{l}^\top . \tag{2.1}
$$

Under this transformation, the linear constraints and objective functions of both models are identical. The structural distinction lies solely in the conic lifting: while (SDP-direct) requires the lifted matrix to be positive semidefinite ($\overline{\mathsf{W}} \in \mathcal{S}^{n+1}_+$), problem (SDP-shifted) imposes the stricter condition that the shifted lifted matrix lies in the doubly nonnegative cone, $\overline{\mathsf{X}} \in \mathcal{D}^{n+1}$.

Since $\mathcal{D}^{n+1} \subset \mathcal{S}^{n+1}_+$, the shifted formulation essentially appends entrywise non-negativity constraints to the standard SDP relaxation. This observation leads directly to the following dominance result.

Proposition 2.2. *The optimal value of the shifted relaxation provides a lower bound at least as tight as that of the direct relaxation:*

$$
\text{val}\,(\text{SDP-direct}) \;\le\; \text{val}\,(\text{SDP-shifted}).
$$

Proof. Let $(\mathsf{X}, \mathbf{x}, \mathbf{v})$ be feasible for (SDP-shifted). The affine mapping (2.1) yields a point $(\mathsf{W}, \mathbf{w})$ that satisfies the linear constraints of (SDP-direct) by construction. Furthermore, since $\mathsf{X} \in \mathcal{D}^{n+1} \subset \mathcal{S}^{n+1}_+$, the positive semidefiniteness of W follows from the linearity of the transformation. Thus, the feasible set of (SDP-shifted) maps into a subset of the feasible set of (SDP-direct), implying the inequality in objective values. □

We now establish conditions under which this relaxation gap vanishes. For any feasible solution $(\mathsf{W}, \mathbf{w})$ of (SDP-direct), let $\mathsf{M} := \mathsf{W} - \mathbf{w}\mathbf{w}^\top \succeq \mathsf{O}$ denote the convexity gap. In terms of the shifted variables, the lifted matrix decomposes as $\mathsf{X} = \mathbf{x}\mathbf{x}^\top + \mathsf{M}$. Consequently, the stronger DNN requirement in (SDP-shifted) is satisfied if and only if X is entrywise non-negative. This reduces the equivalence question to a specific structural property of the optimal solution.

Assumption 2.1. *There exists an optimal solution* $(\mathsf{W}^*, \mathbf{w}^*, \mathbf{v}^*)$ *of* (SDP-direct) *such that the associated shifted lifted matrix is non-negative. That is, with* $\mathsf{M}^* := \mathsf{W}^* - \mathbf{w}^*(\mathbf{w}^*)^\top$, *we have:*

$$X^*_{ij} \;=\; (w^*_i - l_i)(w^*_j - l_j) + M^*_{ij} \;\geq\; 0, \qquad \forall i, j \in [1:n]$$

Proposition 2.3. *If Assumption 2.1 holds, then the two relaxations are equivalent:*

$$\text{val}\,(\text{SDP-direct}) \;=\; \text{val}\,(\text{SDP-shifted}).$$

Proof. Suppose Assumption 2.1 holds. We construct a candidate solution for (SDP-shifted) using the optimal variables from (SDP-direct) via the transformation $\mathbf{x}^* := \mathbf{w}^* - \mathbf{l}$ and $\mathsf{X}^* := \mathsf{W}^* - \mathbf{l}(\mathbf{x}^*)^\top - \mathbf{x}^*\mathbf{l}^\top - \mathbf{l}\mathbf{l}^\top$. By the assumption, $\mathsf{X}^* \geq \mathsf{O}$. Moreover, since W^* is PSD, which ensures X^* is PSD. Thus, $\mathsf{X}^* \in \mathcal{D}^{n+1}$, making the candidate solution feasible for (SDP-shifted). Since the objective values match, the dominance established in Proposition 2.2 becomes an equality. □

Remark 4. Assumption 2.1 is an intrinsic property of the optimal solution, effectively postulating that the geometry of the relaxation aligns with the non-negativity of the DNN cone. Theoretically, this alignment is guaranteed in the limit of zero correlation (diagonal Σ), where the relaxation becomes exact and the rank-1 solution $\mathsf{W}^* = \mathbf{w}^*(\mathbf{w}^*)^\top$ naturally satisfies $(w^*_i - l_i)(w^*_j - l_j) \geq 0$. By perturbation arguments, this property extends to instances where Σ is diagonally dominant or possesses a Stieltjes-like inverse—common features in diversified equity markets where idiosyncratic variances often outweigh pairwise correlations. In such regimes, the computationally cheaper (SDP-direct) inherits the tightness of the heavier (SDP-shifted) without the cost of $O(n^2)$ additional scalar inequalities.

The global semidefinite constraint in (SDP-direct) involves a matrix of dimension $(n+1)\times(n+1)$, rendering the formulation computationally prohibitive for large-scale instances (e.g., $n > 1000$). To overcome this bottleneck, we propose a scalable approximation that exploits the factorization of the covariance matrix, sacrificing the global consistency of the off-diagonal entries in exchange for computational speed.

Since Σ is symmetric positive semidefinite, it admits a factorization $\Sigma = \mathsf{F}\mathsf{F}^\top$ (e.g., via Cholesky decomposition), where $\mathsf{F} \in \mathbb{R}^{n\times k}$ and $k \leq n$. The portfolio variance can thus be expressed as:

$$\langle \Sigma, \mathsf{W} \rangle = \langle \mathsf{F}\mathsf{F}^\top, \mathsf{W} \rangle = \|\mathsf{F}^\top \mathbf{w}\|_2^2.$$

We introduce the auxiliary vector $\mathbf{z} \in \mathbb{R}^k$ to represent the factor exposure, defined as $\mathbf{z} := \mathsf{F}^\top \mathbf{w}$. The variance minimization reduces to minimizing the Euclidean norm $\|\mathbf{z}\|_2^2 = \sum_{j=1}^k z_j^2$. We linearize this objective by introducing scalar auxiliary variables ζ_j for $j \in [1:k]$ and imposing the hypograph constraints $\zeta_j \geq z_j^2$, which are representable via rotated second-order cones (or equivalently, 2×2 PSD blocks):

$$\begin{pmatrix} 1 & z_j \\ z_j & \zeta_j \end{pmatrix} \succeq \mathsf{O}, \qquad j \in [1:k].$$

Under this transformation, the variance term in the objective function is replaced by the linear sum $\sum_{j=1}^{k} \zeta_j$.

We retain the local 3×3 semidefinite blocks from (SDP-direct) to enforce the logical relationship between the continuous weights w_i and the cardinality indicators v_i. This yields the relaxation:

$$
\begin{aligned}
\min_{\mathbf{w},\mathbf{v},\mathbf{z},\zeta,\mathbf{W}} \quad & \sum_{j=1}^{k} \zeta_j \\
\text{s.t.} \quad & \mathbf{z} = \mathsf{F}^\top \mathbf{w}, \\
& \mathbf{e}^\top \mathbf{w} = 1, \qquad \mu^\top \mathbf{w} = r, \\
& \mathbf{l} \le \mathbf{w} \le \mathbf{u}, \\
& \mathbf{e}^\top \mathbf{v} = n - s, \\
& \begin{pmatrix} 1 & w_i & v_i \\ w_i & W_{ii} & 0 \\ v_i & 0 & v_i \end{pmatrix} \in \mathcal{S}_+^3, \qquad \forall i \in [1:n], \\
& \begin{pmatrix} 1 & z_j \\ z_j & \zeta_j \end{pmatrix} \in \mathcal{S}_+^2, \qquad \forall j \in [1:k].
\end{aligned}
\tag{SDP-compact}
$$

Problem (SDP-compact) is a relaxation of (SDP-direct). By discarding the global PSD requirement on the lifted matrix W, we reduce the complexity from imposing constraints on $O(n^2)$ variables to $O(n)$ variables.While this approach sacrifices theoretical tightness by ignoring off-diagonal correlation consistency, it gains significant computational tractability, offering a scalable solution for high-dimensional cardinality-constrained problems.

Conclusion. This subsection establishes a rigorous hierarchy of conic approximations for the CCMV problem. We proved that the (SDP-direct) formulation eliminates the computational redundancy of (SDP-sign-decomposition) without compromising bound quality. While the (SDP-shifted) relaxation theoretically dominates the direct approach, we identified structural conditions under which the two are equivalent, justifying the use of the lighter formulation in practice. Finally, for large-scale instances, the (SDP-compact) relaxation leverages covariance factorization to trade global correlation consistency for scalability. Collectively, these formulations provide a unified framework that allows the optimizer to adapt the relaxation strength to the problem dimension and covariance structure.

2.4 SDP-based Branch-and-Bound Algorithm

In this section, we propose a branch-and-bound (B&B) framework that integrates the SDP relaxations developed above to globally solve the CCMV problem (CCMV). We branch on the variable $\mathbf{v}$ in the search tree; once $\mathbf{v}$ becomes binary at a node, the corresponding relaxation is exact.

A B&B algorithm is essentially determined by three design components: (i) the strategy for computing lower and upper bounds at each explored node; (ii) the branching rule; and (iii) the node-selection (search) strategy. We first describe our bound-generation mechanisms.

2.4.1 Lower Bound

Lower bounds are crucial in B&B, as they certify the suboptimality of the current incumbent and enable pruning of subtrees. At a given node, we define the optimal value of the chosen relaxation as the *local lower bound*. The *global lower bound* is then the minimum of these local values over all active (unexplored) nodes. A strong global lower bound allows early fathoming of nodes and prevents uncontrolled growth of the search tree. The choice of relaxation at each node directly affects both the tightness of the lower bound and the computational cost, so a carefully designed selection strategy is essential.

In the case of (CCMV) *without* short-selling, our default choice is the relaxation (DNN-Box), solved by an off-the-shelf SDP solver such as `MOSEK`, which relies on an interior-point method. Recent advances in low-rank augmented Lagrangian methods (ALM) for doubly nonnegative relaxations have also been reported in the literature Hou et al. (2025a,b), and we recommend these as promising alternatives in large-scale settings.

Our computational experience indicates that the spectral spread of the local objective Hessian has a pronounced impact on the quality and robustness of the relaxation (DNN-Box). We therefore introduce a user-specified threshold τ on the range of eigenvalues of the local Hessian. If this range exceeds τ, we bypass the more expensive DNN-based relaxation and solve instead the compact relaxation (SDP-compact); otherwise, we stick to (DNN-Box). In this way, nodes associated with ill-conditioned quadratic terms are handled by the more scalable second-order-cone-based model.

In the case of (CCMV) *with* short-selling, we adopt a similar rule. When the eigenvalue range at a node is larger than τ, we again solve the compact relaxation (SDP-compact); otherwise we solve the stronger semidefinite relaxation (SDP-direct). This adaptive scheme balances tightness and computational effort across the tree.

2.4.2 Upper Bound

The convex relaxations developed above not only provide lower bounds at each node of the B&B tree, but also yield structural information that can be exploited to construct strong upper bounds efficiently. Central to this approach is the surrogate vector v^*, which arises in the relaxation as a continuous proxy for the support of an optimal sparse portfolio.

Support information encoded by v^*. For each asset i, the local semidefinite constraint

$$\begin{pmatrix} 1 & w_i & v_i \\ w_i & W_{ii} & 0 \\ v_i & 0 & v_i \end{pmatrix} \succeq \mathsf{O}\,, \qquad i \in [1:n]\,,$$

enforces $0 \le v_i \le 1$ and couples v_i with (w_i, W_{ii}) so that v_i acts as a convex surrogate of the binary indicator $\mathbf{1}_{\{i\in[1:n]:w_i\neq 0\}}$. At optimality, variance minimization typically drives $W_{ii} \approx w_i^2$, which pushes v_i toward 1 for assets playing an active role in the portfolio and toward 0 for negligible assets. Together with the constraint $\mathbf{e}^\top \mathbf{v} = n - s$, the optimizer allocates large v_i^* values to the assets most important for risk reduction. Thus the ordering of the components of v^* provides a natural importance ranking.

Iterative relaxation-guided reduction. At each node, we exploit this ranking to shrink the asset universe before solving the original nonconvex problem. Starting from the full relaxation, we sort the assets by v_i^* and discard those with the smallest scores. We then re-solve the relaxation on the reduced index set and obtain an updated v^*. This process is repeated for a few rounds. Assets that consistently exhibit small v_i^* values across iterations can be removed safely, while the remaining set S contains those most relevant for constructing a high-quality sparse portfolio.

Reduced feasibility problem. Once a sufficiently small set S has been identified, we compute an upper bound by solving a reduced cardinality-constrained problem restricted to the assets in S. Let $|S| = s + h$ with $h \ge 0$ representing an over-selection margin. We solve

$$\begin{aligned} \min_{\mathbf{w}_S\in\mathbb{R}^{s+h},\, \mathbf{v}\in\{0,1\}^{s+h}} \quad & \mathbf{w}_S^\top \Sigma \mathbf{w}_S \\ \text{s.t.} \quad & \mathbf{e}^\top \mathbf{w}_S = 1\,, \qquad \mu^\top \mathbf{w}_S = r\,, \\ & \mathbf{l} \le \mathbf{w}_S \le \mathbf{u}\,, \\ & \mathbf{w}_{Si}\, v_i = 0\,, \qquad i \in [1:s+h]\,, \\ & \mathbf{e}^\top \mathbf{v} = h\,, \end{aligned} \tag{CC2MV-S}$$

which enforces that exactly h assets in S are set to zero. A feasible solution for the original problem is then obtained by

$$w_i^* = \begin{cases} 0, & i \notin S, \\ (w_S^*)_i, & i \in S, \end{cases}$$

and its objective value serves as a valid upper bound for the node.

Theoretical justification. We now justify the validity and effectiveness of this procedure.

Proposition 2.4 (Validity of $\mathbf{v}^*$-guided upper bounds)**.** *Let $S \subseteq [1:n]$ be any index set and let $\widehat{\mathbf{w}}_S$ be feasible for (CC2MV-S). Extend $\widehat{\mathbf{w}}_S$ by zeros on $[1:n] \setminus S$. Then the extended vector $\widehat{\mathbf{w}}$ is feasible for the original node and*

$$f(\widehat{\mathbf{w}}) \ge z^*,$$

where z^* is the true optimal value at the node. Hence any feasible point constructed from (CC2MV-S) yields a valid upper bound.

To understand why v^* identifies promising sets S, we impose the following idealized assumption, under which the relaxation is fully informative.

Assumption 2.2 (Exact surrogate at a node)**.** *The relaxation at a node admits an optimal solution* $(\mathsf{W}^*, \mathbf{w}^*, \mathbf{v}^*)$ *such that (i)* $\mathbf{v}^*$ *is integral with* $\sum_i v_i^* = n - s$*; (ii)* $v_i^* = 1$ *if and only if* $w_i^* \neq 0$*; (iii) the relaxation is tight, i.e.,* $f(\mathbf{w}^*) = z^*$*.*

Proposition 2.5 (Optimal support recovery under ideal surrogate)**.** *Under Assumption 2.2, the set* $S^* := \{i : v_i^* = 1\}$ *satisfies* $|S^*| \leq s$ *and*

$$z^*(S^*) = z^*.$$

Thus retaining the indices with the largest values of v_i^* *preserves an optimal support of the original sparse portfolio problem.*

Although $\mathbf{v}^*$ may not be integral in practice, it remains a reliable continuous importance score, and the assets with the largest v_i^* consistently correspond to those most influential in high-quality sparse portfolios. Consequently, the iterative $\mathbf{v}^*$-guided reduction followed by solving (CC2MV-S) yields tight upper bounds at substantially reduced computational cost.

2.4.3 Branching Rule

A natural branching variable is a component of the surrogate vector $\mathbf{v}^*$ obtained from the relaxation. Since v_i is designed to approximate the binary indicator $\mathbf{1}\{w_i \neq 0\}$, fractional values of v_i^* signal ambiguity regarding whether asset i should be active in the optimal sparse portfolio. A simple and effective rule is therefore to branch on the most fractional component of $\mathbf{v}^*$:

$$i^* \;=\; \arg\min_{i \in [1:n]} |v_i^* - 0.5|.$$

This targets the variable for which the relaxation is least decisive about the inclusion ($v_i = 1$) or exclusion ($v_i = 0$) of the asset, and branching on i^* typically yields the largest reduction in the feasible region.

More refined rules may exploit additional structure provided by the relaxation. For example, one may combine sparsity information with the magnitude of the shifted weight variable $\mathbf{x}^*$ and branch on

$$i^* \;=\; \arg\max_{i \in [1:n]} \frac{(x_i^*)^2}{1 - v_i^*},$$

which gives priority to assets that exhibit large exposure (as measured by $|x_i^*|$) but whose selection status remains uncertain ($v_i^* \notin \{0, 1\}$). The numerator reflects the asset's contribution to portfolio variance, whereas the denominator penalizes variables that are already nearly fixed by the relaxation. Such hybrid criteria can be particularly effective in CCMV problems, where the interaction between variance reduction (captured by x^*) and sparsity promotion (captured by $\mathbf{v}^*$) plays a key role.

Other sophisticated rules, including pseudocost or strong-branching variants adapted to semidefinite relaxations, are also possible. The pair $(\mathbf{x}^*, \mathbf{v}^*)$ encapsulates both magnitude-based and sparsity-based information, making the SDP relaxations well suited for designing effective branching strategies in cardinality-constrained portfolio optimization.

2.4.4 Search Rule

We employ a dynamic node-selection strategy that adapts to the progress of the B&B search. At each iteration, we record the most recent node that produced an improvement in either the global lower bound (via a tighter relaxation) or the global upper bound (via a stronger feasible solution).

If the latest improvement was in the *lower bound*, we switch to a *best-bound* strategy: among all active nodes, we select the one with the smallest local lower bound. This focuses computational efforts on nodes whose relaxations appear most promising for further tightening the global bound and thus speeds up convergence from below.

Conversely, if the latest improvement was in the *upper bound*, we switch to a *best-incumbent* (best-optimal-solution) strategy: among active nodes, we select the one with the lowest local upper bound, as obtained from the reduced feasibility problem. This guides the search toward regions most likely to produce an improved incumbent portfolio, thereby enhancing the global upper bound.

This adaptive switching rule balances exploration and exploitation. The algorithm focuses on lower-bound improvement when the relaxation is highly informative and shifts emphasis to upper-bound improvement when promising candidate supports emerge. Such a dynamic strategy is well suited for CCMV problems, where the strength of relaxations and the quality of feasible portfolios may vary significantly across nodes due to both sparsity patterns and risk characteristics.

2.5 Empirical Results

In this section, we examine our methods in different settings and discuss how they can be used and what we can learn from real-world data. We will mainly focus on two parts. In the first subsection, we will report results from the Global Minimum-Variance Portfolio. In this part, we isolate the uncertainty stemming from covariance estimation to facilitate comparison with alternative approaches. In the second part, we report results for the Mean–Variance Portfolio, which gives us a flavor of how cardinality performs when we include return prediction. In general, our setup can accommodate additional constraints that investors may wish to impose; we leave this for future research.

We use daily data obtained from the Center for Research in Security Prices (CRSP), covering the period from 01/01/1995 to 12/29/2023. The empirical analysis follows a *rolling-sample* approach. Consistent with common practice, 21 consecutive trading days are treated as one "month." The out-of-sample period spans from 01/14/2000 to 12/29/2023, resulting in a total of 287 "months" (6,027 trading days). Portfolios are rebalanced monthly, with the number of shares held remaining constant within each

"month," implying no transaction costs during this period. We denote the investment dates by $t = 1, \ldots, 287$.

For each combination (T, N), the investment universe is defined as the set of N stocks with complete return histories over both the most recent T trading days and the subsequent 21 trading days. At each investment date t, returns from the previous T days are used to estimate the covariance matrix $\widehat{\boldsymbol{\Sigma}}_t$ required to implement a given strategy. The estimated parameters determine the portfolio weights $\widehat{\boldsymbol{w}}_t$ for month $t+1$. This rolling window advances monthly, adding 21 new observations and discarding the oldest 21. The outcome of this procedure is a time series of 6,027 out-of-sample daily returns generated by each portfolio strategy. The following portfolios are included in our empirical study:

1. **Sample:** The estimator $\hat{\Sigma}$ is the sample covariance matrix.
2. **1/N:** Equal-weighted portfolio.
3. **GMV-LS:** Global minimum variance portfolio. The estimator $\hat{\Sigma}$ is the shrinkage estimator proposed by Ledoit and Wolf (2004).
4. **GMV-QIS:** Global minimum variance portfolio. The estimator $\hat{\Sigma}$ is the shrinkage estimator proposed by Ledoit and Wolf (2022b).
5. **BNB-s:** Global minimum variance portfolio. The estimator $\hat{\Sigma}$ is the sample covariance matrix, and the cardinality constraint requires the number of selected stocks to be less than or equal to s.
6. **BNB-LSs:** Global minimum variance portfolio. The estimator $\hat{\Sigma}$ is the shrinkage estimator proposed by Ledoit and Wolf (2004), with a cardinality constraint requiring the number of selected stocks to be less than or equal to s.
7. **MV-LS:** Mean-variance portfolio. The estimator $\hat{\Sigma}$ is the shrinkage estimator proposed by Ledoit and Wolf (2004).
8. **MV-QIS:** Mean-variance portfolio. The estimator $\hat{\Sigma}$ is the shrinkage estimator proposed by Ledoit and Wolf (2022b).
9. **MVBNB-s:** Mean-variance portfolio. The estimator $\hat{\Sigma}$ is the sample covariance matrix, with a cardinality constraint requiring the number of selected stocks to be less than or equal to s.

We evaluate portfolio performance using the following measures:

- **Annualized Volatility**: Out-of-sample annualized volatility is calculated as:

$$\hat{\sigma}_{\text{annual}} = \hat{\sigma}_{\text{daily}} \times \sqrt{252}, \tag{2.2}$$

  where the $\hat{\sigma}_{\text{daily}}$ is daily sample standard deviation. To test whether the volatility of two strategies is statistically different, we also compute the p-value of the difference, following the approach suggested by Ledoit and Wolf (2008).

- **Maximum Drawdown (MDD)**: The largest peak-to-trough decline in portfolio value:

$$\text{MDD} = \max_{\tau \in [0,T]} \left[\frac{\sup_{t \in [0,\tau]} P_t - P_\tau}{\sup_{t \in [0,\tau]} P_t} \right] \tag{2.3}$$

  where P_t denotes the portfolio value at time t.

- **Sharpe Ratio**: Risk-adjusted return using total volatility:

$$\text{SR} = \frac{E[R_p - R_f]}{\sigma_p} \tag{2.4}$$

  where R_p is portfolio return, R_f is risk-free rate, and σ_p is portfolio volatility. To test whether the Sharpe ratios of two strategies are statistically different, we again compute the p-value of the difference, using the approach suggested by Ledoit and Wolf (2008).

- **Sortino Ratio**: Risk-adjusted return using downside volatility:

$$\text{SoR} = \frac{E[R_p - R_f]}{\sigma_{\text{downside}}} \tag{2.5}$$

  where σ_{downside} considers only returns below a target.

- **Average Assets**: The mean number of assets held in the portfolio:

$$\bar{N} = \frac{1}{T} \sum_{t=1}^{T} \|\mathbf{w}_t\|_0 \tag{2.6}$$

  where $\|\mathbf{w}_t\|_0$ counts non-zero weights at time t.

- **Turnover**: To quantify the trading activity required to maintain each portfolio strategy, we calculate the portfolio turnover as:

$$\text{Turnover} = \sum_{j=1}^{N_{t+1}} \left| \hat{w}_{j,t+1} - \hat{w}_{j,t^+} \right|, \tag{2.7}$$

  where $\hat{w}_{j,t^+}$ and $\hat{w}_{j,t+1}$ denote the portfolio weights of asset j before and after rebalancing at $t+1$, respectively. To compute turnover, we consider the union of the investment universes at t and $t+1$, since the portfolio is constructed from the N largest-capitalization stocks at each rebalancing date. As a result, some firms may drop out while new ones enter, implying $N_{t+1} \geq N$ and that N_{t+1} generally varies over time.

- **CE**: The certainty-equivalent (CE) return represents the risk-free rate that an investor would accept instead of adopting a given portfolio strategy:

$$\widehat{\text{CE}} = \widehat{\mu}_p - \frac{\gamma}{2} \widehat{\sigma}_p^2, \tag{2.8}$$

  where γ is the coefficient of relative risk aversion. We report results for $\gamma = 4$. For each portfolio, $\widehat{\mu}_p$ and $\widehat{\sigma}_p$ are defined as the sample mean of excess returns and the sample portfolio volatility, respectively.

2.5.1 Sample Covariance Estimation as Our Benchmark

When T is Large for Covariance Estimation

In this section, we aim to demonstrate that for longer periods T, the sample covariance matrix $\hat{\Sigma}$ becomes an unbiased and a more reliable estimator. As a benchmark, we first consider the case where $N = 100$ and $T = 1250$, providing a long time series that enables accurate estimation of $\hat{\Sigma}$. This setting allows us to evaluate how cardinality constraints can assist investors who need to manage a small number of stocks.

Typically, investors have limited attention and resources, making it impractical to analyze all available stocks in detail. Our approach offers an interpretable and data-driven tool that helps investors select a manageable subset of stocks for in-depth analysis. In Table 2.1, we show that even with a small set of stocks, one can achieve a level of diversification and standard deviation comparable to that of a 100-stock portfolio. This finding indicates that the marginal benefit of including more stocks diminishes beyond a certain point. By combining modern statistical techniques with cardinality constraints, investors can construct well-diversified, low-risk portfolios using a relatively small number of assets. This provides practical guidance for initial stock selection without compromising risk control.

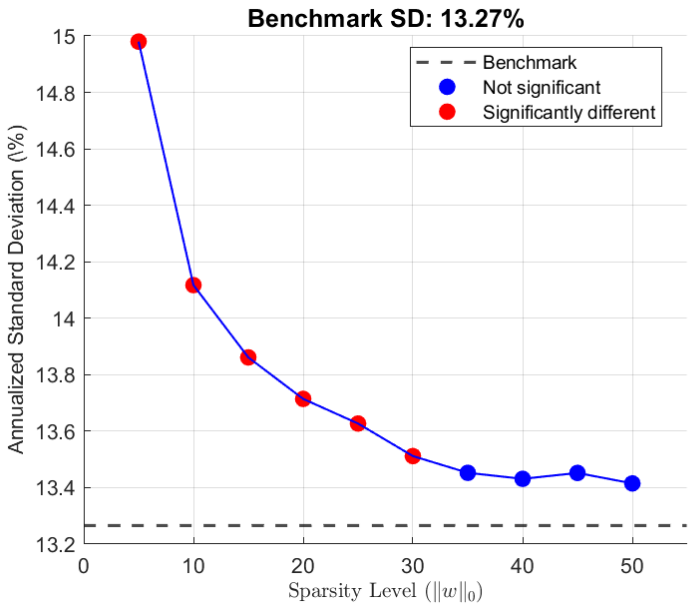
Table 2.1: Performance comparison of different portfolio strategies ($N = 100$, $T = 1250$).

Method	Standard deviation	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Turnover	CE	Avg. N
Benchmark							
EW	0.1932	0.5545	0.3612	0.4606	0.1036	-0.0049	100
GMV-Sample	0.1327	0.3577	0.4050	0.5267	0.7132	0.0185	100
GMV-LS	0.1322	0.3448	0.4178	0.5397	0.6350	0.0203	100
GMV-QIS	0.1315	0.3481	0.3983	0.5153	0.6213	0.0178	100
Our methods							
BNB-5	0.1498	0.3887	0.4072	0.5336	0.3250	0.0161	5
BNB-10	0.1412	0.3709	0.3841	0.4996	0.4125	0.0144	10
BNB-15	0.1386	0.3707	0.3802	0.4893	0.4932	0.0143	15
BNB-20	0.1371	0.3844	0.4102	0.5232	0.5648	0.0186	20
BNB-25	0.1363	0.3883	0.4097	0.5237	0.6355	0.0187	25
BNB-30	0.1351	0.3829	0.3914	0.5000	0.6319	0.0164	30
BNB-35	0.1345	0.3779	0.4160	0.5347	0.6677	0.0198	35
BNB-40	0.1343	0.3753	0.3718	0.4788	0.7052	0.0139	40
BNB-45	0.1345	0.3803	0.4111	0.5295	0.7187	0.0191	45
BNB-50	0.1342	0.3778	0.4007	0.5155	0.7373	0.0178	50

Our results highlight that market movements are often driven by a small subset of influential stocks. In traditional finance literature, techniques such as factor models or Principal Component Analysis (PCA) are commonly employed to understand underlying

market dynamics. However, these approaches are typically tailored for sophisticated investors and may lack intuitive interpretability for broader audiences.

Figure 2.1: Portfolio Volatility under Cardinality Constraints ($N = 100$, $T = 1250$).



Portfolio volatility based on the top 100 market-capitalization stocks, using historical data from 2000 to 2023, constrained by a maximum number of assets. Red nodes denote significance at the 0.01 level compared to the GMV portfolio without cardinality constraints.

In contrast, our approach offers direct interpretability by identifying which specific stocks are crucial for risk control and which stock sets are optimal for investment. This practical insight empowers investors to make more informed and strategic decisions. Additionally, our approach contributes to the advancement of fintech by providing actionable, data-driven tools that bridge the gap between complex financial models and practical investment strategies.

In Figure 2.1, we provide a detailed plot where red dots indicate that the portfolio is statistically different from the benchmark portfolio, which is the global minimum variance portfolio using the sample covariance estimator. In contrast, white dots indicate that the difference is statistically insignificant from the benchmark. We introduce a framework that allows investors to combine our algorithm with statistical tests to determine whether they are in a trade-off phase. We refer to this as the sparsity-risk trade-off. Due to the underlying structure, investors can benefit from a “free lunch", achieving the same level of diversification with a smaller set of stocks, up to a certain point. We observe that a portfolio with just 35 stocks is already well-diversified. However, if an investor seeks to manage an even smaller set of stocks, they must accept a higher level of risk. Our framework provides guidance on this process. As shown in our results, if an investor

is comfortable with a certain level of additional risk, we can offer a comparably small portfolio with as few as 5 stocks. From Table 2.1, we see that this smaller portfolio can achieve similar levels of maximum drawdown and Sharpe ratio. Most portfolios even achieve a lower average turnover, which implies lower transaction costs. The most interesting insight is that a small set of stocks can achieve a level of diversification comparable to that of the full sample covariance portfolio. When using a shrinkage estimator, the risk measures are similar. Since the time series is long, the shrinkage method offers limited additional benefits. However, it can still help reduce turnover and stabilize the benchmark portfolio.

When $N = 500$

After discussing the small-scale setting, one may ask: how does our algorithm perform when we have a large set of stocks? In this section, we investigate how the results change as N increases. We present two tables that separate the cases of a long time series T and a short time series, respectively. As is well known in the literature, when N is large and T is small, the sample covariance matrix becomes highly unstable. This instability is one of the main reasons why many investors prefer equal-weighted portfolios over modern optimization methods.

Table 2.2: Performance Comparison of Different Portfolio Strategies ($N = 500, T = 1250$)

Method	Standard deviation	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Turnover	CE	Avg. N
Benchmark							
EW	0.2020	0.5680	0.4615	0.5873	0.1005	0.0116	500
GMV-Sample	0.1170	0.3083	0.4383	0.5878	2.5060	0.0239	500
GMV-LS	0.1111	0.2902	0.5003	0.6592	1.8859	0.0309	500
GMV-QIS	0.1057	0.3093	0.5682	0.7297	1.1989	0.0377	500
Our methods							
BNB-5	0.1463	0.3757	0.5380	0.6937	0.4460	0.0359	5
BNB-10	0.1319	0.3420	0.5225	0.6835	0.6533	0.0341	10
BNB-15	0.1242	0.3275	0.5658	0.7526	0.7058	0.0394	15
BNB-20	0.1209	0.3328	0.5909	0.7876	0.8769	0.0422	20
BNB-25	0.1190	0.3055	0.5911	0.7818	0.8791	0.0420	25
BNB-30	0.1165	0.3338	0.5832	0.7613	0.9267	0.0408	30
BNB-35	0.1164	0.3333	0.5834	0.7621	0.9887	0.0408	35
BNB-40	0.1163	0.3388	0.6331	0.8235	1.0535	0.0466	40
BNB-45	0.1156	0.3415	0.5793	0.7546	1.1042	0.0402	45
BNB-50	0.1151	0.3396	0.5655	0.7389	1.1160	0.0386	50

By comparing the GMV-Sample portfolios in Table 2.2 and Table 2.3, we observe that the standard deviation and other risk measures change dramatically simply due to having

fewer observations to estimate the covariance matrix. This motivates the use of shrinkage estimators, as a better estimate of $\hat{\Sigma}$ leads to better-controlled portfolios.

First, we focus on Table 2.2. In this table, we observe that the GMV-Sample portfolio still exhibits a relatively low standard deviation. However, it already suffers from the high dimensionality, which reduces its Sharpe ratio compared to the equal-weight portfolio and leads to a high turnover. As expected, both the linear and nonlinear shrinkage estimators of the covariance matrix significantly improve portfolio performance, primarily by maintaining a similar level of risk while substantially increasing the Sharpe ratio. Furthermore, our method shows that simply adding a cardinality constraint can also produce a low-risk portfolio with a higher Sharpe ratio and lower turnover. Therefore, we conclude that when estimation uncertainty is high, one can either use a better covariance matrix estimator or select a smaller subset of assets—both approaches help mitigate the impact of estimation errors.

Table 2.3: Performance Comparison of Different Portfolio Strategies ($N = 500, T = 600$)

Method	Standard deviation	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Turnover	CE	Avg. N
Benchmark							
EW	0.2020	0.5680	0.4615	0.5873	0.1005	0.0116	500
GMV-Sample	0.1958	0.6267	0.2254	0.3349	12.3525	-0.0325	500
GMV-LS	0.1172	0.3129	0.5037	0.7005	3.2095	0.0316	500
GMV-QIS	0.1053	0.2811	0.5150	0.6663	1.4803	0.0320	500
Our methods							
BNB-5	0.1464	0.3709	0.4150	0.5612	0.9835	0.0179	5
BNB-10	0.1381	0.4518	0.4485	0.6131	1.2169	0.0238	10
BNB-15	0.1331	0.4444	0.5097	0.6808	1.4043	0.0324	15
BNB-20	0.1325	0.3448	0.4850	0.6397	1.7152	0.0291	20
BNB-25	0.1295	0.3214	0.6056	0.8030	1.8637	0.0449	25
BNB-30	0.1275	0.3248	0.6210	0.8405	2.0481	0.0467	30
BNB-35	0.1261	0.3386	0.6500	0.8874	2.2273	0.0502	35
BNB-40	0.1239	0.3126	0.6487	0.8975	2.3708	0.0497	40
BNB-45	0.1255	0.3484	0.5468	0.7478	2.5241	0.0371	45
BNB-50	0.1244	0.3829	0.5744	0.7895	2.6910	0.0405	50

From Table 2.3, we can see that in the high-dimensionality setup, the GMV-Sample portfolio no longer yields meaningful results. Shrinkage is necessary in this setup to achieve good performance, and GMV-QIS performed very well. However, we observe that simply adding a cardinality constraint to GMV-Sample significantly improves portfolio diversification and greatly reduces the turnover, which makes the sample covariance matrix useful for portfolio management. Furthermore, we have seen that with just around 35 stocks, we can achieve quite a good Sharpe ratio and a low-risk portfolio. Thus, we can conclude that cardinality constraints help investors avoid the influence of high estimation

error even when the input data is very disadvantageous. One might be interested to see what happens if we combine modern shrinkage methods with cardinality constraints. We will provide a detailed analysis and some insights in the following section.

Cardinality constraints show another approach to improving portfolio performance in high-dimensional settings. There are several practical reasons for imposing such constraints. Firstly, they can reduce transaction costs associated with trading and managing a large number of assets. Secondly, they can lower the management and monitoring overhead required for extensive portfolios. Thirdly, limiting the number of holdings can improve the interpretability and implementability of the portfolio strategy. This section implies that in situations of high uncertainty, selecting a smaller set of assets might be an effective way to mitigate the influence of estimation error than relying solely on improving the covariance matrix estimate through shrinkage.

From Table 2.3, we explore the complexity–risk trade-off. Our observations suggest that investors' preference for a small set of stocks can be interpreted as a strategy to mitigate estimation error. Ideally, if the true covariance matrix were known, the optimal diversified portfolio could be constructed directly. However, in practice, the covariance matrix must be estimated, and using these estimates in portfolio construction introduces substantial estimation errors. We have shown that a smaller set of assets can still yield a well-diversified portfolio. This finding is also consistent with empirical evidence that investors typically do not hold a large number of stocks. Our method could therefore provide a practical guideline for selecting stocks to include in a portfolio.

2.5.2 Combining Cardinality Constraints with Shrinkage Estimators

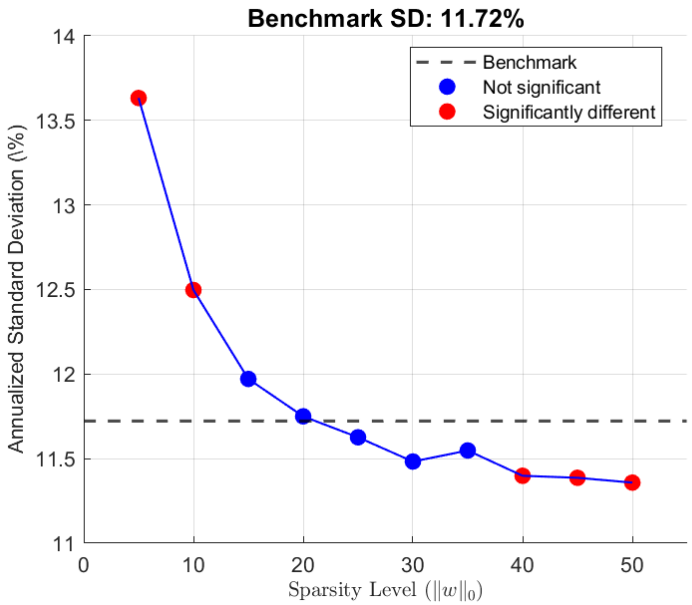
In the previous section, we showed that cardinality constraints are also useful when dealing with short time series data. In such cases, the covariance matrix is unstable—as is well known, it suffers from the limitations of a small dataset and can lead to poor portfolio decisions. By selecting a smaller dimension, such as 30, we can significantly reduce the standard deviation. When the dataset is small compared to the universe of stocks, statisticians typically apply shrinkage techniques.

In this section, we are also interested in exploring what can be achieved by combining both strategies. We find that combining linear shrinkage with cardinality constraints yields portfolios with similar risk levels and maximum drawdowns to those of GMV-LS, but with higher Sharpe ratios. Moreover, we observe better Sortino ratios and lower turnover. The results are even comparable to those of GMV-QIS, but with far fewer assets to monitor. In short, combining cardinality constraints with sophisticated shrinkage methods yields better results than using shrinkage alone. The cardinality constraint focuses on dimensionality reduction while preserving the structure of the chosen estimator, unlike methods such as LASSO or other norm-based regularizations that alter the estimation process. Thus, incorporating a cardinality constraint allows us to identify redundant assets while maintaining the integrity of the covariance matrix estimator. This enables us to benefit from the advantages of the shrinkage estimator while managing a smaller set of assets, thereby reducing turnover and, in practice, lowering transaction costs.

Table 2.4: Performance Comparison of Different Portfolio Strategies ($N = 500, T = 600$)

Method	Standard deviation	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Turnover	CE	Avg. N
Benchmark							
EW	0.2020	0.5680	0.4615	0.5873	0.1005	-0.0049	500
GMV-Sample	0.1958	0.6267	0.2254	0.3349	12.3525	-0.0325	500
GMV-LS	0.1172	0.3129	0.5037	0.7005	3.2095	0.0316	500
GMV-QIS	0.1053	0.2811	0.5150	0.6663	1.4803	0.0320	500
Our methods							
BNB-LS5	0.1363	0.2806	0.6968	0.9483	0.5808	0.0578	5
BNB-LS10	0.1250	0.2945	0.5162	0.7043	0.7195	0.0333	10
BNB-LS15	0.1197	0.2703	0.5008	0.6821	0.8356	0.0313	15
BNB-LS20	0.1175	0.2751	0.5457	0.7379	0.9286	0.0365	20
BNB-LS25	0.1163	0.2766	0.4911	0.6642	1.0128	0.0301	25
BNB-LS30	0.1148	0.2713	0.5768	0.7832	1.1525	0.0399	30
BNB-LS35	0.1155	0.3046	0.5597	0.7444	1.2504	0.0380	35
BNB-LS40	0.1140	0.3258	0.5812	0.7870	1.3448	0.0403	40
BNB-LS45	0.1139	0.3593	0.6194	0.8313	1.4290	0.0446	45
BNB-LS50	0.1136	0.3915	0.5501	0.7397	1.5112	0.0367	50

Figure 2.2: Portfolio Volatility under Crdinality Constraints ($N = 500$, $T = 600$).



Portfolio volatility based on the top 500 market-capitalization stocks, using historical data from 2000 to 2023, constrained by a maximum number of assets. Red nodes denote significance at the 0.05 level compared to the GMV-LS portfolio without cardinality constraints.

From Figure 2.2, we observe that a sparsity level above 15 is already statistically indistinguishable from GMV-LS. We also uncover an interesting pattern. When the sparsity level is very low, there is a clear trade-off between sparsity and risk. In contrast, when the number of selected stocks exceeds approximately 35, it is possible to construct a portfolio with significantly lower risk than the fully invested portfolio, albeit at the cost of holding more assets.

Comparing these results with the earlier findings in Table 2.1, we conclude that short time series lead to unstable covariance estimators and, consequently, high portfolio turnover. However, when a well-designed shrinkage estimator is employed, it is possible to achieve well-diversified portfolios using a much smaller subset of stocks.

That said, if an investor is willing to incorporate modern shrinkage estimators, they can construct an even better-diversified portfolio. The trade-off, therefore, is as follows: for investors who prefer not to use sophisticated shrinkage techniques, reducing the portfolio size can help manage estimation risk. However, when shrinkage estimators are used, the challenge becomes balancing sparsity against risk, as discussed in the previous section.

2.5.3 Markowitz Portfolio with Momentum Scores

We now consider a Markowitz portfolio incorporating a signal. Following Engle et al. (2019), we employ the momentum factor of Jegadeesh and Titman (1993). For each stock and investment period t, momentum is defined as the geometric average of the past 252 daily returns, excluding the most recent 21. The vector of momentum measures on each investment date constitutes the return-predictive signal μ in model (CCMV). The target expected return R is defined as the arithmetic mean momentum of an equal-weighted portfolio comprising the top-quintile stocks ranked by momentum.

Table 2.5 shows results consistent with Table 2.1. Imposing a cardinality constraint improves risk control for the MV-Sample method, which relies on the sample covariance matrix. With sufficiently long time series, the sample covariance is less affected by the curse of dimensionality, and thus MV-LS and MV-QIS offer only marginal improvements. However, the cardinality constraint enables lower-risk portfolios using fewer assets, which slightly enhances both the Sharpe and Sortino ratios, indicating reduced downside volatility. Relative to Table 2.1, the mean-variance portfolio exhibits higher standard deviation but also a higher Sharpe ratio. Incorporating a return-predictive signal further improves the Sortino ratio, though it also increases portfolio turnover, potentially raising transaction costs.

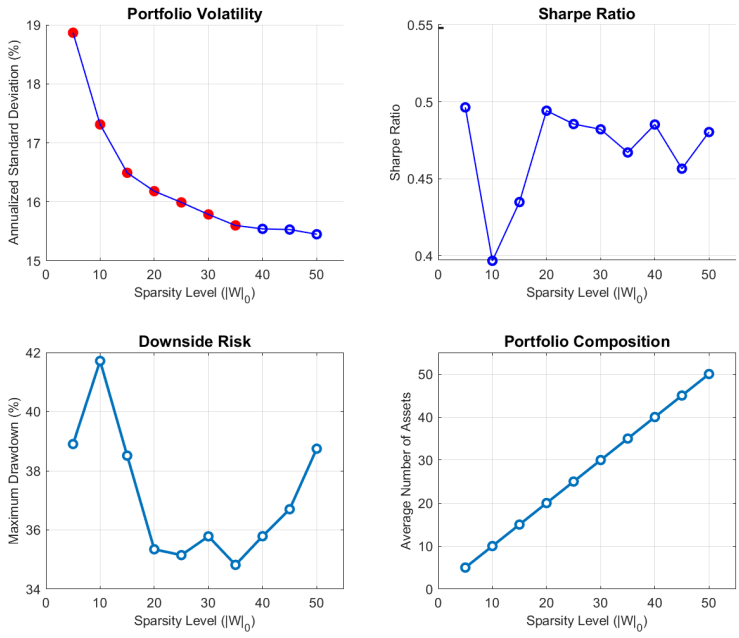
From Figure 2.3, we observe that portfolios with 40 or more stocks exhibit volatility levels comparable to those with 100 assets. The Sharpe ratio also stabilizes beyond this point. The maximum drawdown follows a U-shaped pattern, reaching its minimum between 20 and 40 assets. In terms of portfolio composition, we find that stocks frequently reach the cardinality upper bound imposed.

Compared to Figure 2.1, the inclusion of return-predictive signals improves the Sharpe ratio but also slightly alters the optimal sparsity structure. While Figure 2.1 suggests that a portfolio of 35 stocks is already well-diversified, the current setting requires around 40 assets to yield statistically distinct performance from a full 100-asset portfolio.

Table 2.5: Performance Comparison of Different Portfolio Strategies

Method	Standard deviation	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Turnover	CE	Avg. N
Benchmark							
EW	0.1932	0.5545	0.3612	0.4606	0.1036	-0.0049	100
MV-Sample	0.1530	0.3608	0.4244	0.5681	1.3188	0.0181	100
MV-LS	0.1526	0.3630	0.4418	0.5893	1.2172	0.0208	100
MV-QIS	0.1519	0.3742	0.4227	0.5621	1.2022	0.0181	100
Our methods							
MVBNB-5	0.1887	0.3890	0.4964	0.6579	1.1237	0.0225	5
MVBNB-10	0.1731	0.4172	0.3966	0.5357	1.2063	0.0087	10
MVBNB-15	0.1649	0.3851	0.4348	0.5806	1.1827	0.0173	15
MVBNB-20	0.1618	0.3535	0.4942	0.6525	1.2443	0.0276	20
MVBNB-25	0.1599	0.3515	0.4855	0.6418	1.2634	0.0265	25
MVBNB-30	0.1579	0.3578	0.4821	0.6393	1.3048	0.0263	30
MVBNB-35	0.1560	0.3482	0.4670	0.6225	1.3447	0.0242	35
MVBNB-40	0.1554	0.3579	0.4852	0.6454	1.3498	0.0271	40
MVBNB-45	0.1553	0.3670	0.4565	0.6076	1.3459	0.0227	45
MVBNB-50	0.1545	0.3875	0.4803	0.6393	1.3660	0.0265	50

Figure 2.3: MV Portfolio Measures Under Cardinality Constraints ($N = 100$, $T = 1250$).



Portfolio volatility based on the top 100 market-capitalization stocks, using historical data from 2000 to 2023, constrained by a maximum number of assets. Red nodes denote significance at the 0.01 level compared to the MV portfolio without cardinality constraints.

Our framework combines a tractable optimization algorithm with model-based statistical testing, enabling tailored portfolio construction that meets investor-specific preferences while providing robust, interpretable insights.

When $N = 500$

In this section, we investigate how the results change as N increases. We present two tables that distinguish between the cases of a long time series T and a short time series, respectively. As is well known in the literature, when N is large and T is small, the sample covariance matrix becomes highly unstable. This instability is one of the main reasons why many investors prefer equal-weighted portfolios over modern optimization methods.

By comparing the MV-Sample portfolios in Table **??** and Table **??**, we observe that the standard deviation and other risk measures change dramatically simply due to having fewer observations to estimate the covariance matrix. This observation motivates the use of shrinkage estimators, as a better estimate of $\hat{\Sigma}$ leads to better-controlled portfolios. As expected, both linear and nonlinear shrinkage estimators significantly improve portfolio performance, primarily by maintaining a similar level of risk while substantially increasing the Sharpe ratio.

Table 2.6: Performance comparison of different portfolio strategies ($N = 500$, $T = 1250$, mean-variance).

Method	Standard deviation	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Turnover	CE	Avg. N
Benchmark							
EW	0.2020	0.5680	0.4615	0.5873	0.1005	0.0116	500
MV-Sample	0.1312	0.3381	0.4072	0.5458	3.0996	0.0190	500
MV-LS	0.1253	0.3203	0.4740	0.6274	2.4262	0.0280	500
MV-QIS	0.1205	0.3385	0.5476	0.7096	1.6688	0.0369	500
Our methods							
MVBNB-5	0.1925	0.4664	0.1513	0.1943	1.3934	-0.0450	5
MVBNB-10	0.1695	0.4141	0.3400	0.4425	1.4688	0.0002	10
MVBNB-15	0.1605	0.3694	0.2725	0.3510	1.4985	-0.0078	15
MVBNB-20	0.1545	0.3573	0.3238	0.4280	1.6051	0.0023	20
MVBNB-25	0.1493	0.3505	0.3823	0.4995	1.6415	0.0125	25
MVBNB-30	0.1447	0.3507	0.5082	0.6691	1.6823	0.0317	30
MVBNB-35	0.1438	0.3508	0.4619	0.6015	1.7697	0.0251	35
MVBNB-40	0.1419	0.3561	0.4632	0.6089	1.8231	0.0255	40
MVBNB-45	0.1410	0.3974	0.4708	0.6189	1.8696	0.0266	45
MVBNB-50	0.1384	0.3808	0.4056	0.5302	1.9262	0.0178	50

Table 2.7: Performance Comparison of different portfolio strategies ($N = 500$, $T = 600$, mean-variance).

Method	Standard deviation	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Turnover	CE	Avg. N
Benchmark							
EW	0.2020	0.5680	0.4615	0.5873	0.1005	0.0116	500
MV-Sample	0.2222	0.5890	0.2451	0.3687	14.1334	-0.0443	500
MV-LS	0.1340	0.3196	0.4841	0.6839	3.8018	0.0290	500
MV-QIS	0.1204	0.3133	0.5185	0.6784	1.8935	0.0334	500
Our methods							
MVBNB-5	0.2095	0.4035	0.6152	0.8534	2.0794	0.0411	5
MVBNB-10	0.1788	0.2977	0.5646	0.7891	2.1493	0.0370	10
MVBNB-15	0.1680	0.4167	0.5202	0.7203	2.3738	0.0309	15
MVBNB-20	0.1625	0.3901	0.4513	0.6215	2.5560	0.0205	20
MVBNB-25	0.1605	0.3958	0.5178	0.7131	2.7148	0.0316	25
MVBNB-30	0.1582	0.4815	0.5053	0.6811	2.8796	0.0299	30
MVBNB-35	0.1511	0.4587	0.5433	0.7455	2.9418	0.0364	35
MVBNB-40	0.1525	0.3659	0.6440	0.8855	3.1748	0.0517	40
MVBNB-45	0.1513	0.3594	0.5807	0.8071	3.3074	0.0421	45
MVBNB-50	0.1502	0.3495	0.5800	0.8110	3.4666	0.0420	50

Table 2.8: ($N = 500$, $T = 600$, MV with LS).

Method	Standard deviation	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Turnover	CE	Avg. N
Benchmark							
EW	0.2020	0.5680	0.4615	0.5873	0.1005	0.0116	500
MV-Sample	0.2222	0.5890	0.2451	0.3687	14.1334	-0.0443	500
MV-LS	0.1340	0.3196	0.4841	0.6839	3.8018	0.0290	500
MV-QIS	0.1204	0.3133	0.5185	0.6784	1.8935	0.0334	500
Our methods							
MVBNB-LS5	0.1895	0.4359	0.1408	0.1856	1.5420	-0.0451	5
MVBNB-LS10	0.1634	0.4225	0.1993	0.2753	1.5668	-0.0208	10
MVBNB-LS15	0.1512	0.3558	0.2538	0.3488	1.5762	-0.0074	15
MVBNB-LS20	0.1459	0.2788	0.4171	0.5690	1.6646	0.0183	20
MVBNB-LS25	0.1427	0.2979	0.4324	0.5866	1.7279	0.0210	25
MVBNB-LS30	0.1397	0.3179	0.5390	0.7420	1.8120	0.0363	30
MVBNB-LS35	0.1373	0.3384	0.4569	0.6262	1.8690	0.0250	35
MVBNB-LS40	0.1365	0.2947	0.4670	0.6446	1.9496	0.0265	40
MVBNB-LS45	0.1349	0.3226	0.4642	0.6379	2.0460	0.0262	45
MVBNB-LS50	0.1342	0.2906	0.5559	0.7645	2.1190	0.0386	50

Furthermore, our method shows that simply adding a cardinality constraint can also produce low-risk portfolios with higher Sharpe ratios and lower turnover. In the short time-series setting shown in Table 2.6, restricting the number of stocks alone already leads to substantial improvements when data are limited. In particular, our method achieves a higher certainty-equivalent (CE) return than MV-QIS. Moreover, the performance of our approach is relatively insensitive to the specific covariance matrix estimator used.

Therefore, when estimation uncertainty is high, one can either employ a better covariance matrix estimator or select a smaller subset of assets—both approaches help mitigate the impact of estimation errors. A natural question is what happens when these two approaches are combined. As shown in Table 2.7, combining shrinkage estimation with a cardinality constraint yields further improvements, allowing us to outperform MV-LS by focusing on a smaller set of stocks.

2.5.4 Industry Composition Analysis

A natural question after imposing a cardinality constraint is how well the resulting smaller portfolio represents the industry composition of the original investment universe. To investigate this, we obtain industry codes from the CRSP database and map `siccd` codes to industry classifications following Bali et al. (2017). We then compare the proportion of industry composition in the portfolio after selecting 50 stocks from the original universe of 500 stocks. Specifically, we consider a scenario with a long estimation window ($T = 1250$ trading days) to estimate the sample covariance matrix and apply our algorithm to construct a 50-stock portfolio based on the global minimum variance portfolio from the 500-stock universe.

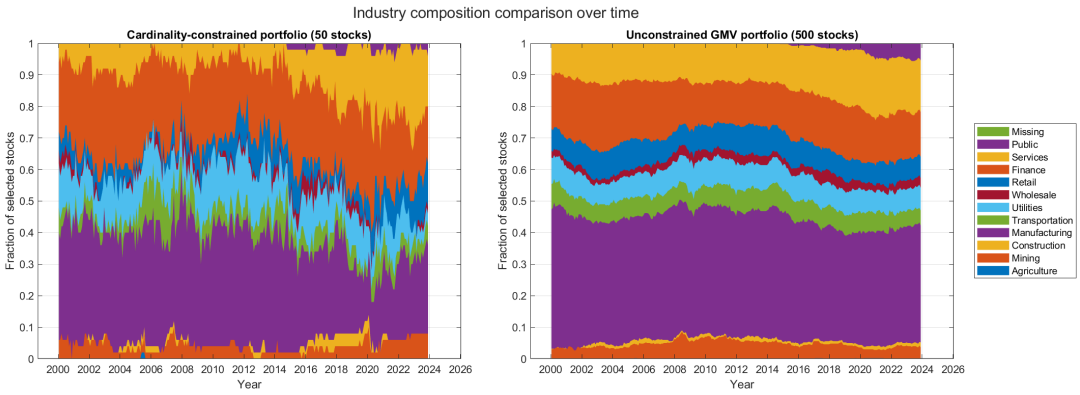


Figure 2.4: Industry composition over time for the cardinality-constrained portfolio (50 stocks) compared with the original 500-stock portfolio.

Figure 2.4 illustrates the industry composition of the selected portfolio compared with that of the original portfolio over time. Overall, the proportions of stocks across industries remain remarkably similar between the two portfolios throughout the sample period. This suggests that the cardinality constraint does not significantly distort the

structural composition of the portfolio across industries. Instead, the optimization procedure effectively reduces the dimensionality of the portfolio while preserving the economic structure of the investment universe. In particular, the selected portfolio does not become overly concentrated in a small subset of industries but maintains a diversified representation that closely mirrors the original universe. These findings indicate that cardinality-constrained optimization can substantially reduce the number of assets in the portfolio while retaining the key structural characteristics of the original portfolio.

This result suggests that the cardinality constraint primarily acts as a dimensionality reduction mechanism rather than fundamentally altering the economic structure of the portfolio.

2.6 Conclusion

Markowitz's portfolio theory emphasizes the construction of well-diversified portfolios, yet the question of how many assets constitute sufficient diversification remains open. While increasing the number of stocks can reduce idiosyncratic risk, it also raises the risk of estimation error and exposes portfolio optimization to the curse of dimensionality. This has led to substantial research focused on improving covariance and return estimators to manage large-scale portfolios effectively.

At the same time, Keynes's advocacy for concentrated investments remains relevant. Empirical studies show that many investors do not significantly expand portfolio size as the investable universe grows. Instead, they often prefer to focus on a manageable number of assets they can thoroughly analyze, potentially underutilizing modern quantitative tools designed for broad diversification.

Our framework bridges these two perspectives. It exploits available information to construct diversified portfolios while also providing a user-friendly framework for investors who prefer to manage a smaller number of assets. More importantly, our model provides a direct answer to the question: How many stocks are enough to achieve meaningful diversification? By combining cardinality constraints with formal statistical testing, the model identifies the point at which the marginal benefit of diversification diminishes. It thus helps investors quantify the trade-offs involved in managing smaller portfolios and the associated "price" of reduced diversification.

Importantly, the proposed cardinality constraints do not materially distort the structural composition of portfolios across industries. Instead, the optimization procedure primarily acts as a dimensionality reduction mechanism, preserving the broader economic structure of the investment universe while reducing portfolio complexity. This finding is particularly relevant for individual investors, who often prefer manageable portfolios without sacrificing the core benefits of diversification.

Overall, our findings suggest that Keynes's view remains highly relevant: a carefully selected subset of assets can often achieve risk-adjusted performance comparable to that of a substantially larger portfolio. Finally, the proposed framework also contributes to the growing FinTech literature by providing a flexible methodology that can be extended to incorporate transaction costs, short-selling constraints, and tail-risk considerations,

making it practical for real-world portfolio construction.

Appendix to Chapter 2

A - Additional Tables

We also provide Table 2.9, similar to Tables 2.1 and 2.2 but with $N = 1000$ stocks, to give interested readers a sense of the results when scaling to a larger investment universe. Since a large N may cause instability in the covariance matrix estimation, we additionally report the results obtained by combining linear shrinkage with cardinality constraints in Table 2.10.

Table 2.9: Performance Comparison ($N = 1000$, $T = 1250$).

Method	Standard deviation	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Turnover	Avg. N
Benchmark						
EW	0.2136	0.5719	0.4741	0.6159	0.1036	1000
GMV-Sample	0.1546	0.2956	0.4421	0.6317	9.0032	1000
GMV-LS	0.1074	0.2915	0.6915	0.9392	3.1164	1000
GMV-QIS	0.0967	0.2797	0.7954	1.0190	1.3127	1000
Our methods						
BNB-5	0.1424	0.4082	0.3448	0.4359	0.5199	5
BNB-10	0.1322	0.3785	0.4640	0.6049	0.6741	10
BNB-15	0.1271	0.3678	0.4615	0.5973	0.7826	15
BNB-20	0.1248	0.3411	0.4761	0.6213	0.9244	20
BNB-25	0.1225	0.3389	0.5406	0.7201	1.0355	25
BNB-30	0.1210	0.3222	0.5367	0.7154	1.1944	30
BNB-35	0.1204	0.3224	0.5913	0.7998	1.2844	35
BNB-40	0.1196	0.3234	0.5693	0.7710	1.3938	40
BNB-45	0.1189	0.3226	0.6361	0.8571	1.4656	45
BNB-50	0.1175	0.3189	0.6230	0.8376	1.5604	50

Table 2.10: Performance Comparison with Linear Shrinkage ($N = 1000$, $T = 1250$).

Method	Standard deviation	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Turnover	Avg. N
Benchmark						
EW	0.2136	0.5719	0.4741	0.6159	0.1036	1000
GMV-Sample	0.1546	0.2956	0.4421	0.6317	9.0032	1000
GMV-LS	0.1074	0.2915	0.6915	0.9392	3.1164	1000
GMV-QIS	0.0967	0.2797	0.7954	1.0190	1.3127	1000
Our methods						
BNB-LS5	0.1420	0.3830	0.5299	0.6738	0.3159	5
BNB-LS10	0.1288	0.4745	0.4187	0.5329	0.3858	10
BNB-LS15	0.1223	0.4096	0.5151	0.6620	0.4718	15
BNB-LS20	0.1181	0.4041	0.5414	0.6982	0.5604	20
BNB-LS25	0.1167	0.3606	0.5811	0.7530	0.6522	25
BNB-LS30	0.1141	0.3509	0.5930	0.7647	0.6973	30
BNB-LS35	0.1127	0.3272	0.5783	0.7498	0.7511	35
BNB-LS40	0.1119	0.2770	0.6434	0.8429	0.8068	40
BNB-LS45	0.1104	0.2723	0.6293	0.8246	0.8325	45
BNB-LS50	0.1096	0.2774	0.6114	0.7966	0.9077	50

3 Is Correlation Neglect Bad for Portfolio Diversification?

Abstract. While correlations play a central role in Markowitz portfolio selection, evidence shows that investors often neglect them, relying on simple heuristics rather than Pearson correlation. Standard theory suggests that incorporating correlations should improve performance, yet out-of-sample results frequently favor strategies that ignore them. This paper asks: Is correlation neglect always harmful, and which aspects of correlation truly matter? I propose a transformation that isolates the directional component of correlations and show that both fully ignoring and fully relying on correlation are suboptimal. Empirically, the directional component captures the most relevant information for diversification and improves portfolio performance. By distinguishing between the beneficial and irrelevant components of correlation coefficients, the paper provides a framework for constructing more robust portfolios.

Keywords: Correlation neglect; Portfolio optimization; Kendall's τ correlation; Portfolio diversification.

3.1 Introduction

Modern portfolio theory, pioneered by Markowitz (1952b), is a cornerstone of financial economics. Central to its implementation is the covariance matrix of asset returns, which dictates the benefits of diversification. The correlation structure, in particular, is critical for constructing optimal portfolios. Accurately estimating and utilizing the correlation matrix is therefore considered a fundamental challenge for investors and asset managers. Given its theoretical importance, a line of research in finance investigates "correlation neglect," a cognitive bias where investors tend to ignore or underweight correlation information when making portfolio decisions (Eyster and Weizsäcker, 2010; Kallir and Sonsino, 2009; Enke and Zimmermann, 2019; Ungeheuer and Weber, 2021). This line of work characterizes the neglect of correlation as a heuristic error and seeks to "de-bias" investors, for instance by improving the experiential presentation of correlation data to encourage its use (Laudenbach et al., 2023). The resulting prescription is straightforward: investors should devote greater attention to correlation.

In stark contrast, a line of research in finance raises serious doubts about the practical utility of sample correlation matrices. It is well-documented that the sample covariance matrix, $\widehat{\mathbf{\Sigma}}$, suffers from severe estimation errors, leading to poor portfolio optimization performance based on these estimates (DeMiguel et al., 2009b). In fact, many optimal portfolios constructed using sample covariance matrices can be easily outperformed by

simple heuristic strategies. For example, the equal-weighted $(1/N)$ portfolio, which implicitly ignores correlations, has been shown to consistently outperform optimized portfolios that rely on noisy sample estimates. Moreover, Kirby and Ostdiek (2012) demonstrate that the primary source of this instability arises not from the estimation of individual asset volatilities but rather from noisy pairwise correlation estimates. When focusing solely on volatility information, it is possible to construct portfolios that outperform the equal-weighted benchmark while avoiding extreme portfolio weights. From this perspective, it may even be rational for investors to disregard sample correlations altogether.

This tension gives rise to a fundamental paradox. On the one hand, a large body of literature urges investors to incorporate correlation information that they are often prone to neglect. On the other hand, empirical evidence suggests that relying on sample correlations can be counterproductive. How can these two perspectives be reconciled? If the full correlation matrix is too noisy to be reliable, yet ignoring it entirely discards valuable information, an important question arises: which components of correlation provide stable diversification benefits, and which are dominated by estimation error?

In this paper, we bridge this gap by proposing a simple yet powerful decomposition of the correlation coefficient into two distinct components: its magnitude (the strength of co-movement) and its rank-based directional component. We show empirically that the directional component of correlation is remarkably stable and robust, even when estimated from short time series or across large cross-sections of assets. In contrast, the magnitude of correlation is highly unstable and constitutes the primary source of estimation error that undermines portfolio optimization, particularly in low signal-to-noise environments. By isolating the stable directional information, we construct portfolios that preserve the diversification benefits of correlations while discarding their noisy elements.

To further explain these findings, we adopt both simulation-based and theoretical perspectives. We show that correlation neglect naturally arises in environments where relevant information must be estimated from noisy samples and where uncertainty about the future violates the assumptions of classical rational decision theory. Under such conditions, using the full correlation matrix often requires extremely large samples to achieve reliable asymptotic results. Counterintuitively, ignoring part of the correlation structure can sometimes improve out-of-sample performance, a phenomenon known as the less-is-more effect: the relationship between the amount of information used and the resulting accuracy forms an inverse U-shape curve.

Neglecting correlations introduces some specification error, but it can substantially reduce the variance caused by sampling noise, thereby improving robustness. Our analysis provides guidance on how investors can selectively exploit correlation information without overfitting to unreliable estimates. In particular, we show that focusing on the directional component offers a tractable and effective solution: it captures the essential information needed for diversification, requires minimal additional computation, and can be seamlessly incorporated into investment applications. For instance, by exploiting the well-known link between Kendall's τ and Pearson correlation, we can extract the directional component of correlation coefficients. This approach enables investors to make better portfolio decisions without relying on unstable correlation magnitudes.

We also analyze the statistical properties of our proposed correlation matrix through a spectral analysis of its eigenvalues. Our results show that the eigenvalues of the new matrix are more tightly centered around their true values. These findings shed light on how correlation neglect influences portfolio decisions: by effectively cleaning the extreme eigenvalues of the correlation matrix, the matrix inversion becomes more stable, which in turn produces less extreme portfolio weights and enhances the robustness of diversification.

Our research contributes to the literature in three key ways. First, we introduce a new method for portfolio construction that relies solely on the directional component of the correlation matrix. We show that this approach can be interpreted as a theoretically grounded and intuitive form of elementwise shrinkage on the sample correlation matrix. Our empirical analysis demonstrates that this method yields superior out-of-sample performance compared to the $1/N$ rule, full correlation neglect, and well-known shrinkage estimators such as Ledoit and Wolf (2004).

Second, we contribute to the literature on FinTech and financial education by providing a practical framework that helps investors incorporate robust correlation information into their investment decisions. Rather than simply encouraging reliance on the sample correlation estimator, our findings suggest that investors should focus on the stable, directional component of asset relationships. Our results help explain why investors might intuitively distrust complex correlation measures and instead rely on simpler heuristics, while also providing a practical pathway to improve their decision-making.

Finally, we provide a clearer understanding of why and how shrinkage estimators work. By decomposing the correlation matrix into its magnitude and directional components, we demonstrate that the magnitude benefits most from shrinkage, while the directional component contains valuable information that should be preserved. This leads to a more targeted and effective approach to managing estimation error in portfolio choice. Our simulation and empirical results further reveal an important mechanism: overestimated correlations are the primary drivers of extreme negative weights and unstable portfolio allocations, which in turn hurt out-of-sample performance. Interestingly, correlation neglect benefits from intentionally sacrificing some model accuracy, as it is more robust to unpredictable future events and estimation errors.

Our study connects to several strands of literature. First, by interpreting correlation neglect as an elementwise shrinkage function for correlation matrix estimation, our work naturally relates to the literature on shrinkage estimators. Ledoit and Wolf (2022a) provide a comprehensive review of the past two decades' reseach, highlighting how shrinkage methods improve the statistical properties of estimators and enhance portfolio optimization. Second, our approach is also connected to the use of thresholding operators as regularization penalties (Rothman et al., 2009; Bickel and Levina, 2008; Karoui, 2008). Unlike shrinkage methods that primarily target eigenvalues, thresholding directly regularizes individual elements of the covariance matrix. A key advantage of thresholding is that it imposes essentially no computational burden, making it attractive for problems in very high dimensions and real-time applications. We show that our directional component shares similar properties, underscoring its relevance in high-dimensional portfolio problems.

Our paper is also closely connected to the literature on Kendall's τ correlation matrices,

as we exploit the relationship between Kendall's τ and Pearson correlation to isolate the directional component of the correlation coefficient. Prior studies (Edirisinghe and Zhou, 2014; Espana et al., 2024) demonstrate that Kendall's τ improves portfolio optimization by enhancing the estimation of both eigenvalues and eigenvectors in data-poor regimes, thereby yielding better covariance matrix estimators. Related research on the empirical spectral distribution of Kendall's τ rank-based correlation matrices (Bandeira et al., 2017; Bao, 2019) further supports our approach and helps explain why focusing solely on the directional component of Pearson correlation leads to favorable statistical properties.

Furthermore, our work connects to Gerber et al. (2022), who propose the Gerber statistic as a robust co-movement measure for covariance matrix estimation in portfolio construction. Their approach recognizes that very large or extremely small movements can distort correlation estimates and therefore relies on a more robust co-movement measure in a mean-variance setting. In our paper, we formally establish the connection between rank-based correlation and traditional Pearson correlation, analyzing which component drives the improvement in estimation accuracy and portfolio performance.

Finally, our study relates to the empirical findings of Ungeheuer and Weber (2021), who show experimentally that investors understand dependence, but not necessarily in terms of Pearson correlation. Participants behave as if they apply a simple counting heuristic based on the frequency of comovement. In our framework, we formally define this heuristic using Kendall's τ and connect it to Pearson correlation. This connection provides an intuitive explanation of why correlation neglect can act as a robust method for addressing real-world diversification tasks, especially under high estimation uncertainty.

The remainder of the paper is organized as follows. Section 3.2 reviews the correlation neglect models. Section 3.3 presents the empirical results. Section 3.4 describes the simulation studies. Section 3.5 concludes.

3.2 Description of the Correlation Neglect Models Considered

In this section, we discuss various models that capture how investors deal with correlation coefficients and provide a brief introduction to each model. Our focus lies on the diversification effects that arise from different ways of measuring correlations between stocks. We use R_t to denote the N-vector of returns on the N risky assets available for investment at date t. Let $\mathbf{\Sigma}_t$ denote the corresponding $N \times N$ variance–covariance matrix of returns, with its sample estimate given by $\hat{\mathbf{\Sigma}}_t$.

To facilitate comparison across different strategies, we consider an investor who chooses the portfolio of risky assets that minimizes the variance of returns:

$$\min_{\boldsymbol{w}_t} \ \boldsymbol{w}_t^\top \mathbf{\Sigma}_t \boldsymbol{w}_t \quad \text{s.t.} \quad \boldsymbol{w}_t^\top \mathbf{1} = 1, \tag{3.1}$$

where $\boldsymbol{w}_t$ denotes the vector of portfolio weights at time t. The optimal portfolio weights are then given by:

$$\boldsymbol{w}_t = \frac{\mathbf{\Sigma}_t^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{\Sigma}_t^{-1}\mathbf{1}}. \tag{3.2}$$

Since the covariance matrix can be decomposed into volatility and correlation components, we have:

$$\mathbf{\Sigma}_t = \begin{bmatrix} \sigma_{x_1,t} & & & 0 \\ & \sigma_{x_2,t} & & \\ & & \ddots & \\ 0 & & & \sigma_{x_n,t} \end{bmatrix} \begin{bmatrix} 1 & \rho_{x_1,x_2,t} & \cdots & \rho_{x_1,x_n,t} \\ \rho_{x_1,x_2,t} & 1 & \cdots & \rho_{x_2,x_n,t} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{x_1,x_n,t} & \rho_{x_2,x_n,t} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_{x_1,t} & & & 0 \\ & \sigma_{x_2,t} & & \\ & & \ddots & \\ 0 & & & \sigma_{x_n,t} \end{bmatrix}. \tag{3.3}$$

Our analysis mainly focuses on how different treatments of the correlation component influence asset allocation and the resulting degree of diversification.

Table 3.1: List of various correlation-neglect models considered

#	Model	Abbreviation
Naïve		
1.	$1/N$ portfolio with rebalancing (benchmark strategy)	EW or $1/N$
Full Correlation Neglect		
2.	Volatility timing (Kirby and Ostdiek, 2012)	VT
Partial Correlation Neglect		
3.	Directional correlation via Kendall's τ (parametric): $\tau = \frac{2}{\pi}\arcsin(\rho)$	K's τ^2
Naïve & Partial Correlation Neglect		
4.	Linear shrinkage estimator (Ledoit and Wolf, 2004)	LW
No Correlation Neglect		
5.	Sample Pearson Correlation	SampleC

In the following, we first present out-of-sample results that illustrate how correlation neglect affects portfolio decisions and then provide detailed summary statistics on the distribution of portfolio weights across stocks. We show that the directional (rank-based) information contained in the correlation matrix is relatively stable, whereas the magnitude component is highly sensitive and often causes risk reduction to fail out-of-sample.

To quantify this effect, we proceed in several steps. First, we analyze the effects of ignoring correlation information entirely and examine how this affects portfolio weights compared to an equal-weight portfolio. Next, we consider the case of only incorporating directional (rank-based) information and investigate its impact on portfolio weight adjustments. For this part, we propose a *parametric approach* to isolate directional information from correlations by exploiting the connection between Kendall's τ and Pearson correlation. This method is computationally efficient and avoids the long calculation times required by the standard nonparametric Kendall's τ estimator.

Finally, we compare our results with well-known shrinkage estimators to show how portfolio weights are adjusted under different regularization methods. Importantly, our

goal is not to propose an optimal shrinkage estimator that minimizes estimation error. Instead, we aim to provide a method that achieves risk reduction comparable to an equal-weight portfolio while offering a more interpretable and practical framework. This allows us to offer clearer guidance on which aspects of correlation are most relevant when constructing portfolios designed to manage risk.

3.2.1 Naïve Portfolio

The naïve equal-weight ("EW" or "$1/N$") strategy allocates an equal portfolio weight of $\mathbf{w}_t = 1/N$ to each of the N risky assets. This approach requires neither estimation nor optimization and disregards all information about return correlations and volatilities. For comparison with the optimized weights in Equation (2), the $1/N$ strategy can be interpreted as ignoring all correlations in the covariance matrix $\mathbf{\Sigma}_t$ and imposing the restriction that all assets have identical volatility. Under this approach, the covariance estimator becomes

$$\widehat{\mathbf{\Sigma}}_{EW,t} = \bar{\sigma}^2 \mathbf{I},$$

where $\bar{\sigma}^2$ denotes the cross-sectional average of the sample variances. In this formulation, the investor effectively ignores both the correlation structure and cross-sectional variation in volatility. DeMiguel et al. (2009b) show that this simple rule can outperform a range of optimized portfolio strategies, particularly in the presence of estimation error.

3.2.2 Full Correlation Neglect

Kirby and Ostdiek (2012) identify the correlation estimator as a primary driver of high turnover and transaction costs in variance-based asset allocation strategies, frictions that can erode the theoretical benefits of portfolio optimization. They propose a simple yet powerful solution: setting all pairwise correlations between risky-asset returns to zero. This approach results in a diagonal covariance matrix,

$$\widehat{\mathbf{\Sigma}}_{VT,t} = \text{diag}(\hat{\sigma}^2_{1,t}, \hat{\sigma}^2_{2,t}, \ldots, \hat{\sigma}^2_{N,t}),$$

and a strategy termed *volatility timing*.

The rationale behind this method is to treat the zeroing out of the off-diagonal elements of $\widehat{\mathbf{\Sigma}}_t$ as an aggressive form of shrinkage. While this discards correlation information, it simplifies estimation by reducing the number of parameters from $N(N-1)/2$ to N. The key insight is that the significant reduction in estimation risk can outweigh the information loss, leading to more stable portfolio weights and reduced turnover. Consequently, this volatility timing strategy can outperform naïve diversification, even in the presence of significant transaction costs.

3.2.3 Partial Correlation Neglect

In this section, we introduce a correlation measure that isolates the *directional component* of stock correlation while deliberately ignoring the *magnitude* of their co-movements.

We employ Kendall's τ, a nonparametric statistic that quantifies the ordinal association between two variables by comparing their pairwise rankings. Intuitively, Kendall's τ captures how often two assets move in the same direction versus in opposite directions. This perspective allows us to view the standard Pearson correlation as comprising two components: a directional component (captured by Kendall's τ) and a magnitude component. Our analysis focuses on the directional component's effect on portfolio weights and risk reduction.

For two jointly distributed random variables (X, Y) with independent copies (X_1, Y_1) and (X_2, Y_2), the population Kendall's τ is

$$\tau = \mathbb{P}\big((X_1 - X_2)(Y_1 - Y_2) > 0\big) - \mathbb{P}\big((X_1 - X_2)(Y_1 - Y_2) < 0\big),$$

which measures the difference in probabilities of two pairs being concordant versus discordant.

Given a bivariate sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, the sample Kendall's τ is

$$\hat{\tau} = \frac{2}{n(n-1)} \sum_{i<j} \operatorname{sgn}(X_i - X_j) \operatorname{sgn}(Y_i - Y_j),$$

where

$$\operatorname{sgn}(x) = \begin{cases} +1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

A pair of observations (X_i, Y_i) and (X_j, Y_j) is concordant if the assets move in the same direction (e.g., $X_i > X_j$ and $Y_i > Y_j$), and discordant if they move in opposite directions. The statistic $\hat{\tau}$ ranges in $[-1, 1]$, with $\hat{\tau} = 1$ indicating perfect concordance, $\hat{\tau} = -1$ perfect discordance, and $\hat{\tau} = 0$ no association.

The Kendall's τ correlation matrix $\mathbf{T}$ generalizes this to multiple assets, with each entry $\mathbf{T}_{ij}$ representing the τ between the i-th and j-th variables.

Connection to Pearson Correlation

A natural question is how this directional correlation relates to the standard Pearson correlation. In financial applications, stock returns are often modeled as elliptically distributed (e.g., Normal or Student's t). In this setting, the two measures are linked by the following proposition:

PROPOSITION 3.1 (Kendall's τ and Pearson Correlation)**.** *Let* $\mathbf{X} = (X_1, X_2)$ *follow an elliptical distribution. Then Kendall's* τ *and Pearson correlation* ρ *satisfy*

$$\tau = \frac{2}{\pi} \arcsin(\rho).$$

This proposition follows directly from Theorem 2 in Lindskog et al. (2003). The result implies that focusing solely on the directional component corresponds to a *nonlinear*

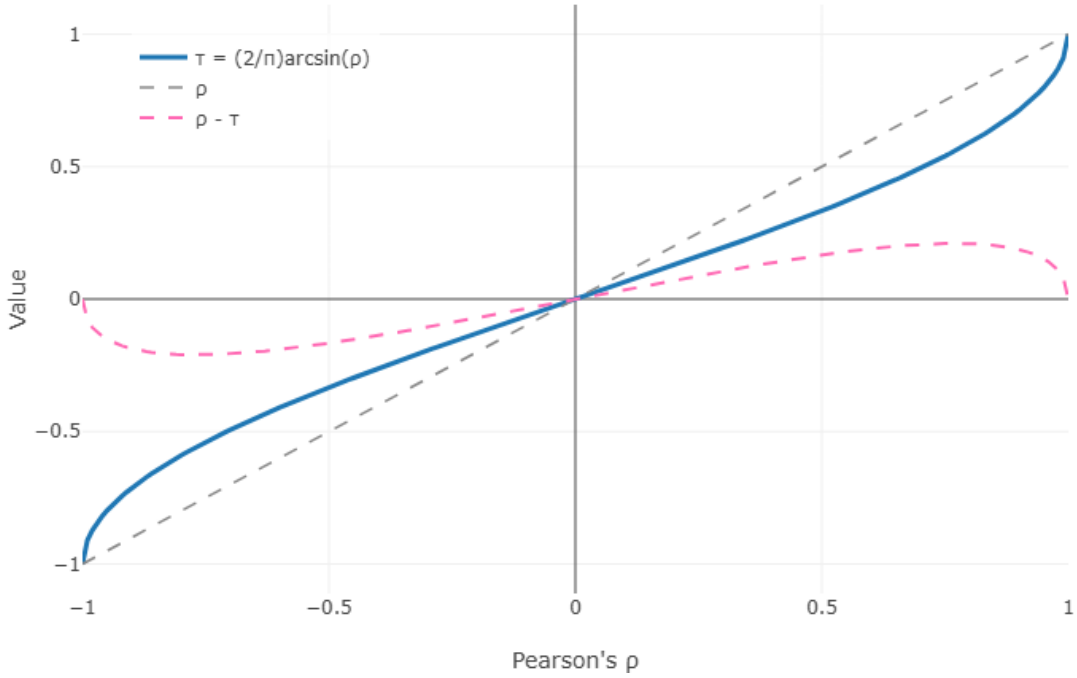


Figure 3.1: Relationship between Pearson's ρ and Kendall's τ

shrinkage of Pearson correlation toward zero. The transformation is mild for low correlations but becomes more pronounced as $|\rho|$ increases. Only when $\rho \in \{-1, 0, 1\}$ do the two measures coincide.

Figure 3.1 illustrates this effect, showing how extreme correlations are shrunk more strongly under Kendall's τ. Accordingly, we define the estimator

$$\hat{\tau}_{ij} = \frac{2}{\pi} \arcsin(\hat{\rho}_{ij}),$$

which isolates the directional information in the Pearson correlation $\hat{\rho}_{ij}$ while suppressing its magnitude.

The subsequent analysis investigates whether this transformation enhances portfolio diversification. In this case, our covariance estimator becomes

$$\hat{\mathbf{\Sigma}}_{\tau,t} = \begin{bmatrix} \hat{\sigma}_{1,t} & & & 0 \\ & \hat{\sigma}_{2,t} & & \\ & & \ddots & \\ 0 & & & \hat{\sigma}_{n,t} \end{bmatrix} \begin{bmatrix} 1 & \hat{\tau}_{12,t} & \cdots & \hat{\tau}_{1n,t} \\ \hat{\tau}_{12,t} & 1 & \cdots & \hat{\tau}_{2n,t} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\tau}_{1n,t} & \hat{\tau}_{2n,t} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \hat{\sigma}_{1,t} & & & 0 \\ & \hat{\sigma}_{2,t} & & \\ & & \ddots & \\ 0 & & & \hat{\sigma}_{n,t} \end{bmatrix}.$$

3.2.4 Naïve & Partial Correlation Neglect

Ledoit and Wolf (2004) propose a widely used linear shrinkage estimator for the covariance

matrix:

$$\widehat{\mathbf{\Sigma}}_{\mathrm{LW},t} = \lambda\bar{\sigma}^2\mathbf{I} + (1-\lambda)\widehat{\mathbf{\Sigma}}_t,$$

where $\bar{\sigma}^2$ denotes the cross-sectional average of the sample variances. This estimator shrinks the sample covariance matrix toward a scaled identity matrix, thereby reducing sensitivity to noisy correlation estimates. It can be interpreted as a linear combination of an equal-weighted portfolio (imposing identical volatilities and zero correlations) and the sample-based estimator. In this sense, it partially ignores correlation structure while simultaneously enforcing a homoskedasticity constraint, which improves estimation stability in high-dimensional settings.

3.2.5 No Correlation Neglect

To implement this model, we adopt the classic *plug-in* approach, whereby the optimization problem in Equation (2) is solved using the sample estimate of the covariance matrix, denoted by $\widehat{\Sigma}_t$. We refer to this strategy as the *sample-based minimum variance portfolio*. While this approach is commonly used to capture the correlation structure among assets, our analysis highlights that the correlation component plays a dominant role in driving high portfolio turnover, extreme weights, and large short positions, which ultimately deteriorate portfolio performance. These findings suggest that caution is warranted when relying on sample correlations in diversification-based strategies, particularly in the presence of estimation error and transaction costs, as documented in the literature.

3.3 Empirical Results

3.3.1 Methodology for Evaluating Performance

This section aims to investigate the influence of correlation neglect on real asset allocation performance across different setups and to demonstrate how our proposed method can be applied to empirical data to derive insights from observed market behavior. The analysis focuses on two main objectives: (i) evaluating the performance of the Global Minimum Variance Portfolio (GMVP) under different types of correlation neglect, and (ii) examining the impact of correlation neglect on portfolio weights.

We use daily data obtained from the Center for Research in Security Prices (CRSP), covering the period from 01/01/1995 to 12/29/2023. The empirical analysis follows a *rolling-sample* approach. Consistent with common practice, 21 consecutive trading days are treated as one "month." The out-of-sample period spans from 01/14/2000 to 12/29/2023, resulting in a total of 287 "months" (6,027 trading days). Portfolios are rebalanced monthly, with the number of shares held remaining constant within each "month," implying no transaction costs during this period. We denote the investment dates by $t = 1, \ldots, 287$.

For each combination (T, N), the investment universe is defined as the set of N stocks with complete return histories over both the most recent T trading days and the subsequent 21 trading days. At each investment date t, returns from the previous T days are used to

estimate the covariance matrix $\widehat{\boldsymbol{\Sigma}}_t$ required to implement a given strategy. The estimated parameters determine the portfolio weights $\widehat{\boldsymbol{w}}_t$ for month $t+1$. This rolling window advances monthly, adding 21 new observations and discarding the oldest 21. The outcome of this procedure is a time series of 6,027 out-of-sample daily returns generated by each portfolio strategy.

Given the out-of-sample returns, we evaluate portfolio performance using five metrics:

1. Out-of-sample annualized volatility is calculated as:

$$\hat{\sigma}_{\text{annual}} = \hat{\sigma}_{\text{daily}} \times \sqrt{252}, \tag{3.4}$$

   where the $\hat{\sigma}_{\text{daily}}$ is the daily sample standard deviation. To test whether the volatility of two strategies is statistically different, we also compute the p-value of the difference, following the approach suggested by Ledoit and Wolf (2008).

2. The out-of-sample Sharpe ratio of each portfolio is defined as the sample mean of excess returns, $\hat{\mu}_p$, divided by portfolio volatility, $\hat{\sigma}_p$:

$$\widehat{\text{SR}} = \frac{\hat{\mu}_p}{\hat{\sigma}_p}. \tag{3.5}$$

   To test whether the Sharpe ratios of two strategies are statistically different, we again compute the p-value of the difference, using the approach suggested by Ledoit and Wolf (2008).

3. Out-of-sample maximum drawdown (MDD), the largest peak-to-trough decline in portfolio value, is computed as:

$$\text{MDD} = \max_{\tau \in [0,T]} \left[\frac{\sup_{t \in [0,\tau]} P_t - P_\tau}{\sup_{t \in [0,\tau]} P_t} \right] \tag{3.6}$$

   where P_t denotes the portfolio value at time t.

4. The certainty-equivalent (CE) return represents the risk-free rate that an investor would accept instead of adopting a given portfolio strategy:

$$\widehat{\text{CE}} = \widehat{\mu}_p - \frac{\gamma}{2}\widehat{\sigma}_p^2, \tag{3.7}$$

   where γ is the coefficient of relative risk aversion. We report results for $\gamma = 4$. To test whether the CE returns from two strategies are statistically different, we also compute the p-value of the difference, using the approach suggested by DeMiguel et al. (2009b).

5. Portfolio turnover measures the trading activity required to maintain a strategy:

$$\text{Turnover} = \sum_{j=1}^{N_{t+1}} \left| \hat{w}_{j,t+1} - \hat{w}_{j,t^+} \right|, \tag{3.8}$$

where $\hat{w}_{j,t^+}$ and $\hat{w}_{j,t+1}$ denote the portfolio weights of asset j before and after rebalancing at $t+1$, respectively. To compute turnover, we consider the union of the investment universes at t and $t+1$, since the portfolio is constructed from the N largest-capitalization stocks at each rebalancing date. As a result, some firms may drop out while new ones enter, implying $N_{t+1} \geq N$ and that N_{t+1} generally varies over time.

3.3.2 Empirical Results

We first investigate a challenging but realistic scenario for asset allocation: a "short sample" setting where the number of time-series observations is only slightly larger than the number of assets ($T/N = 1.05$). This environment is notorious for producing severe estimation error in the covariance matrix.

The portfolio constructed using the raw sample covariance matrix (SampleC) performs disastrously. It exhibits extremely high volatility, large drawdowns, and deeply negative certainty-equivalent (CE) returns, implying that a risk-averse investor would pay to avoid this strategy. Moreover, its extremely high turnover indicates unstable and erratic weight reallocations from month to month. This confirms that directly inverting the sample covariance matrix in high-dimensional settings is essentially a recipe for *"error maximization"*: the optimizer chases noise rather than meaningful signal.

In contrast, strategies that neglect correlation information perform considerably better. The equally weighted (EW) portfolio provides a reasonable baseline, but the volatility-timing (VT) strategy, which uses only the diagonal elements of the covariance matrix, yields a clear improvement. By weighting assets inversely to their variances, the VT portfolio achieves significantly lower risk and a higher Sharpe ratio.

This demonstrates that even a naïve form of "correlation neglect"—ignoring off-diagonal elements entirely—can be more effective than relying on noisy correlation estimates. Statistical tests further show that partially incorporating correlation information improves portfolio risk control. In lower-dimensional settings, the gain in Sharpe ratio is modest and often not statistically significant, since reduced risk sometimes comes at the expense of lower returns. However, as the investment universe expands, the benefit of correlation information becomes more pronounced, and the improvement in the Sharpe ratio becomes statistically significant.

The most sophisticated methods deliver the strongest results. Both the Ledoit-Wolf (LW) shrinkage estimator and our proposed rank-based Kendall's τ (K's τ) strategy significantly outperform simpler heuristics. Our K's τ^2 strategy [1], which isolates the most robust directional component of correlation, consistently yields one of the highest Sharpe ratios and CE returns while maintaining a lower turnover than LW.

Statistical tests confirm that K's τ^2 portfolios extract stable and valuable information from the correlation structure. Among all methods, the closest competitor is VT, which

[1] For readers interested in how the parametric Kendall's τ compares to the nonparametric version, we report the results in the appendix. Since our main focus is on how different levels of correlation neglect influence portfolio decisions, we primarily rely on the parametric version, which can be interpreted as neglecting correlations toward zero.

Table 3.2: Performance Comparison of Different Portfolio Strategies

Method	Standard deviation	Sharpe ratio	Maximum drawdown	CE	Avg. Turnover
N=100, T=105					
K's τ^2	12.35	0.50	31.49	3.14	1.37
EW	19.32	0.36	55.45	-0.49	0.10
	(0.00)	(0.39)		(0.11)	
VT	16.48	0.43	45.71	1.64	0.17
	(0.00)	(0.59)		(0.25)	
LW	13.53	0.45	39.98	2.44	2.30
	(0.00)	(0.44)		(0.22)	
SampleC	49.24	0.28	81.32	−34.54	29.11
	(0.00)	(0.37)		(0.00)	
N=500, T=525					
K's τ^2	10.24	0.57	29.02	3.71	1.12
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.57)		(0.22)	
VT	17.39	0.54	49.74	3.40	0.09
	(0.00)	(0.88)		(0.45)	
LW	11.88	0.42	33.85	2.12	3.38
	(0.00)	(0.14)		(0.10)	
SampleC	36.03	0.17	81.26	−19.69	36.11
	(0.00)	(0.10)		(0.00)	
N=1000, T=1050					
K's τ^2	9.30	0.81	26.99	5.83	0.98
EW	21.36	0.47	57.19	1.00	0.10
	(0.00)	(0.09)		(0.09)	
VT	18.41	0.57	51.35	3.64	0.08
	(0.00)	(0.17)		(0.23)	
LW	10.84	0.64	28.04	4.61	3.42
	(0.00)	(0.12)		(0.14)	
SampleC	29.85	0.34	65.68	−7.62	33.60
	(0.00)	(0.05)		(0.01)	

Note: All values in parentheses are p-values comparing the model with the benchmark K's τ^2.

often shows larger p-values in tests of Sharpe ratio differences. This suggests that volatility alone carries strong predictive information for portfolio construction, and ignoring correlations entirely still leads to relatively good outcomes. However, volatility-based approaches sacrifice some diversification potential.

The strength of K's τ^2 lies in regularizing the correlation matrix, taming noise while preserving useful signals. By focusing on the directional component, it balances signal extraction with portfolio stability. While correlation-based methods generally produce higher turnover than the $1/N$ strategy, they significantly improve risk-adjusted performance.

Table 3.3: Performance Comparison of Different Portfolio Strategies

Method	Standard deviation	Sharpe ratio	Maximum drawdown	CE	Avg. Turnover
N=500, T=525					
K's τ^2	10.24	0.57	29.02	3.71	1.12
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.57)		(0.22)	
VT	17.39	0.54	49.74	3.40	0.09
	(0.00)	(0.88)		(0.45)	
LW	11.88	0.42	33.85	2.12	3.38
	(0.00)	(0.14)		(0.10)	
SampleC	36.03	0.17	81.26	−19.69	36.11
	(0.00)	(0.10)		(0.00)	
N=500, T=600					
K's τ^2	10.26	0.62	28.54	4.28	1.05
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.40)		(0.18)	
VT	17.42	0.55	49.82	3.44	0.09
	(0.00)	(0.64)		(0.38)	
LW	11.72	0.50	31.29	3.16	3.21
	(0.00)	(0.26)		(0.18)	
SampleC	19.58	0.23	62.67	−3.25	12.35
	(0.00)	(0.05)		(0.01)	
N=500, T=1250					
K's τ^2	10.54	0.67	29.46	4.85	0.71
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.26)		(0.13)	
VT	17.60	0.55	50.72	3.40	0.08
	(0.00)	(0.43)		(0.29)	
LW	11.11	0.50	29.02	3.09	1.89
	(0.00)	(0.04)		(0.03)	
SampleC	11.70	0.44	30.83	2.39	2.51
	(0.00)	(0.03)		(0.02)	

All values in parentheses are p-values comparing the model with the benchmark K's τ^2.

Next, we examine whether simply increasing the length of the estimation window can overcome estimation challenges. For a fixed portfolio size of $N = 500$, we expand the sample from $T = 525$ to $T = 1250$.

The findings in Table 3.3 reveal a nuanced picture:

- As T increases, the performance of the SampleC strategy *improves dramatically*. By $T = 1250$, its volatility and Sharpe ratio become respectable.
- However, consistent with the findings of DeMiguel et al. (2009b), SampleC still fails to outperform simpler, more robust strategies. Achieving reliable results with SampleC requires unrealistically long time series.

The key insight from Table 3.3 is the remarkable stability of strategies related to correlation-neglect approaches. The performance of VT and especially K's τ^2 remains exceptionally consistent across different time-series lengths. While longer samples slightly reduce turnover for K's τ^2, the overall risk-return profile remains stable.

Interestingly, VT and K's τ^2 yield statistically similar Sharpe ratios and CE returns, but K's τ^2 achieves lower volatility. This result arises because volatility alone provides strong signals about the risk-return trade-off, but correlation information further enhances volatility control. VT portfolios exhibit slightly higher volatility but achieve comparable returns, while K's τ^2 extracts additional diversification benefits.

Even though SampleC improves with longer data windows, it still tends to overestimate correlations and overweight certain stocks, which ultimately hurts returns. Counterintuitively, ignoring noisy correlation information often produces better portfolios than using SampleC directly.

This analysis highlights several crucial points:

- When sample sizes are small, naïve inversion of the sample covariance matrix leads to unstable and poorly performing portfolios.
- Simple correlation-neglect strategies, such as VT, already deliver significant benefits.
- Among advanced methods, K's τ^2 stands out by isolating a stable directional signal, yielding strong performance with lower turnover and greater robustness.

For investors who are uncertain whether they have "enough" data, focusing on directional correlation provides a reliable and effective approach to portfolio construction. We have also conducted the empirical analysis using the full Markowitz setup with a predictive signal, and the patterns are very similar. Readers interested in these results may refer to Appendix 3.5.

3.3.3 The Effect of Correlation Neglect on Portfolio Weights

In this section, we conduct a detailed analysis of the summary statistics for the portfolio weights generated by various allocation strategies. Building on the preceding finding that neglecting the correlation component can mitigate estimation error and improve

out-of-sample performance, we now seek to understand the mechanism through which this improvement is achieved. Specifically, we investigate how different treatments of the covariance matrix influence the resulting portfolio weights. The central hypothesis is that estimation errors, particularly in the off-diagonal elements (correlations) of the sample covariance matrix, induce extreme long and short positions, which in turn degrade portfolio performance. To characterize the portfolios, we compute five key summary statistics:

1. **Average Minimum Weight**: The mean of the lowest portfolio weight assigned to any single asset across all rebalancing periods.

2. **Average Maximum Weight**: The mean of the highest portfolio weight assigned to any single asset across all rebalancing periods.

3. **Average Percentage of Positive Weights**: The mean proportion of assets held in long positions across all rebalancing periods. We report this measure because Jagannathan and Ma (2003) show that imposing no-short-sale constraints stabilizes portfolio weights. We examine whether correlation neglect operates through a similar channel-by reducing short positions and thereby stabilizing portfolio weights.

4. **Average Concentration**: The mean Herfindahl index, a measure of weight concentration defined as $H = \sum_{i=1}^{N} w_i^2$. A higher index indicates a more concentrated portfolio.

5. **Average Short Position Size**: The mean absolute value of all negative weights, representing the average magnitude of short-selling across rebalancing periods.

These statistics are averaged across the 287 monthly rebalancing dates in our out-of-sample period. They collectively provide a comprehensive picture of the portfolio's sparsity, diversification, and reliance on leverage through short-selling.

Table 3.4 reports these results for the short-sample case where the time-series length T and the cross-sectional dimension N satisfy $T/N = 1.05$. This data-limited scenario is common in practice and magnifies the impact of estimation error. The results, presented in Table 3.4, are striking. The findings in Table 3.4 reveal stark differences in the characteristics of the portfolios constructed using varying levels of correlation neglect.

The portfolio based on the sample covariance matrix (SampleC), which directly inverts the sample covariance matrix, exhibits severe instability. The average minimum and maximum weights are extreme, often exceeding 100% in magnitude. This indicates that the portfolio takes massive, highly leveraged positions in a few assets. This is confirmed by the astronomical Herfindahl Index (12.27), signifying extreme concentration, and a total short position that is over 10 times the portfolio's net value. Concurrently, the average short position size is substantial, highlighting an aggressive and risky allocation strategy driven by spurious correlations in the data. These extreme weights are a classic symptom of estimation error in an ill-conditioned covariance matrix, where small changes in input parameters lead to dramatic shifts in the optimal portfolio. This behavior is

Table 3.4: Summary Statistics of Weights of Different Portfolio Strategies

	Avg Min	Avg Max	Avg. Pos. w	Avg. Conc.	Avg. Size of Short
N=100, T=105					
K's τ^2	−0.0562	0.1703	57.88	0.1339	0.6941
EW	0.0100	0.0100	100.00	0.0100	0.0000
VT	0.0016	0.0327	100.00	0.0136	0.0000
LW	−0.0970	0.1413	59.31	0.2085	1.2100
SampleC	−1.1021	1.1661	51.41	12.2742	10.2765
N=500, T=525					
K's τ^2	−0.0283	0.0729	52.64	0.0662	1.4705
EW	0.0020	0.0020	100.00	0.0020	0.0000
VT	0.0002	0.0073	100.00	0.0027	0.0000
LW	−0.0567	0.0748	53.37	0.1923	3.2240
SampleC	−0.4595	0.4724	50.79	4.5141	15.9578
N=1000, T=1050					
K's τ^2	−0.0193	0.0559	51.24	0.0462	1.7550
EW	0.0010	0.0010	100.00	0.0010	0.0000
VT	0.0001	0.0042	100.00	0.0014	0.0000
LW	−0.0378	0.0557	52.58	0.1367	3.8644
SampleC	−0.2571	0.2619	50.65	2.3843	16.4928

a classic symptom of "error maximization," where the optimizer aggressively exploits spurious correlations found in noisy data.

In stark contrast, the volatility-timing (VT) strategy, which completely ignores correlation information and weights assets based solely on their volatility, produces highly stable weights. The weights are constrained, with a complete absence of short positions. This approach is analogous to the equally weighted (EW) portfolio in its simplicity and robustness. By ignoring the noisy correlation structure, the VT strategy avoids the optimization error that plagues the sample-based approach, leading to the superior out-of-sample performance documented in Table 3.2.

The strategy based on Kendall's τ^2, which utilizes only the directional component of correlation, offers a compelling intermediate solution. This method implicitly restricts the magnitude of short positions while still permitting significant long positions. The asymmetric effect on long and short weights is evident from the summary statistics. The concentration ratio, as measured by the Herfindahl index, is significantly reduced compared to the sample covariance strategy. This behavior can be interpreted as a form of implicit regularization or shrinkage. By focusing on the more robust, rank-based measure of co-movement, the model filters out much of the estimation noise associated with linear

correlation, thereby preventing the optimization from assigning extreme negative weights.

This shrinkage-like effect is also observed in the portfolio constructed using the Ledoit-Wolf (LW) shrinkage estimator. The LW method explicitly regularizes the covariance matrix by shrinking the sample estimates towards a more structured target. This process effectively dampens extreme weights, reduces short-selling intensity, and increases portfolio diversification, as evidenced by the lower Herfindahl index. Both the Kendall's τ^2 and LW approaches demonstrate that moderating the influence of estimated correlations is critical for achieving well-diversified and robust portfolios.

Furthermore, our analysis shows that as the number of assets in the portfolio increases, the limitations of the sample covariance matrix become more pronounced. Although larger portfolios based on the sample matrix exhibit slightly less extreme maximum weights, they tend to concentrate leverage into a single substantial short position. This indicates that as portfolio dimensionality grows, the risk of estimation errors leading to dangerously concentrated short positions also increases.

This analysis reveals a key channel through which estimation error harms portfolio performance: the overestimation of correlation coefficients in finite samples induces extreme negative weights, leading to poorly diversified and unstable portfolios. Strategies that neglect correlation (like VT) or employ robust estimators that implicitly or explicitly shrink the correlation structure (like Kendall's τ^2 and Ledoit-Wolf) effectively mitigate this problem. This finding is consistent with the work of Jagannathan and Ma (2003), who demonstrate that imposing no-short-sale constraints can be an effective remedy for estimation error in practice. Our results suggest that correlation neglect acts similarly to such a constraint, primarily by preventing the large, erroneous short positions that arise from noisy correlation estimates. Ultimately, we find that the directional component of correlation is a more stable and reliable input for portfolio decisions, especially when data is limited. By focusing on this robust feature, investors can construct better-diversified portfolios and enhance out-of-sample performance.

We next investigate whether simply having more data can solve the problem. In Table 3.5, we fix the number of assets at $N = 500$ and increase the time-series length from $T = 525$ ($T/N = 1$) to $T = 1250$ ($T/N = 2.5$).

As expected, increasing the amount of data helps the SampleC strategy. The average extreme weights decrease, and the concentration index drops from 4.51 to a more moderate 0.36. However, even with 2.5 times more data points than assets, the portfolio remains significantly more concentrated and leveraged compared to those constructed using alternative correlation neglect methods. This highlights that simply relying on "brute-force" longer samples is often insufficient to resolve the underlying estimation issues.

The most important result from this table is the remarkable stability of the K's τ^2 strategies. Across all data lengths ($T = 525, 600, 1250$), the summary statistics for the K's τ^2 portfolios remain almost unchanged. This provides strong evidence supporting our central thesis: the directional component of correlation is a stable signal that is largely insensitive to the noise induced by limited data. By isolating this signal, we develop a portfolio construction method that remains robust in both high-dimensional and short-sample environments.

Table 3.5: Summary Statistics of Weights of Different Portfolio Strategies

	Avg. Min	Avg. Max	Avg. Pos. w	Avg. Conc.	Avg. Size of Short
N=500, T=525					
K's τ^2	−0.0283	0.0729	52.64	0.0662	1.4705
EW	0.0020	0.0020	100.00	0.0020	0.0000
VT	0.0002	0.0073	100.00	0.0027	0.0000
LW	−0.0567	0.0748	53.37	0.1923	3.2240
SampleC	−0.4595	0.4724	50.79	4.5141	15.9578
N=500, T=600					
K's τ^2	−0.0287	0.0738	52.74	0.0659	1.4610
EW	0.0020	0.0020	100.00	0.0020	0.0000
VT	0.0002	0.0072	100.00	0.0027	0.0000
LW	−0.0586	0.0792	53.33	0.1978	3.2592
SampleC	−0.2398	0.2528	51.21	1.1949	8.0010
N=500, T=1250					
K's τ^2	−0.0304	0.0739	53.20	0.0621	1.3651
EW	0.0020	0.0020	100.00	0.0020	0.0000
VT	0.0002	0.0068	100.00	0.0027	0.0000
LW	−0.0699	0.0923	53.19	0.1761	2.8480
SampleC	−0.1946	0.2043	52.57	0.3576	3.5903

Our analysis of portfolio weights reveals the precise channel through which estimation error harms performance: noisy, overestimated sample correlations lead to extreme and heavily leveraged negative weights. This finding is consistent with Jagannathan and Ma (2003), who show that no-short-sale constraints can improve performance by mitigating the impact of estimation error.

Strategies that neglect correlation act as a blunt but effective tool, similar to a no-short-sale constraint. However, our proposed method, based on Kendall's τ, offers a more nuanced solution. By retaining the stable directional information in correlations while discarding the noisy magnitude component, it effectively regularizes the portfolio, preventing extreme weights and enhancing diversification. This approach provides a robust and reliable method for portfolio construction, especially when data is limited. We also check the mechanisim under the full Markowitz setup with a predictive signal and obtain similar patterns; see Appendix 3.5 for details.

3.4 Simulation Studies

The results in Section 3.3 demonstrate that correlation neglect influences diversification in real data. To further investigate this mechanism, we employ simulated data to analyze how the performance of the strategies considered in our empirical analysis varies with the number of assets (N) and the length of the estimation window (T). The main advantage of using simulated data is that their economic and statistical properties are fully specified. Specifically, we generate returns from a simple single-factor model, assuming that they are independently and identically distributed over time and follow a normal distribution. Since the normal distribution is elliptical, this framework provides a clearer understanding of how correlation neglect affects portfolio diversification. Moreover, it ensures that the results are not confounded by empirical irregularities, such as small-firm effects, calendar effects, momentum, mean reversion, or the fat tails typically observed in real financial data.

Our approach for simulating returns, as well as our choice of parameter values, follows (Craig MacKinlay and Pástor, 2000; DeMiguel et al., 2009b). We assume that the market consists of a risk-free asset and N risky assets, which are driven by K underlying factors. The excess returns of the remaining $N - K$ risky assets are generated according to the factor model: $R_{a,t} = \alpha + BR_{b,t} + \epsilon_t$, where $R_{a,t}$ is the vector of excess asset returns, α is the vector of mispricing coefficients, B is the factor loadings matrix, $R_{b,t}$ is the vector of excess returns on the factor portfolios, $R_b \sim N(\mu_b, \Sigma_b)$, and ϵ_t is the vector of idiosyncratic noise, $\epsilon \sim N(0, \Sigma_\epsilon)$, which is assumed to be independent of the factor portfolios.

Following the setup of DeMiguel et al. (2009b), we assume that the risk-free rate follows a normal distribution with an annual mean of 2% and a standard deviation of 2%. We further assume that there is only one factor ($K = 1$), whose annual excess return has a mean of 8% and a standard deviation of 16%. The mispricing term α is set to zero, and the factor loadings B are evenly distributed between 0.5 and 1.5. The variance-covariance matrix of the idiosyncratic noise, Σ_ϵ, is assumed to be diagonal, with elements drawn from a uniform distribution on $[0.10, 0.30]$, resulting in a cross-sectional average annual idiosyncratic volatility of 20%. We also report results for different volatility levels to examine their impact on portfolio weights.

We consider scenarios with the number of assets set to $N \in \{100, 500, 1000\}$ and estimation window lengths $T \in \{505, 1250, 2500\}$, corresponding to short, medium, and long data setups, respectively. For each configuration, Monte Carlo sampling is used to generate daily return data matching the out-of-sample length in the empirical section.

From the simulation results, we observe that the sample correlation coefficient produces estimates close to the true diversification outcome only when idiosyncratic volatility is low and sufficiently long time-series samples are available-roughly ten times the cross-sectional dimension. However, once idiosyncratic volatility increases relative to factor variance (as is typically the case in financial markets with a low signal-to-noise ratio), the performance of the sample correlation coefficient deteriorates significantly. In such settings, much longer data histories are required to obtain reliable results. When the available sample

Table 3.6: Standard deviation for simulated data.

	$N = 50$			$N = 100$			$N = 500$		
Strategy	$T = 505$	$T = 1250$	$T = 2500$	$T = 505$	$T = 1250$	$T = 2500$	$T = 505$	$T = 1250$	$T = 2500$
Noise Level					20%				
K's τ^2	8.12	8.04	8.02	6.55	6.42	6.37	3.50	3.29	3.19
True	7.03	7.03	7.03	5.62	5.62	5.62	2.72	2.72	2.72
LW	7.61	7.28	7.15	6.52	6.00	5.80	5.27	3.61	3.13
SampleC	7.39	7.19	7.12	6.30	5.86	5.73	29.97	3.51	3.03
Noise Level					50%				
K's τ^2	14.70	14.53	14.45	13.27	12.93	12.79	8.76	8.08	7.79
True	13.33	13.33	13.33	11.70	11.70	11.70	7.01	7.01	7.01
LW	15.40	14.79	14.32	13.75	13.11	12.70	9.72	8.91	8.15
SampleC	14.05	13.65	13.51	13.15	12.25	11.94	87.61	9.07	7.84
Noise Level					80%				
K's τ^2	16.25	16.04	15.96	15.74	15.31	15.14	12.67	11.66	11.22
True	14.94	14.94	14.94	13.98	13.98	13.98	10.02	10.02	10.02
LW	19.09	18.68	18.25	17.05	16.61	16.30	12.65	11.89	11.48
SampleC	15.78	15.31	15.14	15.72	14.62	14.27	360.40	12.95	11.20

is short or the cross-sectional dimension is large, the estimates become extremely noisy, leading to unstable and highly concentrated portfolio allocations.

By contrast, the directional component of correlation remains remarkably stable across different sample lengths and cross-sectional dimensions. Although it necessarily discards part of the correlation structure and thereby introduces specification error, it consistently yields results close to the true benchmark. As illustrated in Figures 3.13 and 3.14 (provided in the appendix for space considerations), this stability translates into more diversified portfolio weights and substantially fewer short positions. Conceptually, this reflects a trade-off between specification error and sampling error: by ignoring part of the correlation information, we incur specification error but benefit from more stable portfolios that are robust to sample length and dimensionality.

A comparison with the Ledoit-Wolf (LW)[2] shrinkage estimator reveals an interesting distinction. Shrinkage methods stabilize the covariance estimator but shrink both variances and correlations, which leads to lower performance relative to the directional component. This finding reinforces our central result: the correlation matrix is the primary source of sampling error and extreme weights, more so than variance estimates. Correlation neglect therefore acts as an implicit shrinkage mechanism. By down-weighting correlations, it regularizes extreme portfolio weights and reduces reliance on short positions, ultimately

[2]If one follows Ledoit and Wolf (2003) and shrinks toward a one-factor market model-where the factor is defined as the cross-sectional average return-LW consistently outperforms the sample covariance in any scenario. Note that in our simulations some LW results appear weaker than expected. This mainly arises from the simulation design: when we raise idiosyncratic noise to very high levels, we essentially add a vector of noise independent of the factor portfolios. In this case, if we use the single-parameter shrinkage method of Ledoit and Wolf (2004), which pushes the covariance matrix toward equal variances. This misspecifies volatility (which is informative), so performance suffers. This is consistent with our earlier discussion of why volatility-timing strategies outperform equal-weighting.

producing lower-risk portfolios, consistent with the findings of Jagannathan and Ma (2003).

Finally, to further illustrate the role of the estimation window size, we provide additional plots in the appendix showing the portfolio risk of different methods as T increases. The results reveal that both K's τ^2 and LW remain relatively stable across different sample sizes, whereas SampleC performs poorly in small samples and only begins to outperform K's τ^2 when the sample size reaches approximately five to ten times the cross-sectional dimension. This finding confirms that extremely long histories are required for SampleC to achieve its true performance potential, consistent with insights from previous literature (DeMiguel et al., 2009b).

Taken together, the simulation study offers a deeper understanding of correlation neglect. It can be interpreted as a form of shrinkage that, while introducing specification error, enhances portfolio performance by mitigating extreme weights and reducing short exposures. The exceptional stability of the directional component provides a practical guideline for practitioners seeking to incorporate correlation information under limited data availability or in noisy market environments.

3.4.1 Empirical Spectral Distribution

Modern Portfolio Theory (MPT), pioneered by Markowitz, provides a theoretically elegant framework for constructing optimal portfolios based on estimated means, variances, and correlations of asset returns. However, its application in practice has been limited by a fundamental challenge: estimating the correlation matrix accurately in high-dimensional, data-poor environments.

As our previous simulation results demonstrate, obtaining reliable correlation estimates typically requires the sample size T to be at least five to ten times larger than the number of assets N. In real-world applications, where the cross-sectional dimension is large and time-series data are limited, such conditions rarely hold. Consequently, the sample correlation matrix often suffers from severe estimation noise, leading to unstable portfolio weights and poor out-of-sample performance.

To better understand the effects of data limitations on correlation estimation, researchers have increasingly relied on Random Matrix Theory (RMT; Marvcenko and Pastur (1967)), which provides powerful insights into the spectral properties of sample covariance and correlation matrices. RMT characterizes the limiting spectral distribution of eigenvalues under the null hypothesis of independent and identically distributed (i.i.d.) returns. This theoretical benchmark allows us to distinguish between eigenvalue components dominated by signal and those dominated by noise, thereby guiding the development of methods for regularizing or denoising empirical correlation matrices.

Building on these insights, a rich body of literature has emerged to improve correlation estimation in high-dimensional settings. Techniques such as eigenvalue cleaning, shrinkage, and factor-based approaches have been proposed and extensively studied. For comprehensive reviews, see Bun et al. (2016) and Bouchaud and Potters (2009). In particular, many shrinkage estimators, including the widely used Ledoit-Wolf methods, are grounded

in RMT principles and provide theoretically justified improvements over the naïve sample estimator (Ledoit and Wolf (2022a)).

In this section, we investigate how our proposed method affects the spectral distribution of eigenvalues of the correlation matrix. By comparing our approach with standard shrinkage and eigenvalue-cleaning techniques, we highlight its effectiveness in reducing estimation noise while preserving essential dependency structures among assets. This analysis provides a deeper understanding of the mechanism through which our method improves portfolio diversification and risk management in high-dimensional environments.

The Marvcenko–Pastur Law and Eigenvalue Distributions

The resulting cross-sectional distribution of the sample eigenvalues is known as the Marvcenko–Pastur (MP) law (Marvcenko and Pastur, 1967), which provides a theoretical benchmark for understanding the behavior of sample eigenvalues in high-dimensional settings. Consider a data matrix $\mathbf{X} \in \mathbb{R}^{T\times N}$, where returns are i.i.d. with zero mean and variance 1. Define the sample correlation matrix as:$\mathbf{S} = \frac{1}{T}\mathbf{X}^\top\mathbf{X}$. When both N and T grow to infinity with their ratio converging to $\gamma = N/T$, the empirical spectral distribution of $\mathbf{S}$ converges almost surely to the MP distribution with density:

$$f_{\mathrm{MP}}(\lambda) = \begin{cases} \dfrac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\gamma\lambda}, & \lambda \in [\lambda_-, \lambda_+], \\ 0, & \text{otherwise}, \end{cases}$$

where the lower and upper bounds of the support are: $\lambda_\pm = \left(1 \pm \sqrt{\gamma}\right)^2$.

The MP law therefore characterizes the eigenvalue spectrum of the sample correlation matrix under the null hypothesis of no correlation. When γ is large (i.e., T is small relative to N), the eigenvalues become highly dispersed, and many accumulate near zero, leading to numerical instability in matrix inversion. Conversely, when $\gamma \to 0$ (data-rich environments), the eigenvalues concentrate tightly around 1, recovering the population correlation structure.

This result provides the theoretical foundation for modern techniques in correlation estimation and eigenvalue cleaning. In our context, we leverage the MP framework to understand how focusing on the directional component of correlations affects the eigenvalue distribution and, consequently, the stability of portfolio weights.

PROPOSITION 3.2 (Marvcenko–Pastur Law for the Directional Componentof Pearson Correlation)**.** *Let $\mathbf{X} \in \mathbb{R}^{T\times N}$ be a data matrix with i.i.d. entries of zero mean and variance* 1*, and assume $\mathbf{X}$ follows an elliptical distribution. Define the sample Pearson correlation matrix $\hat{\rho}_{ij}$ and construct the directional correlation matrix $\mathbf{D}$ as:*

$$D_{ij} = \frac{2}{\pi}\arcsin(\hat{\rho}_{ij}).$$

Then, as $N, T \to \infty$ with $N/T \to \gamma > 0$, the empirical spectral distribution of $\mathbf{D}$ converges in probability to:

$$\frac{2}{3}Y_\gamma + \frac{1}{3},$$

where Y_γ follows the standard Marvcenko–Pastur distribution with parameter γ as defined above.

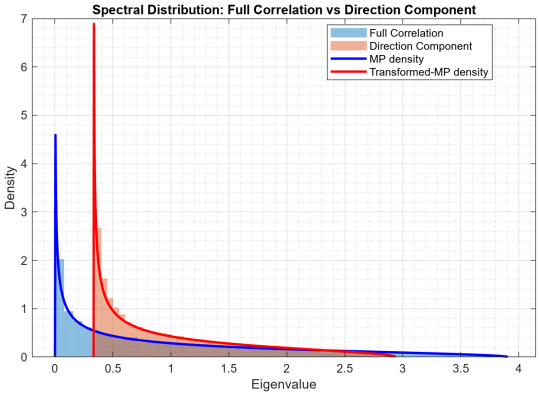
Proof. As $T \to \infty$, we have $\hat{\rho}_{ij} \to \rho_{ij}$. From Proposition **??**, under the elliptical distribution assumption, we have the relationship between Kendall's τ and Pearson's ρ:

$$\tau_{ij} = \frac{2}{\pi} \arcsin(\rho_{ij}).$$

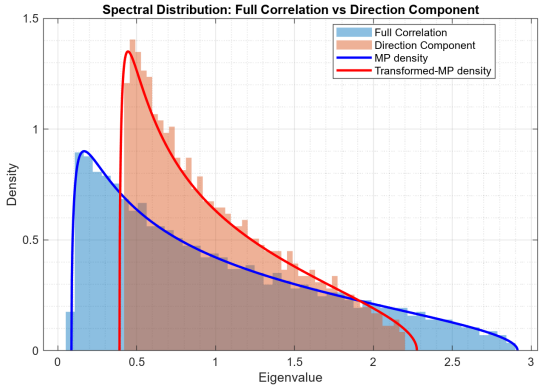
Since $\frac{2}{\pi} \arcsin(1) = 1$, the diagonal elements of $\mathbf{D}$ equal one, and the off-diagonal elements correspond to the Kendall's τ correlation matrix. By Bandeira et al. (2017), the empirical spectral distribution of the Kendall's τ matrix converges to:

$$\frac{2}{3} Y_\gamma + \frac{1}{3},$$

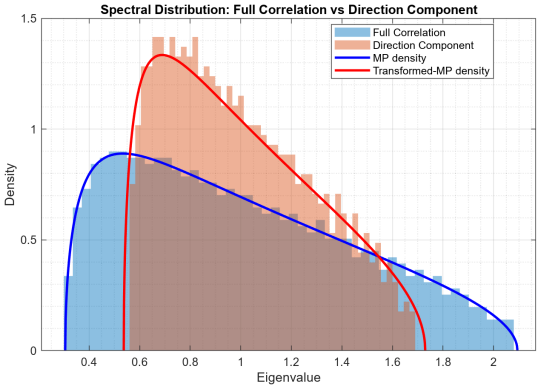
where Y_γ follows the standard Marvcenko–Pastur distribution for the Pearson correlation matrix $\mathbf{X}$. □



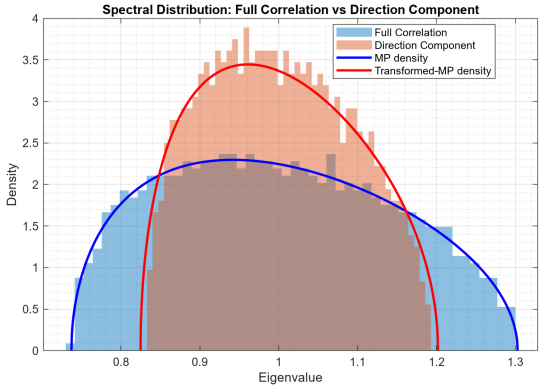
A. $N = 1000$, $T = 1050$, $\gamma \approx 0.95$



B. $N = 1000$, $T = 2000$, $\gamma = 0.5$



C. $N = 1000$, $T = 5000$, $\gamma = 0.2$



D. $N = 1000$, $T = 50000$, $\gamma = 0.02$

Figure 3.2: Comparison of empirical eigenvalue distributions and theoretical MP densities under different values of γ. The black solid line is the original MP density, and the red dashed line is the affine-transformed MP density.

From Fig. 3.2, we observe that in data-poor settings (small T relative to N), the eigenvalues of the sample correlation matrix are highly concentrated near zero. Since portfolio weights are proportional to the inverse of the covariance matrix, this concentration induces extremely unstable and extreme weights.

However, when we focus only on the directional component of the correlation—obtained via the arcsin transformation of Kendall's τ—the eigenvalue distribution becomes significantly shrunk toward 1. This stabilization of the spectrum leads to much more stable portfolio weights, even when data are limited.

Our simulation further illustrates this mechanism by fixing $N = 100$ and increasing T. When T is extremely large, the eigenvalues of the Pearson correlation matrix naturally become well distributed around 1, and the directional component yields a spectrum even closer to 1. This explains the earlier finding: in data-rich environments, Kendall's τ may introduce a small model misspecification error relative to Pearson correlation, but its impact on portfolio performance is negligible because the estimation error disappears when sufficient data are available.

In all other cases, however, focusing only on the directional component offers a practical and robust solution. This section identifies the channel through which extreme portfolio weights arise—namely, the distortion of the eigenvalue distribution in high-dimensional, data-poor setups. From this perspective, shrinking correlations is effectively equivalent to shrinking the eigenvalue distribution, which aligns with other eigenvalue-cleaning techniques in the literature. Therefore, applying the arcsin transformation to isolate the directional component of correlation provides a theoretically justified and numerically stable approach to covariance estimation, avoiding the need for extremely long time-series data.

3.5 Conclusion

The financial literature presents conflicting views on the role of correlation in portfolio choice. On the one hand, behavioral research identifies "correlation neglect" as a cognitive bias, suggesting that investors should be encouraged to incorporate correlation information into their decisions. On the other hand, a large body of empirical evidence demonstrates that portfolios constructed using sample correlation estimates often perform poorly out of sample, implying that ignoring these correlations-as simple heuristics do-can be a rational response to estimation error.

This paper reconciles this tension by showing that the sample correlation coefficient is not a monolithic signal. We propose a transformation that isolates the directional component of correlations and show that both ignoring correlations entirely and relying on them fully are suboptimal. Our central finding is that the directional component is remarkably stable and robust, even in high-dimensional settings with limited data. In contrast, the magnitude is the primary source of estimation error, which destabilizes portfolio weights and deteriorates out-of-sample performance.

We systematically analyze how correlation neglect influences portfolio diversification. Using simulation studies, we demonstrate that ignoring the magnitude of correlations

introduces some specification error but substantially reduces the variance caused by sampling noise, thereby improving robustness. A spectral analysis of the correlation matrix further reveals the mechanism: by effectively shrinking extreme eigenvalues, correlation neglect stabilizes the matrix inversion required for portfolio optimization, which in turn reduces extreme portfolio weights. Our empirical analysis confirms that this mechanism also operates in real financial data, where our proposed method consistently delivers improved diversification and risk-adjusted performance.

The implications of our findings are threefold. First, for financial econometrics, our approach offers a new and intuitive form of elementwise shrinkage on the correlation matrix, providing a theoretical foundation for discarding noisy components while retaining stable, valuable information. Second, for behavioral finance, our work refines the debate on correlation neglect: investors' skepticism toward sample correlation estimates may not reflect a cognitive bias but rather a rational response to excessive noise. The prescriptive takeaway is not to indiscriminately "use correlation" but instead to focus on its stable directional component. Third, for practitioners, our proposed method is simple to implement and provides a practical pathway to harness the benefits of diversification without being misled by unstable correlation estimates.

In conclusion, the question is not whether to use correlation, but how. By disentangling the stable signal from the volatile noise within the correlation matrix, this paper provides both a theoretically grounded and practically effective framework for building more robust, well-diversified, and stable portfolios.

Appendix to Chapter 3

A. Markowitz Portfolio with Momentum

In this section, we turn our attention to the Markowitz portfolio with a return-predictive signal. We report the results from Section 3.3 under the full Markowitz portfolio setup. Specifically, we incorporate the momentum factor of Jegadeesh and Titman (1993). For a given investment period t and a given stock, momentum is defined as the geometric average of the previous 252 daily returns, excluding the most recent 21 returns. Collecting the individual momentum measures of all N stocks in the investment universe yields the return-predictive signal $\boldsymbol{m}$.

In the absence of short-sale constraints, the optimization problem for the Markowitz portfolio with a momentum signal is formulated as

$$\begin{aligned} \min_{\boldsymbol{w}_t} \quad & \boldsymbol{w}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{w}_t \\ \text{s.t.} \quad & \boldsymbol{w}_t^\top \boldsymbol{m} = b, \\ & \boldsymbol{w}_t^\top \mathbf{1} = 1. \end{aligned}$$

Here, b denotes the target expected return. We set b equal to the arithmetic average return of the equally weighted portfolio composed of the top-quintile stocks ranked by momentum $\boldsymbol{m}$. This same value of b is used as the target expected return for the

subsequent portfolio analyses. The problem admits a closed-form analytical solution $\boldsymbol{w}_t = \lambda_1 \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu} + \lambda_2 \boldsymbol{\Sigma}_t^{-1}\mathbf{1}$, where $\lambda_1 = (Cb - B)/D$ and $\lambda_2 = (A - Bb)/D$, with $A = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}$, $B = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}_t^{-1}\mathbf{1}$, $C = \mathbf{1}^\top \boldsymbol{\Sigma}_t^{-1}\mathbf{1}$, and $D = AC - B^2$.

Table 3.7: Performance Comparison of Different Portfolio Strategies (Mean-Variance)

Method	Standard deviation	Sharpe ratio	Maximum drawdown	CE	Avg. Turnover
N=100, T=105					
K's τ^2	14.75	0.53	42.63	3.54	1.95
EW	19.32	0.36	55.45	-0.49	0.10
	(0.00)	(0.32)		(0.10)	
VT	19.15	0.33	51.41	-1.04	0.69
	(0.00)	(0.12)		(0.03)	
LW	16.15	0.51	50.69	3.08	2.95
	(0.00)	(0.74)		(0.33)	
SampleC	51.44	0.23	87.33	−41.23	31.23
	(0.00)	(0.19)		(0.00)	
N=500, T=525					
K's τ^2	11.79	0.54	31.34	3.58	1.63
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.70)		(0.24)	
VT	18.94	0.47	43.33	1.71	0.63
	(0.00)	(0.65)		(0.24)	
LW	13.62	0.40	33.71	1.73	4.00
	(0.00)	(0.16)		(0.08)	
SampleC	41.38	0.16	80.91	−27.42	41.55
	(0.00)	(0.11)		(0.00)	
N=1000, T=1050					
K's τ^2	10.67	0.70	27.31	5.19	1.50
EW	21.36	0.47	57.19	1.00	0.10
	(0.00)	(0.29)		(0.13)	
VT	19.54	0.48	43.75	1.74	0.62
	(0.00)	(0.17)		(0.12)	
LW	12.17	0.51	26.80	3.22	4.02
	(0.00)	(0.06)		(0.05)	
SampleC	35.58	0.19	82.41	−18.44	40.19
	(0.00)	(0.03)		(0.00)	

Note: All values in parentheses are p-values comparing the model with the benchmark K's τ^2.

Table 3.8: Performance Comparison of Different Portfolio Strategies (Mean-Variance)

Method	Standard deviation	Sharpe ratio	Maximum drawdown	CE	Avg. Turnover
N=500, T=525					
K's τ^2	11.79	0.54	31.34	3.58	1.63
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.70)		(0.24)	
VT	18.94	0.47	43.33	1.71	0.63
	(0.00)	(0.65)		(0.24)	
LW	13.62	0.40	33.71	1.73	4.00
	(0.00)	(0.16)		(0.08)	
SampleC	41.38	0.16	80.91	−27.42	41.55
	(0.00)	(0.11)		(0.00)	
N=500, T=600					
K's τ^2	11.79	0.59	30.70	4.16	1.56
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.53)		(0.19)	
VT	18.92	0.47	42.96	1.75	0.63
	(0.00)	(0.45)		(0.18)	
LW	13.40	0.48	31.96	2.90	3.80
	(0.00)	(0.29)		(0.17)	
SampleC	22.22	0.25	58.90	−4.43	14.13
	(0.00)	(0.08)		(0.01)	
N=500, T=1250					
K's τ^2	12.05	0.64	31.78	4.76	1.26
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.38)		(0.14)	
VT	18.94	0.48	43.02	1.86	0.61
	(0.00)	(0.27)		(0.13)	
LW	12.53	0.47	32.03	2.80	2.43
	(0.00)	(0.04)		(0.02)	
SampleC	13.12	0.41	33.81	1.90	3.10
	(0.00)	(0.02)		(0.01)	

All values in parentheses are p-values comparing the model with the benchmark K's τ^2.

Table 3.9: Summary Statistics of Weights of Different Portfolio Strategies (Mean-Variance)

	Avg Min	Avg Max	Avg.. Pos. w	Avg. Conc.	Avg. Size of Short
N=100, T=105					
K's τ^2	−0.09	0.19	56.27	0.21	1.06
EW	0.01	0.01	100.00	0.01	0.00
VT	−0.03	0.06	71.60	0.04	0.27
LW	−0.12	0.17	57.52	0.30	1.60
SampleC	−1.21	1.28	51.27	14.52	11.07
N=500, T=525					
K's τ^2	−0.03	0.08	52.55	0.08	1.75
EW	0.00	0.00	100.00	0.00	0.00
VT	−0.01	0.02	72.00	0.01	0.28
LW	−0.06	0.08	53.06	0.23	3.62
SampleC	−0.55	0.56	50.54	6.09	18.53
N=1000, T=1050					
K's τ^2	−0.02	0.06	51.34	0.05	2.00
EW	0.00	0.00	100.00	0.00	0.00
VT	−0.01	0.01	72.59	0.00	0.28
LW	−0.04	0.06	52.42	0.16	4.25
SampleC	−0.31	0.31	50.55	3.33	19.60

Table 3.10: Summary Statistics of Weights of Different Portfolio Strategies (Mean-Variance)

	Avg Min	Avg Max	Avg.. Pos. w	Avg. Conc.	Avg. Size of Short
N=500, T=525					
K's τ^2	−0.03	0.08	52.55	0.08	1.75
EW	0.00	0.00	100.00	0.00	0.00
VT	−0.01	0.02	72.00	0.01	0.28
LW	−0.06	0.08	53.06	0.23	3.62
SampleC	−0.55	0.56	50.54	6.09	18.53
N=500, T=600					
K's τ^2	−0.03	0.08	52.65	0.08	1.73
EW	0.00	0.00	100.00	0.00	0.00
VT	−0.01	0.02	72.26	0.01	0.28
LW	−0.06	0.08	53.04	0.23	3.62
SampleC	−0.28	0.29	51.10	1.53	9.03
N=500, T=1250					
K's τ^2	−0.03	0.07	52.90	0.07	1.61
EW	0.00	0.00	100.00	0.00	0.00
VT	−0.01	0.02	73.27	0.01	0.27
LW	−0.07	0.09	53.11	0.20	3.11
SampleC	−0.20	0.21	52.56	0.39	3.89

B. Full GMVP Comparison (including K's τ^1)

Table 3.11: Performance Comparison of Different Portfolio Strategies (GMVP)

Method	Standard deviation	Sharpe ratio	Maximum drawdown	CE	Avg. Turnover
N=100, T=105					
K's τ^1	12.76	0.43	28.87	2.20	1.60
	(0.00)	(0.11)		(0.06)	
K's τ^2	12.35	0.50	31.49	3.14	1.37
EW	19.32	0.36	55.45	-0.49	0.10
	(0.00)	(0.39)		(0.11)	
VT	16.48	0.43	45.71	1.64	0.17
	(0.00)	(0.59)		(0.25)	
LW	13.53	0.45	39.98	2.44	2.30
	(0.00)	(0.44)		(0.22)	
SampleC	49.24	0.28	81.32	−34.54	29.11
	(0.00)	(0.37)		(0.00)	
N=500, T=525					
K's τ^1	10.95	0.56	32.23	3.72	1.29
	(0.00)	(0.88)		(0.50)	
K's τ^2	10.24	0.57	29.02	3.71	1.12
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.57)		(0.22)	
VT	17.39	0.54	49.74	3.40	0.09
	(0.00)	(0.88)		(0.45)	
LW	11.88	0.42	33.85	2.12	3.38
	(0.00)	(0.14)		(0.10)	
SampleC	36.03	0.17	81.26	−19.69	36.11
	(0.00)	(0.10)		(0.00)	
N=1000, T=1050					
K's τ^1	10.12	0.84	30.44	6.48	1.09
	(0.00)	(0.65)		(0.16)	
K's τ^2	9.30	0.81	26.99	5.83	0.98
EW	21.36	0.47	57.19	1.00	0.10
	(0.00)	(0.09)		(0.09)	
VT	18.41	0.57	51.35	3.64	0.08
	(0.00)	(0.17)		(0.23)	
LW	10.84	0.64	28.04	4.61	3.42
	(0.00)	(0.12)		(0.14)	
SampleC	29.85	0.34	65.68	−7.62	33.60
	(0.00)	(0.05)		(0.01)	

Notes: All values in parentheses are p-values comparing the model with the benchmark K's τ^2. K's τ^1 uses the nonparametric version to calculate Kendall's τ correlation.

Table 3.12: Performance Comparison of Different Portfolio Strategies (GMVP)

Method	Standard deviation	Sharpe ratio	Maximum drawdown	CE	Avg. Turnover
N=500, T=525					
K's τ^1	10.95	0.56	32.23	3.72	1.29
	(0.00)	(0.88)		(0.50)	
K's τ^2	10.24	0.57	29.02	3.71	1.12
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.57)		(0.22)	
VT	17.39	0.54	49.74	3.40	0.09
	(0.00)	(0.88)		(0.45)	
LW	11.88	0.42	33.85	2.12	3.38
	(0.00)	(0.14)		(0.10)	
SampleC	36.03	0.17	81.26	−19.69	36.11
	(0.00)	(0.10)		(0.00)	
N=500, T=600					
K's τ^1	10.95	0.59	31.45	4.04	1.19
	(0.00)	(0.54)		(0.36)	
K's τ^2	10.26	0.62	28.54	4.28	1.05
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.40)		(0.18)	
VT	17.42	0.55	49.82	3.44	0.09
	(0.00)	(0.64)		(0.38)	
LW	11.72	0.50	31.d29	3.16	3.21
	(0.00)	(0.26)		(0.18)	
SampleC	19.58	0.23	62.67	−3.25	12.35
	(0.00)	(0.05)		(0.01)	
N=500, T=1250					
K's τ^1	11.18	0.73	31.88	5.67	0.78
	(0.00)	(0.28)		(0.10)	
K's τ^2	10.54	0.67	29.46	4.85	0.71
EW	20.20	0.46	56.80	1.16	0.10
	(0.00)	(0.26)		(0.13)	
VT	17.60	0.55	50.72	3.40	0.08
	(0.00)	(0.43)		(0.29)	
LW	11.11	0.50	29.02	3.09	1.89
	(0.00)	(0.04)		(0.03)	
SampleC	11.70	0.44	30.83	2.39	2.51
	(0.00)	(0.03)		(0.02)	

Notes: All values in parentheses are p-values comparing the model with the benchmark K's τ^2. K's τ^1 uses the nonparametric version to calculate Kendall's τ correlation.

C. Simulation — Herfindahl Index and Short Positions Size

Table 3.13: Herfindahl index for simulated data.

	$N = 50$			$N = 100$			$N = 500$		
Strategy	$T = 505$	$T = 1250$	$T = 2500$	$T = 505$	$T = 1250$	$T = 2500$	$T = 505$	$T = 1250$	$T = 2500$
Noise Level					20%				
K's τ^2	0.15	0.15	0.15	0.07	0.07	0.07	0.03	0.02	0.02
True	1.76	1.76	1.76	1.53	1.53	1.53	1.17	1.17	1.17
LW	0.45	0.79	1.09	0.21	0.37	0.62	0.09	0.06	0.09
SampleC	1.73	1.71	1.72	1.59	1.55	1.55	7.47	1.26	1.24
Noise Level					50%				
K's τ^2	0.12	0.10	0.10	0.26	0.24	0.23	0.81	0.76	0.73
True	1.09	1.09	1.09	1.10	1.10	1.10	1.05	1.05	1.05
LW	0.05	0.09	0.16	0.04	0.05	0.08	0.03	0.03	0.03
SampleC	1.13	1.11	1.10	1.16	1.14	1.14	6.19	1.10	1.10
Noise Level					80%				
K's τ^2	0.34	0.34	0.34	0.17	0.17	0.18	0.02	0.02	0.02
True	1.02	1.02	1.02	1.03	1.03	1.03	1.02	1.02	1.02
LW	0.02	0.03	0.04	0.01	0.02	0.02	0.01	0.01	0.01
SampleC	1.06	1.05	1.04	1.09	1.07	1.07	4.98	1.05	1.06

Table 3.14: Average short position size for simulated data.

	$N = 50$			$N = 100$			$N = 500$		
Strategy	$T = 505$	$T = 1250$	$T = 2500$	$T = 505$	$T = 1250$	$T = 2500$	$T = 505$	$T = 1250$	$T = 2500$
Noise Level					20%				
K's τ^2	0.58	0.58	0.57	0.75	0.74	0.73	1.11	1.05	1.02
True	1.27	1.27	1.27	1.38	1.38	1.38	1.51	1.51	1.51
LW	0.96	1.06	1.14	1.05	1.09	1.16	2.09	1.40	1.25
SampleC	1.31	1.29	1.28	1.49	1.41	1.39	14.23	1.87	1.66
Noise Level					50%				
K's τ^2	0.19	0.18	0.18	0.12	0.12	0.12	0.05	0.04	0.04
True	0.49	0.49	0.49	0.72	0.72	0.72	1.24	1.24	1.24
LW	0.14	0.17	0.21	0.33	0.35	0.38	1.07	1.02	0.92
SampleC	0.55	0.52	0.51	0.86	0.78	0.76	13.66	1.59	1.38
Noise Level					80%				
K's τ^2	0.04	0.02	0.02	0.10	0.08	0.07	0.54	0.48	0.45
True	0.23	0.23	0.23	0.38	0.38	0.38	0.95	0.95	0.95
LW	0.00	0.00	0.00	0.02	0.04	0.05	0.34	0.49	0.53
SampleC	0.28	0.25	0.25	0.52	0.44	0.42	12.17	1.29	1.09

Bibliography

Acharya, V., Engle, R., and Richardson, M. (2012). Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review*, 102(3):59–64.

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies*, 30(1):2–47.

Adrian, T. and Brunnermeier, M. (2011). Covar. Technical report, National Bureau of Economic Research.

Adrian, T. and Brunnermeier, M. (2016). Covar. *American Economic Review*, 106(7):1705–1741.

Aielli, G. P. (2013). Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics*, 31(3):282–299.

Ao, M., Yingying, L., and Zheng, X. (2019). Approaching mean-variance efficiency for large portfolios. *The Review of Financial Studies*, 32(7):2890–2919.

Armanious, A. (2024). Too-systemic-to-fail: Empirical comparison of systemic risk measures in the eurozone financial system. *Journal of Financial Stability*, 73:101273.

Bali, T., Engle, R., and Murray, S. (2017). *Empirical Asset Pricing: The Cross Section of Stock Returns: An Overview*, pages 1–8. John Wiley Sons, Ltd.

Bandeira, A. S., Lodhia, A., and Rigollet, P. (2017). Marvcenko-Pastur law for Kendall's tau. *Electronic Communications in Probability*, 22(none):1 – 7.

Banulescu-Radu, D., Hurlin, C., Leymarie, J., and Scaillet, O. (2021). Backtesting marginal expected shortfall and related systemic risk measures. *Management Science*, 67(9):5730–5754.

Bao, Z. (2019). Tracy-Widom limit for Kendall's tau. *Annals of Statistics*, 47(6):3504 – 3532.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate garch models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.

Benoit, S., Colliard, J.-E., Hurlin, C., and Pérignon, C. (2017). Where the risks lie: A survey on systemic risk. *Review of Finance*, 21(1):109–152.

Berkes, I. and Horváth, L. (2003). Limit results for the empirical process of squared residuals in garch models. *Stochastic Processes and their Applications*, 105(2):271–298.

Beutner, E., Heinemann, A., and Smeekes, S. (2021). A justification of conditional confidence intervals. *Electronic Journal of Statistics*, 15:2517–2565.

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *Annals of Statistics*, 36(6):2577–2604.

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society*, 12(5):788–792.

Boldin, M. (2000). On empirical processes in heteroscedastic time series and their use for hypothesis testing and estimation. *Mathematical Methods of Statistics*, 9(1):65–89.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *Review of Economics and Statistics*, pages 498–505.

Bomze, I. M., D'Onofrio, F., Palagi, L., and Peng, B. (2025a). Feature selection in linear support vector machines via a hard cardinality constraint: A scalable conic decomposition approach. *European Journal of Operational Research*.

Bomze, I. M. and Peng, B. (2023). Conic formulation of qpccs applied to truly sparse qps. *Computational Optimization and Applications*, 84(3):703–735.

Bomze, I. M., Peng, B., Qiu, Y., and Yıldırım, E. A. (2025b). On tractable convex relaxations of standard quadratic optimization problems under sparsity constraints. *Journal of Optimization Theory and Applications*, 204(3):38.

Bouchaud, J.-P. and Potters, M. (2009). Financial applications of random matrix theory: a short review. *arXiv preprint arXiv:0910.1205*.

Brownlees, C. and Engle, R. F. (2017). Srisk: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies*, 30(1):48–79.

Bun, J., Bouchaud, J.-P., and Potters, M. (2016). My beautiful laundrette: Cleaning correlation matrices for portfolio optimization.

Burer, S. (2009). On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120(2):479–495.

Calvet, L. E., Campbell, J. Y., and Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy*, 115(5):707–747.

Campbell, J. Y. (2006). Household finance. *Journal of Finance*, 61(4):1553–1604.

Chen, S. X. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, 6(1):87–107.

Chen, S. X. and Tang, C. Y. (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics*, 3(2):227–255.

Comte, F. and Lieberman, O. (2003). Asymptotic theory for multivariate garch processes. *Journal of Multivariate Analysis*, 84(1):61–84.

Craig MacKinlay, A. and Pástor, L. (2000). Asset pricing models: Implications for expected returns and portfolio selection. *The Review of Financial Studies*, 13(4):883–916.

De Bandt, O. and Hartmann, P. (2002). Systemic risk: A survey, in financial crisis, contagion and the lender of last resort: A book of readings. ed. by c. goodhart, and g. illing.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management science*, 55(5):798–812.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.

Dimitriadis, T. and Hoga, Y. (2024). Dynamic covar modeling. *arXiv preprint arXiv:2206.14275*.

Edirisinghe, C. and Zhou, W. (2014). Portfolio optimization using rank correlation. In *Encyclopedia of Business Analytics and Optimization*, pages 1866–1879. IGI Global Scientific Publishing.

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. and Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory*, 11(1):122–150.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.

Enke, B. and Zimmermann, F. (2019). Correlation neglect in belief formation. *The Review of Economic Studies*, 86(1):313–332.

Espana, T., Coz, V. L., and Smerlak, M. (2024). Kendall correlation coefficients for portfolio optimization. *arXiv preprint arXiv:2410.17366*.

Eyster, E. and Weizsäcker, G. (2010). Correlation neglect in financial decision-making.

Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606.

Financial Stability Board (2011). Policy measures to address systemically important financial institutions. Consultation paper, November.

Financial Stability Board (2022). 2022 list of global systemically important banks (g-sibs). Consultation paper, November.

Francq, C. and Sucarrat, G. (2017). An equation-by-equation estimator of a multivariate log-garch-x model of financial returns. *Journal of Multivariate Analysis*, 153:16–32.

Francq, C. and Thieu, L. Q. (2019). Qml inference for volatility models with covariates. *Econometric Theory*, 35(1):37–72.

Francq, C. and Zakoïan, J.-M. (2012). Qml estimation of a class of multivariate asymmetric garch models. *Econometric Theory*, 28(1):179–206.

Francq, C. and Zakoïan, J.-M. (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics*, 184(1):158–173.

Francq, C. and Zakoïan, J.-M. (2016). Estimating multivariate garch models equation by equation. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 78(3):613–635.

Francq, C. and Zakoïan, J.-M. (2018). Estimation risk for the var of portfolios driven by semi-parametric multivariate models. *Journal of Econometrics*, 205(2):381–401.

Francq, C. and Zakoïan, J.-M. (2025). Inference on dynamic systemic risk measures. *Journal of Econometrics*, 247:105–936.

Gao, F. and Song, F. (2008). Estimation risk in garch var and es estimates. *Econometric Theory*, 24(5):1404–1424.

Gerber, S., Markowitz, H., Ernst, P., Miao, Y., Sargen, P., et al. (2022). The gerber statistic: A robust co-movement measure for portfolio optimization. *Journal of Portfolio Management*, 48(3):87–102.

Girardi, G. and Ergün, A. T. (2013). Systemic risk measurement: Multivariate garch estimation of covar. *Journal of Banking & Finance*, 37(8):3169–3180.

Goetzmann, W. N. and Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, 12(3):433–463.

Green, R. C. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *Journal of Finance*, 47(5):1785–1809.

Hafner, C. M. and Preminger, A. (2009). On asymptotic theory for multivariate garch models. *Journal of Multivariate Analysis*, 100(9):2044–2054.

Han, H. and Kristensen, D. (2014). Asymptotic theory for the qmle in garch-x models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics*, 32(3):416–429.

Hjort, N. L. and Pollard, D. (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806*.

Hoga, Y. (2023). The estimation risk in extreme systemic risk forecasts. *Econometric Theory*, pages 1–50.

Hou, D., Tang, T., and Toh, K.-C. (2025a). A low-rank augmented lagrangian method for doubly nonnegative relaxations of mixed-binary quadratic programs. *Operations Research*.

Hou, D., Tang, T., and Toh, K.-C. (2025b). A low-rank augmented lagrangian method for polyhedral-sdp and moment-sos relaxations of polynomial optimization. *arXiv preprint arXiv:2512.06359*.

Idier, J., Lamé, G., and Mésonnier, J.-S. (2014). How useful is the marginal expected shortfall for the measurement of systemic exposure? a practical assessment. *Journal of Banking & Finance*, 47:134–146.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58(4):1651–1683.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate arch models. *Econometric Theory*, 14(1):70–86.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Kallir, I. and Sonsino, D. (2009). The neglect of correlation in allocation decisions. *Southern Economic Journal*, 75(4):1045–1066.

Karoui, N. E. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Annals of Statistics*, 36(6):2717–2756.

Keynes, J. M. (1983). Keynes as an investor. *The Collected Works of John Maynard Keynes*, 12:1–113.

Khokhlov, V. (2016). Conditional value-at-risk for elliptical distributions. *Evropskỳ vcasopis ekonomiky a managementu*, 2(6):70–79.

Kirby, C. and Ostdiek, B. (2012). It's all in the timing: simple active portfolio strategies that outperform naïve diversification. *Journal of Financial and Quantitative Analysis*, 47(2):437–467.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1):53–89.

Kulperger, R. and Yu, H. (2005). High moment partial sum processes of residuals in garch models and their applications. *Annals of Statistics*, 33(5):2395–2422.

Laudenbach, C., Ungeheuer, M., and Weber, M. (2023). How to alleviate correlation neglect in investment decisions. *Management Science*, 69(6):3400–3414.

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.

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859.

Ledoit, O. and Wolf, M. (2022a). The power of (non-) linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*, 20(1):187–218.

Ledoit, O. and Wolf, M. (2022b). Quadratic shrinkage for large covariance matrices. *Bernoulli*, 28(3):1519–1547.

Lindskog, F., McNeil, A., and Schmock, U. (2003). Kendall's tau for elliptical distributions. In *Credit Risk: Measurement, Evaluation and Management*, pages 149–156. Springer.

Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector arma-garch model. *Econometric Theory*, 19(2):280–310.

Marvcenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457.

Markowitz, H. (1952a). Modern portfolio theory. *Journal of Finance*, 7(11):77–91.

Markowitz, H. (1952b). Portfolio selection. *Journal of Finance*, 7(1):77–91.

Mogstad, M., Romano, J. P., Shaikh, A. M., and Wilhelm, D. (2024). Inference for ranks with applications to mobility across neighbourhoods and academic achievement across countries. *Review of Economic Studies*, 91(1):476–518.

Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234.

Nolde, N. and Zhang, J. (2020). Conditional extremes in asymmetric financial markets. *Journal of Business & Economic Statistics*, 38(1):201–213.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2):388–413.

Perea, M. D. S., Dunne, P. G., Puhl, M., and Reininger, T. (2019). Sovereign bond-backed securities: A var-for-var and marginal expected shortfall assessment. *Journal of Empirical Finance*, 53:33–52.

Rothman, A. J., Levina, E., and Zhu, J. (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104(485):177–186.

Scaillet, O. (2005). Nonparametric estimation of conditional expected shortfall. *Assurances et Gestion des Risques*, 72(4):639–660.

Sheppard, K. (2009). Mfe matlab function reference financial econometrics. *URL https://www. kevinsheppard. com/images/9/95/MFE_ Toolbox_ Documentation. pdf*.

Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate garch models. In *Handbook of Financial Time Series*, pages 201–229. Springer.

Statman, M. (1987). How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis*, 22(3):353–363.

Tse, Y. K. (2000). A test for constant correlations in a multivariate garch model. *Journal of Econometrics*, 98(1):107–127.

Ungeheuer, M. and Weber, M. (2021). The perception of dependence, investment decisions, and stock prices. *Journal of Finance*, 76(2):797–844.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Van Der Vaart, A. W. and Wellner, J. A. (2007). Empirical processes indexed by estimated functions. *Lecture Notes-Monograph Series*, pages 234–252.

Vyazilov, A. E. (1999). The residual empirical distribution function in garch (1, 1) and its application in hypothesis testing. *Russian Mathematical Surveys*, 54(4):856.

Westfall, P. H. and Young, S. S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. John Wiley & Sons.

White, H., Kim, T.-H., and Manganelli, S. (2015). Var for var: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics*, 187(1):169–188.

Zhao, L., Chakrabarti, D., and Muthuraman, K. (2019). Portfolio construction by mitigating error amplification: The bounded-noise portfolio. *Operations Research*, 67(4):965–983.