Erasmus University Rotterdam
Erasmus School of Economics
Master Thesis Quantitative Finance

Factor Covariance Estimation: Static versus Dynamic modelling in a Portfolio Optimisation setting

E.P.M. Gijssel (475841)

Supervisor:	R. Lönn
Supervisor PGGM:	M. van der Meer
Second assessor:	J. Durieux
Date:	25th July 2023

The content of this thesis is the sole responsibility of the author and does not reflect the view of the supervisors, second assessor, PGGM, Erasmus School of Economics or Erasmus University.

Abstract

In this research, we aim to accurately estimate the components of a factor covariance matrix, where the factors are defined as factor mimicking portfolios. We study the performance of static shrinkage methods summarised in Ledoit & Wolf (2022a), dynamic multivariate GARCH models, such as the Dynamic Conditional Correlation model of Engle (2002), and a combination of the two. Their performance is assessed by adopting a portfolio management setting where the covariance estimate is essential to form minimum variance and mean-variance efficient portfolios. We test whether there are possibilities to improve the current covariance estimation method of PGGM.

This paper extends the current literature in three significant ways; it is the first to combine the new shrinkage method of Ledoit & Wolf (2022b) with dynamic models, to consider the performance of the dynamic models in relation to factor covariances and to investigate covariance matrices of this size. Existing literature discusses either a large set of assets or such a set modelled under a parsimonious factor structure.

We find that the shrinkage model with constant correlation gives the best outcomes when considering a relatively small number of factors, while the nonlinear shrinkage estimator of Ledoit & Wolf (2022b) is the top performer in a higher dimension. The new models show improvement possibilities compared to PGGM's current estimator in both dimensions. The empirical findings are supported by a simulation study.

Contents

1 Introduction

Factor investing enhances diversification, generates above-market returns and manages risk. Its way to do so is by targeting broad, persistent, and long-recognised drivers of returns, also known as factor risk premia (Blitz & Vidojevic, 2019). The number of known factors nowadays is high, ranging from factors based on the ticker sign of assets, to factors that follow the Twitter engagement of companies. The factors are split into style and macroeconomic factors (Pagolu et al., 2016). After decades of theoretical and empirical academic research, along with the practice of many asset managers across different asset classes, the footprint of factor investing has grown broader and deeper (Ma et al., 2022).

Factor portfolio research builds upon the foundations laid by Markowitz (1952), which introduces the mean-variance portfolio theory and the CAPM model of Sharpe (1964). Fama & French (1992) find empirical contradictions in the CAPM model and propose a new factor model. This proposition resulted in one of the most well-known papers in finance, where Fama & French suggest a three-factor model consisting of size, book-to-market and excess market return factors to explain returns. As such, a profitable portfolio can be constructed based on return drivers. In practice, factor portfolios that provide exposures to these return drivers/factors are constructed by sorting stocks on certain characteristics and buying those that score high and optionally short-selling those that score low.

The exposures to the different factors for each asset are crucial in constructing a profitable factor portfolio. In order to accomplish this, information about the mean and covariance matrix of the assets within the asset space is required. The covariance matrix serves two primary purposes in quantitative investment management. Firstly, it predicts the covariance between assets within a portfolio and, therefore, the risk of holding such a portfolio into the future. This plays a central role in risk management. Secondly, we require the covariance matrix estimates in mean-variance optimisation (MVO), one of the first well-known techniques in quantitative investment management by Markowitz (1952), which outlines a method for constructing portfolios with maximum expected return per unit of risk.

The estimation of the covariance matrix brings quite some challenges. The most obvious model to employ is the sample covariance which is relatively simple in construction, unbiased and intuitively appealing. However, this method brings a dimensionality problem; a multi-asset model can easily contain more than 2000 assets, requiring the estimation of more than 1 million independent elements. Consequently, the estimates are susceptible to noise and spurious relations (Ledoit & Wolf, 2003). When the number of stocks is considerably large, especially relative to the number of historical return observations, the sample covariance is estimated with much error (Ledoit & Wolf, 2004a).

The problem with mean-variance optimisation is its vulnerability to these errors; the highest weights are placed precisely where the covariance estimates are least reliable. This leads to out-of-sample underperformance (Michaud, 1989). In practice, this means that a fund manager's realised track records will underrepresent its true stock-picking abilities. Also, the sample variance estimate is often singular, making it impossible to perform portfolio optimisation. Therefore, financial literature starts to rely on more sophisticated statistical models and research into this field has increased over the last few years to circumvent the aforementioned problems.

In line with previous literature, this research assumes that the high-dimensional covariance matrix of assets can be captured by the sum of a factor covariance matrix and an idiosyncratic diagonal matrix (Santos & Moura, 2014). Consequently, the dimension of the covariance matrix reduces from approximately 2000 assets to 40 factors. Nevertheless, the estimation of a factor covariance still requires further investigation as estimating an independent factor covariance matrix, the main interest in this research, is not discussed extensively in the existing literature.

This research aims to compare several methods of estimating the factor covariance matrix, weigh their advantages and disadvantages, and give a recommendation of the optimal estimator and its implementation. We do so by evaluating the accuracy of the covariance matrix by a loss function and constructing global minimum variance (GMV) portfolios. In these portfolios, the weights are mostly determined by the covariance matrix and the accuracy of its estimation. We aim to find the model that gives the best portfolio performance in terms of metrics in line with contemporary literature.

The concern with estimating large covariance matrices has been the subject of various research. There are two main kinds of solutions proposed to tackle the high-dimensionality problem. Firstly, a dynamic approach by fitting multivariate GARCH models, as discussed in for example Engle (2002). Secondly, static shrinkage methods summarised in Ledoit & Wolf (2022a). We discuss both the dynamic and static methods and a combination of the two.

In this paper, we focus on Dynamic Conditional Correlation GARCH models, introduced by Engle (2002) and one of its extensions, the Asymmetric Dynamic Conditional Correlation model, that allows for series-specific news impact and symmetries in correlation dynamics (Cappiello et al., 2006). Moreover, we study the Dynamic Equicorrelation model of Engle & Kelly (2012) that assumes that all pairs of returns have the same correlation on a given day, but that the correlation varies over time. The DECO model makes computations easier in higher dimensions compared to the DCC and ADCC models. Lastly, the Generalised Orthogonal GARCH of Van der Weide (2002) is applied that linearly transforms the observed data into a set of uncorrelated components by an orthogonal matrix. Next to the dynamic models, we study linear and nonlinear shrinkage methods combined with different targets, summarised in Ledoit & Wolf (2022a). We consider the identity, one-factor and constant correlation matrix as targets. Additionally, this research aims to apply the linear and nonlinear shrinkage methods in combination with different GARCH models, as discussed in Engle et al. (2019).

We assess the performance of the models by conducting both a simulation study and a backtest on an empirical dataset. We perform a direct evaluation using a loss function described in Engle et al. (2019) and an indirect evaluation by assessing the performance of the global minimum variance (GMV) and Sharpe ratio (SR) portfolios.

Overall, this research finds that the static estimators give better results compared to the dynamic ones. When considering a smaller covariance dimension, linear shrinkage towards the constant correlation matrix performs the best for both the simulation and empirical study, giving the lowest loss function values and the lowest standard deviation for the GMV portfolios. For a higher dimension, the nonlinear shrinkage method shows the most promising results, being consistently the best for the different simulations and the empirical GMV portfolio. Note that the SR portfolios show some interesting results, but are generally not insightful for determining the best estimator due to large estimation errors in the mean in practice.

This research is relevant because many econometric and data-science applications require a reliable covariance matrix estimate. As already discussed before, it is of importance to portfolio managers who require the covariance matrix as the input for finding optimal portfolios, for example in Markowitz's portfolio selection. By improving covariance matrix estimation, fund managers can create a better risk-weighted portfolio and consequently deliver better returns to their clients. In particular, this paper enhances the current estimation method of the factor matrix of PGGM by considering different statistical models. Consequently, improving the portfolio construction method and the performance of the resulting factor-based portfolios in practice.

Although financial professionals widely use covariance estimation in practice, there is little published academic literature available on covariance estimation in a setting of factors and the considered dimension size. Often papers consider a granular set of assets or only one factor, like the CAPM. Additionally, literature discusses modelling the cross-sectional relations of factors, while this paper uses it for portfolio optimisation. This research thus adds to the current academic literature by filling a gap; being one of the first to discuss factor covariances in a portfolio optimisation setting. Moreover, it is one of the first to find suitable models for a covariance matrix of this scale. Lastly, this research is the first to consider the recently discussed quadratic nonlinear shrinkage method of Ledoit & Wolf (2022b) combined with dynamic models.

The remainder of this paper is outlined as follows; Section 2 introduces the considered literature, whereafter Section 3 explains the methodology. The data characteristics are described in Section 4. Next, Sections 5 and 6 discuss the results. The paper ends with a conclusion and discussion in Sections 7 and 8.

2 Literature

The main focus of this paper is to improve the covariance estimation method as described in contemporary literature, as discussed below. This section first briefly introduces factor investing, the discussion about certain factors and the purpose of the covariance in this framework. Afterwards, we dive deeper into the literature devoted to covariance estimation and research on handling the aforementioned curse of dimensionality.

2.1 Factor discussion

This research assumes that the factors that drive asset returns are value, quality, size, momentum, low volatility and the market factor. Literature agrees on the existence of these factors, but there is much debate about where the outperformance originates from (Fama & French, 1992). Value and size are the two most well-known factors and are part of the Fama-French three-factor model of Fama & French (1992). The value factor consists of buying stocks that have relatively low prices compared to their fundamental metrics, also known as buying relatively cheap stocks. Fama & French (1992) document this value premium, showing that value stocks outperform growth stocks in the long run. Some papers attribute this outperformance to risk. For example, Fama & French (1992) mention that value stocks are fundamentally riskier; they have riskier (i.e. more volatile) earnings and profitability compared to growth firms.

Therefore, the higher average returns are compensation for the higher risk.

Others support a behavioral explanation for the outperformance. Among them are Lakonishok et al. (1994), who argue that value strategies yield higher returns because they exploit the suboptimal behaviour of the typical investor. They note that both individual and institutional investors have a preference for growth strategies and an avoidance of value strategies. Potentially because most investors have shorter time horizons than required for value strategies to pay off consistently and overestimate the growth potential of growth stocks. Individuals want high abnormal returns within a few months rather than a slow, steady growth over the years.

The size effect refers to the tendency for small-cap stocks to outperform large-cap stocks over the long term, as discussed by Banz (1981). Similar to value, the root of the outperformance of small stocks has been a topic of debate. Many papers agree that it proves misspecification of the capital asset pricing model rather than inefficient capital markets (Berk, 1995). The author argues that the relative firm size measures risk. He states that the size factor predicts the expected return because of the theoretical risk premium contained in the market characteristics of assets. Namely, firms with lower market values are more likely to be riskier than firms with high market values, indicating a negative correlation in the cross-section.

On the other hand, Stoll & Whaley (1983) refer to liquidity reasons, believing that small caps tend to be less liquid and therefore more expensive to trade. Zhang (2006) attributes the outperformance to information uncertainty and Chan & Chen (1991) to financial stress. Lastly, behavioural arguments are given in Barberis et al. (1998) and Daniel et al. (1998), suggesting that investors overvalue large stocks and undervalue small stocks due to cognitive biases.

Jegadeesh & Titman (1993) argue that strategies that buy stocks that have done well in the past and sell stocks that have performed poorly generate positive returns. Carhart (1997) proposes adding such a momentum factor to the three-factor model. Explanations given for the source of the momentum premium are again risk- and behavioural-based. Risk-based explanations are by Johnson (2002) and Sagi & Seasholes (2007), that argue that winners have gotten riskier due to their past outperformance. Other research, such as Jegadeesh & Titman (2001), suggests that momentum relates to investor herd behaviour, where investors follow trends and buy or sell based on others.

Fama & French (2015) use a theoretical starting point of the dividend discount model and deduce the profitability and investment factors from it. Both are referred to as quality factors. They use theoretic reasoning to motivate the addition of the two factors and test this hypothesis empirircally. This way of deducing the factors is different from their earlier research, Fama & French (1992), where they use a more data-driven approach by applying data sorting and portfolio construction. The quality factor argues that stocks with high profitability tend to have higher returns than stocks with low profitability. Fama & French (2015) find that the five-factor model better explains average returns associated with major anomalies that the three-factor model does not target. Again, some attribute the outperformance to risk, while others believe in a behavioural explanation. Moreover, a firm's quality is difficult to determine, and there is no consensus on the best way to measure it in literature. As such, the quality factor might be the most controversial of the ones discussed in this literature section. In this research, the quality factor consists of both profitability and investment metrics.

2.2 Curse of sample covariance

A factor portfolio should seek to deliver as much exposure as possible to the targeted factors, minimise the exposure to unwanted risk and ensure the portfolio is sufficiently liquid. To find such a portfolio, it is important to understand how factors relate to one another (through their covariances).

Arnott et al. (2019) mention that many investors mistakenly believe they can diversify away most of the risks in factor investing by creating a portfolio of several factors. However, in periods of market stress, factors begin to move in unison such that most diversification benefits possibly disappear (Arnott et al., 2019). Therefore, understanding how covariances change over time and how they are estimated accurately is essential.

The first option to estimate the covariance matrix is the sample covariance matrix. However, Ledoit & Wolf (2004b) state that as the number of parameters to estimate grows, the estimation of the sample covariance becomes inaccurate. The most extreme coefficients in the matrix then tend to take on extreme values not because it is the population value, but because they contain an extreme amount of error (Ledoit & Wolf, 2004a). Michaud (1989) calls this error maximisation in a mean-variance context. A way to solve this is by shrinking the estimates towards a certain target, summarised in Ledoit & Wolf (2022a). Moreover, the sample covariance becomes singular, leading to a matrix that is not positive definite and non-invertible. This makes it impossible to perform a mean-variance optimisation. Silvennoinen & Teräsvirta (2009) propose the GARCH specifications where the model structure implies and ensures positive definiteness. Moreover, it is important to consider multivariate GARCH (MGARCH) models as co-movements in financial returns are inevitable in asset pricing and risk management (Silvennoinen & Teräsvirta, 2009; Bauwens et al., 2006).

2.3 GARCH models

According to Bauwens et al. (2006), MGARCH modelling provides realistic but parsimonious specifications of the covariance matrix, while ensuring its positive definiteness. When determining the specification for the MGARCH, there is a trade-off between flexibility and parsimony. It should be flexible enough to be able to represent the dynamics of the conditional (co-)variances , but parsimonious enough to allow for relatively easy estimation and interpretation of the model parameters (Silvennoinen & Teräsvirta, 2009).

Silvennoinen & Teräsvirta (2009) categorise the MGARCH models into classes. The first are generalisations of the univariate GARCH model that express the conditional covariance matrix directly, for example the BEKK model of Bollerslev et al. (1988). The second is motivated by parsimony and contains linear combinations of univariate GARCH and factor models. The third expresses the conditional variances and correlations in steps, instead of directly modelling the conditional covariance matrix. This third class includes the Constant Conditional Correlation (CCC) model, Dynamic Conditional Correlation (DCC) model and their extensions. Their appeal lies in the intuitive interpretation of the correlations. Caporin & McAleer (2012) investigate the differences between BEKK and DCC models in practical applications. They argue that the DCC model is more flexible than the BEKK model because it defines the conditional variances separately, enabling the choice for different univariate GARCH models. Silvennoinen

& Teräsvirta (2009) add that BEKK models may have already matured and have little room for improvement. Therefore, we focus in this research on the DCC model and its extensions.

The CCC and DCC models are based on the decomposition of the conditional covariance matrix into conditional standard deviations and correlations (Silvennoinen & Teräsvirta, 2009). The models are built hierarchically: one first chooses a univariate GARCH-type model for each conditional variance, afterwards, one models the conditional correlation matrix and imposes positive definiteness (Bauwens et al., 2006).

The simplest multivariate correlation model is the CCC of Bollerslev (1990). It assumes that the conditional correlation matrix is time-invariant (Silvennoinen & Teräsvirta, 2009). The estimation is computationally attractive as it reduces the number of unknown parameters. However, empirical studies have suggested that the assumption of constant conditional correlations may be too restrictive and unrealistic (Silvennoinen & Teräsvirta, 2009; Bauwens et al., 2006). Engle (2002) introduces the DCC model that assumes a time-dependent conditional correlation matrix. The DCC model has the flexibility of univariate GARCH models but the parsimony of the parametric models for the correlations. This is useful when modelling high-dimensional datasets and ensures positive definiteness (Bauwens et al., 2006). A downside of the DCC models is that they are complex and require much computation time. The parameters are therefore often made scalar. Later, a reformulation of the DCC is proposed by Aielli (2013), which alters the specification of the correlation driving process. However, Engle et al. (2019), among others, argue that the correction effects are irrelevant in empirical applications[1].

In the spirit of the DCC model, Engle & Kelly (2012) propose the Dynamic Equicorrelation (DECO) model with a similar setup. The DECO deviates from the DCC by assuming that all pairs of returns have the same correlation on a given day, but the correlation varies over time. This eliminates the difficulties of high-dimensional systems by having simple analytic inverses resulting in a dramatically simplified and sped-up likelihood calculation. Engle & Kelly (2012) argue that when the true model is DCC, DECO makes estimation feasible when the dimension of the system would be too large for DCC to handle in terms of computation time.

Another extension is the Asymmetric Dynamic Conditional Correlation (AG-DCC) model, which introduces asymmetry in the correlation dynamics and allows for series-specific news impact (Cappiello et al., 2006). It enables the investigation of the presence of asymmetric responses to negative returns by testing whether conditional variances, covariances and correlations of assets are sensitive to the sign of past innovations. The asymmetry specification adds flexibility but increases the number of parameters. Cappiello et al. (2006), therefore, suggest implementing restricted versions of the AG-DCC by making the specification diagonal, scalar (ADCC model) or symmetric. This research implements the ADCC model.

The last multivariate GARCH model this paper considers is the Generalized Orthogonal GARCH (GOGARCH) model of Van der Weide (2002). This model is based on the BEKK model of Engle & Kroner (1995) and Orthogonal GARCH (OGARCH) of Alexander (2002). The GOGARCH assumes that observed data can be linearly transformed into a set of uncorrelated components using an invertible matrix (Van der Weide, 2002). Caporin & McAleer (2014) compare the GOGARCH, DCC models and naive estimation methods and find that the naive

[1]In this empirical application, we disregard this correction

methods generally underperform compared to the dynamic models. Moreover, they stress that direct and indirect evaluation methods give different results, such that there is no optimal model.

Estimation of the MGARCH models can become computationally burdensome for large datasets and statistically overestimates the largest eigenvalues while underestimating the smallest ones (Pakel et al., 2021). Therefore, Pakel et al. (2021) propose an estimation method that is computationally fast; the Maximum Composite Likelihood Estimation (MCLE) approach. The method approximates the full-dimensional joint likelihood function using combinations of lower dimensional marginal densities. It calculates the average of bivariate log-likelihood functions for a large selection of asset pairs as the objective function. Pakel et al. (2021) examine their theory for the BEKK and DCC models of Engle & Kroner (1995) and Engle (2002) by comparing the MCLE to the regular Maximum Likelihood Estimation (MLE) of the full-dimensional model. The paper finds that the MLE suffers from bias while the MCLE does not. The MCLE is especially desirable if the full dimensional likelihood is difficult to work with or not straightforward to specify or compute (Pakel et al., 2021). We apply the MCLE in this research because it is faster and performs similarly to the MLE in most cases (Pakel et al., 2021).

2.4 Linear shrinkage

Another solution to improve covariance matrix estimates is to apply shrinkage. It rectifies the in-sample bias by pushing up the eigenvalues or covariance entries that are too small while pulling down the ones that are too large (Engle et al., 2019). One of the drawbacks of this approach is that the estimators depend on the choice of shrinkage target, which is arbitrary (Coqueret & Milhau, 2014).

Ledoit & Wolf (2003) and Ledoit & Wolf (2004a) are the first to introduce shrinkage. The first paper introduces the topic theoretically, while the second focuses on how to implement the methods to add value to active portfolio management. They propose to shrink the unbiased but unstable sample covariance matrix towards a target to obtain a more efficient estimator. A good target should come as close as possible to the true covariance matrix with as few parameters as possible (Ledoit & Wolf, 2022a).

The first target introduced is the identity matrix. Ledoit & Wolf (2022a) note that variables in many applications do not have mean zero, so it is better to base the sample covariance matrix on the demeaned data instead. Secondly, the biased but less variable single index covariance matrix (consisting of one factor: the market index) is introduced as a target by Ledoit & Wolf (2003). Shrinkage takes a weighted average between the target and the sample covariance. The assigned weight controls how much structure is imposed: the larger the weight, the stronger the structure. In the follow-up paper, Ledoit & Wolf (2004a) choose the constant correlation matrix as a shrinkage target.

Ledoit & Wolf (2003) measure the performance by the out-of-sample standard deviation of the global minimum variance portfolio and show that the shrinkage method outperforms other estimators. Ledoit & Wolf (2004a) show that shrinkage reduces tracking error and increases the realised Sharpe ratio of the active portfolio manager compared to the sample covariance matrix.

There are also other ways to implement shrinkage. Jagannathan & Ma (2003) mention that constraining the weights to be non-negative can be interpreted as implying some form of shrinkage imposed from the weights. Constraining portfolios can reduce the risk in estimated optimal portfolios, even when the constraints are wrong (Jagannathan & Ma, 2003).

2.5 Nonlinear shrinkage

The aforementioned papers all investigate linear shrinkage. In recent years, Ledoit & Wolf (2012) introduce nonlinear shrinkage. Linear shrinkage is simpler to understand, derive and implement. Nevertheless, nonlinear shrinkage can deliver another level of performance improvement, especially if overlaid with stylised facts such as time-varying co-volatility or factor models (Ledoit & Wolf, 2022a). Nonlinear shrinkage applies individual shrinkage intensities to every entry of the sample covariance matrix. For example, entries with relatively more sampling error should be moved more to the shrinkage target. The downside of this method is that it is mathematically more challenging because the number of different shrinkage intensities is growing rapidly (Ledoit & Wolf, 2022a). To perform nonlinear shrinkage, one uses a spectral decomposition of the sample covariance and shrinks different eigenvalues with different shrinkage intensities while keeping the eigenvectors the same (Ledoit & Wolf, 2012). There are different variants of nonlinear shrinkage. For example, Ledoit & Wolf (2012) discuss the oracle estimator, Ledoit & Wolf (2020) give an analytical nonlinear shrinkage method based on kernels and Ledoit & Wolf (2022b) introduce the quadratic inverse shrinkage method (QIS). This paper implements the QIS method because it is the optimal method among all nonlinear shrinkage formulas in terms of accuracy, speed and scalability (Ledoit & Wolf, 2022b). Note that the nonlinear shrinkage especially yields significant improvement compared to its linear counterpart when the dimension size is very large compared to the sample size or the population of eigenvalues is very dispersed. In this research, the matrix dimension is smaller than the sample size. Consequently, we consider this strategy but do not expect a large improvement compared to linear shrinkage.

Moreover, Engle et al. (2019) introduce the possibility of applying (non)linear shrinkage combined with a dynamic model, such as the DCC model. They aim to robustify the DCC model against large dimensions. The most important takeaway of the paper is that it gives superior results to apply shrinkage to the unconditional correlation matrix estimated in the second step of the DCC model rather than to the estimated conditional covariance matrix itself. We follow this approach.

3 Methodology

In this section, we first outline the construction of the factors that span the dataset. Then we discuss the considered estimation methods for the factor covariance matrix with the purpose of reducing the estimation error in high dimensions. Afterwards, we outline several shrinkage methods and their application with different targets. This section ends with the evaluation methods and inference.

3.1 Factor Mimicking Portfolio construction

This subsection discusses the construction of the factor mimicking portfolio (FMP) returns which mimick the factor returns. The latter are the returns of interest for the covariance estimation.

Factors are constructed from regular assets. For this, we first consider the division of the returns into factor returns by an exposure matrix X_t that includes the exposures of the factors that are seen as the drivers of return. Note that the factor returns are unobserved and need to be constructed via FMPs. The factors that are considered are Value, Quality, Low Volatility, Size and Momentum, following PGGM. All factors are characterised by different variables that quantify a certain factor and are outlined in Appendix A. In the remainder of this subsection, we address them as variables. The following factor model is used to derive the factors, as described in academic literature,

$$r_t = X_t f_t + \varepsilon_t, \tag{1}$$

where r_t is a vector $(N \times 1)$ of asset returns, X_t the $(N \times K)$ matrix of factor exposures, f_t a $K \times 1$ vector of factor returns and ε_t the $(N \times 1)$ vector of asset-specific returns. This research aims to estimate the covariance of f_t, further denoted as Σ_t.

Firstly, we determine the matrix X_t that consists of the exposures of the factors towards PGGM seeks active exposure. These are also called scores because they describe how well a certain asset scores on a specific factor.

Each factor is measured by different variables, such that we assign a score to each asset for the variables related to a certain factor. We first transform the variables defined in Appendix A, to a uniform distribution. We do so by ranking the assets for a certain variable j, such that each asset gets a score for variable j. This score shows how well a certain asset performs for a certain variable j compared to the other assets. Afterwards, they are transformed using a normal distribution as follows:

$$Z_j = CDF^{-1}_{normal}\Big(\frac{\text{rank}(j)}{N+1}\Big), \tag{2}$$

where Z_j is the normal distributed score of variable j for each asset. This way, we obtain a score centered around zero for each asset for each variable j. Afterwards, the normal distributed score per variable and asset is split into three different exposure components that do not overlap: industry, region and asset specific. This split is introduced because PGGM neutralises the factors for industry and region exposures. The exposures per component to each variable are obtained by the following regression

$$Z_j = 1\alpha + D_{Ind}\delta_{Ind,j} + D_{Reg}\delta_{Reg,j} + \varepsilon_j, \tag{3}$$

$$\text{with } \; w'_{BM} D_{Ind}\delta_{Ind,j} = 0 \;\; \& \;\; w'_{BM} D_{Reg,j}\delta_{Reg,j} = 0, \tag{4}$$

where D indicates a dummy matrix indicating if an asset corresponds to a certain region or industry. Moreover, ε_j captures the asset-specific information for a certain variable j and is the main point of interest. The regression is solved by constrained OLS to obtain ε_j, $\delta_{Ind,j}$ and $\delta_{Reg,j}$. These parameters are also called scores and determine the input of the exposure matrix X_t. The conditions are required to prevent the multicollinearity introduced by the two dummy matrices D. The weights w_{BM} are the known benchmark weights of the assets where the benchmark is the FTSE Developed Index, which is used as a market proxy. We assume that the benchmark weighted sum of all industry effects is zero. Thus, the assumption is that there is no industry effect on average.

Next, we compute the exposure for each asset to each factor from the asset exposures to each variable. We do this for all three components. This calculation consists of a linear combination of the variable-specific exposures for the variables belonging to a factor, as in Appendix A. We end up with a region-, industry- and asset-specific exposure to each factor. These are then scaled by the square root of the number of variables per factor to ensure that the standard deviation stays relatively constant as a function of the number of subfactors. An example of such calculation is as follows for the value exposure of the asset component ε:

$$\varepsilon_{value} = \frac{1}{\sqrt{4}}(\varepsilon_{BP} + \varepsilon_{EBITDAEV} + \varepsilon_{EPNTM} + \varepsilon_{FCFY}), \tag{5}$$

where BP is the Book-to-Price variable, EBITDAEV is the EBITDA-to-EV ratio, EPNTM is the Expected Profit-to-Price ratio and FCFY is the Free Cashflow-to-Price ratio. This way, we obtain 15 factor exposures that are asset-, industry- and region-specific for each asset. Out of these 15, we only use the factor exposures for the 10 style factors for each asset: the five asset-specific (ε) and the five industry scores (δ_{Ind}). These are placed in the columns of X_t. The next procedure only involves these 10 exposures.

Finally, we create Factor-Mimicking portfolios (FMPs) with the property that such a portfolio has exposure 1 to its corresponding factor and exposure 0 to the other factors. The vector of FMP returns FMP_t and the FMP weights $W_{FMP_{t,i}}$ for each factor i are calculated via a WLS regression as follows:

$$FMP_t = (X_t' W_t^{-1} X_t)^{-1} X_t' W_t^{-1} r_t, \tag{6}$$

$$W_{FMP_{t,i}} = W_t X_t (X_t' W_t X_t)^{-1} \iota_K, \tag{7}$$

where X_t denotes the exposure matrix, r_t gives the asset returns and W_t is a diagonal matrix with square root market caps on the diagonals, giving a higher weight to large firms. As a result of this construction, the time-series covariances in the FMP returns are low. Also, the FMP return (FMP_t) is equal to the return of the factor (f_t) when the weights of the FMP (W_{FMP_t}) are multiplied by the total asset return matrix, hence the name factor mimicking portfolio.

In conclusion, we end up with five region and industry-neutral ("Within") and five industry ("Across") FMPs. These 10 FMPs do not have any overlap and are corrected for the sensitivity of the market, such that they can be seen as independent factor sources of return.

The returns of these 10 FMPs are what the subset of the dataset in Section 4 consists of, together with an FMP for two Axioma factors and a constant that reflects the market FMP without any factor influence. This gives the total subset of 13 FMPs. Moreover, the full dataset of 42 FMP return series consists of the 13 FMPs together with 24 industry and 5 region FMPs based on the dummies that indicate if a certain asset belongs to a certain industry or region. To reiterate the goal of this research; the aim is to estimate the covariances between the factor returns f_t in the considered 13 and 42 FMP dimensions. Due to the construction of the FMPs, this corresponds to estimating the covariance of the FMP returns.

3.2 Covariance estimation methods

The first dynamic model that we implement is the Dynamic Conditional Correlation (DCC) model of Engle (2002). The main idea is first to estimate the individual covariance and GARCH model parameters for all the underlying assets and then use these estimates conditionally to estimate the parameters of the dynamic model. This research follows the notation of Engle (2002) where i ranges from 1 to K, indicating the dimension of the covariance matrix and t ranges from 1 to T where T equals the sample size. The model assumes that the twelfth moment is finite (Engle, 2002). The goal is to find an accurate estimate of Σ_t, defined as $\hat{\Sigma}_t$. The DCC model does this as follows:

$$\hat{\Sigma}_t = D_t R_t^{DCC} D_t, \tag{8}$$

where D_t is the $K \times K$ diagonal matrix of time-varying conditional standard deviations from a univariate GARCH model with $\sqrt{h_{i,t}}$ the ith diagonal element and R_t^{DCC} the time-varying conditional correlation matrix of the returns f_t. In the first stage, the univariate GARCH model is fit for each factor to obtain the estimates $h_{i,t}$. Appendix A gives the GARCH formulation. Afterwards, we use the standardised returns $\varepsilon_{i,t} = f_{i,t}/\sqrt{h_{i,t}}$ to estimate the correlation parameters of R_t in line with Engle (2002). By construction, R_t is the conditional covariance matrix of the vector of standardised returns ε_t. The conditional correlation model and the $K \times K$ symmetric positive definite matrix Q_t is modelled by

$$R_t^{DCC} = Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \tag{9}$$

$$Q_t = (1 - \alpha - \beta)S + \alpha(\varepsilon_{t-1}\varepsilon'_{t-1}) + \beta Q_{t-1}, \tag{10}$$

where $Q_t = [q_{ij,t}]$, $Q_t^* = \mathrm{diag}(q_{11,t} \dots q_{KK,t})$ and $S = [s_{ij}]$ is the unconditional correlation matrix of the returns, calculated as $S = \mathrm{E}[\varepsilon_t \varepsilon'_t]$, such that its sample analogue is $\bar{S} = \frac{1}{T}\sum_{t=1}^{T} \varepsilon'_t \varepsilon_t$, and (α, β) are scalars. We require that Q_t is positive definite such that R_t^{DCC} is a correlation matrix with ones on the diagonal and every other element smaller than 1 in absolute value (Engle, 2002). To ensure that Q_t is positive definite, Engle (2002) assumes $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta < 1$ and S is positive definite. I apply the Maximum Composite Likelihood Estimation (MCLE) method to estimate the DCC parameters, which is described by Pakel et al. (2021) and discussed in Appendix A. The MCLE approximates the full-dimensional joint likelihood function using combinations of lower dimensional marginal densities (Pakel et al., 2021). We estimate the parameters in Equation 10 via a rolling window approach and re-estimate them for each period in time.

Furthermore, the Dynamic Equicorrelation (DECO) model of Engle & Kelly (2012) is studied. It attains simple analytic inverses and determinants such that likelihood estimation is simplified and optimisation becomes easier (Engle & Kelly, 2012). Moreover, the DECO model correlations of factors i and j depend on the return histories of all pairs of factors, while for the DCC model it only depends on the return histories of the two factors concerning the correlation. As a result, the DECO model draws from a broader information set that enables the model to capture a pooling aspect (Engle & Kelly, 2012). The method attains the following dynamics where equicorrelation matrix R_t^{DECO} consists of $K \times 1$ vectors of random variables and is positive definite if:

$$R_t^{DECO} = (1-\rho_t)I_K + \rho_t J_K, \qquad \text{where} \qquad \rho_t = \frac{1}{K(K-1)}(\iota' R_t^{DCC}\iota - K), \tag{11}$$

where ρ_t is the equicorrelation, I_K denotes the K-dimensional identity matrix, J_K is a $K \times K$ matrix of ones and R_t^{DCC} gives the correlations of the returns under the DCC model, as in Equation (9). Moreover, the inverse, R_t^{-1}, only exists if $\rho_t \neq 1$ and $\rho_t \neq \frac{-1}{K-1}$. The covariance is estimated similarly to Equation (8),

$$\hat{\Sigma}_t = D_t R_t^{DECO} D_t, \tag{12}$$

where D_t is the diagonal matrix of conditional volatilities $h_{i,t}$, R_t^{DECO} gives the estimate of the equicorrelation matrix and $\hat{\Sigma}_t$ gives the conditional covariance estimate. The model is estimated by MCLE, similar to the DCC model.

Moreover, we implement the scalar Asymmetric Dynamic Conditional Correlation (ADCC) model of Cappiello et al. (2006). The model adapts the DCC model to allow for conditional asymmetries in correlations and accounts for series-specific news impact and smoothing parameters (Cappiello et al., 2006). We alter Equation (10) following Cappiello et al. (2006), where we plug in the estimates for S and N, $\bar{S}$ and $\bar{N}$ respectively:

$$Q_t = (S - \alpha^2 S - \beta^2 S - g^2 N) + \alpha^2 \varepsilon_{t-1}\varepsilon_{t-1}' + g^2 n_{t-1}n_{t-1}' + \beta^2 Q_{t-1}, \tag{13}$$

where α, β and g are scalars, $n_t = I[\varepsilon_t < 0] \circ \varepsilon_t$ with $I[\cdot]$ the $K \times 1$ indicator function equalling one if the argument is true and zero otherwise. The Hadamard product is denoted by $\circ$ and $N = \mathrm{E}[n_t n_t']$, such that $\bar{N} = \frac{1}{T}\sum_{t=1}^{T} n_t' n_t$. The remaining parameters have the same definition as in the DCC model. The sufficient condition to ensure Q_t is positive definite is given by $\alpha^2 + \beta^2 + \delta g^2 < 1$ where δ =maximum eigenvalue $[S^{-1/2}NS^{-1/2}]$[5]. This condition can be evaluated on the sample data and is implemented during the estimation of the conditional correlation (Cappiello et al., 2006). This model follows the same approach as the DCC model for estimating the parameters. Afterwards, the same procedure as for the DCC model is followed to obtain the covariance estimate. We apply the scalar version of all dynamic models to prevent a huge number of parameters and computation time.

Lastly, this research considers the Generalized Orthogonal GARCH model (GOGARCH) model of Van der Weide (2002), a generalisation of the OGARCH model and nested in the more general BEKK model of Bollerslev et al. (1988). The GOGARCH implicitly assumes that the observed data can be linearly transformed in a set of uncorrelated components by an invertible matrix (Van der Weide, 2002). The most important assumption of the GOGARCH is that the observed returns f_t can be expressed as a linear combination of uncorrelated economic components y_t such that

$$f_t = Zy_t, \tag{14}$$

where Z is a $K \times K$ non-singular matrix that links the unobserved components y_t with the observed returns f_t, is constant over time and invertible. This formulation is familiar to a factor model setup. Note that we already consider factor returns, such that we impose a factor structure on a dataset with an already existing factor dynamic. For the GOGARCH(1,1) process that we apply in this research, the economic components follow the distribution $y_t \sim N(0, H_t)$. The definition of H_t is defined in Equation (17). In order to achieve identification and find the linkage Z, we require the matrices P and Λ that give the orthonormal eigenvectors and the eigenvalues of the unconditional covariance matrix ZZ'. The orthogonal matrix P is only guaranteed to coincide with Z when the diagonal elements of H_t are all distinct. Thus, identification problems arise when some of the uncorrelated components have similar unconditional variance (Van der Weide, 2002). The linkage Z is well identified when conditional information is taken into account. If Z links the uncorrelated components y_t with the factor returns f_t, then there exists an orthogonal matrix U such that:

$$Z = P\Lambda^{1/2}U, \tag{15}$$

where P is the orthogonal matrix of eigenvectors and Λ is the matrix of eigenvalues of $\Sigma = ZZ'$ (Van der Weide, 2002). The factor components are then specified as

$$y_t = H_t^{1/2}u_t, \tag{16}$$

where u_t is a random vector process with the characteristics $\mathrm{E}[u_t] = 0$ and $\mathrm{E}[u_t^2] = 1$. Then H_t is defined as

$$H_t = \mathrm{diag}(h_{1,t}, \ldots, h_{K,t}), \tag{17}$$

where $h_{1,t}, \ldots h_{K,t}$ are described by a GARCH(1,1) process, as given in Appendix A, and $H_0 = I_K$ equals the unconditional covariance matrix of the components. This implies that y_t is a covariance-stationary process with mean zero and unconditional variance I_K. The conditional covariance matrix of the factor returns f_t is given by

$$\hat{\Sigma}_t = ZH_tZ'. \tag{18}$$

The model applies the Maximum Likelihood Estimation (MLE) method with a two-step approach, described in Appendix A, following Van der Weide (2002).

3.3 Shrinkage methods

The goal of shrinkage is to shift an unbiased covariance estimator with much variance, towards an estimator with a nonzero bias, but little variance, known as the shrinkage target (Coqueret & Milhau, 2014). In this research, we combine linear and nonlinear shrinkage with simple targets and the dynamic modelling specification. We first introduce the linear shrinkage method for which we follow a similar approach to Ledoit & Wolf (2003) and Ledoit & Wolf (2004a) among others, where they shrink the sample covariance matrix or exponentially weighted moving average covariance matrix $\bar{\Sigma}_t$ towards a target. The true unobserved covariance is denoted by Σ_t and assumed to be positive definite. We aim to find the shrunk estimate $\tilde{\Sigma}_t$. The general form for linear shrinkage is given by

$$\tilde{\Sigma}_t = \psi_t F_t \ + (1 - \psi_t)\bar{\Sigma}_t, \tag{19}$$

where F_t is the shrinkage target and $\psi_t \in [0, 1]$ gives the shrinkage intensity. As shrinkage targets we consider the identity matrix, one-factor market model and the constant correlation matrix.

The one-factor model is defined as the cross-sectional average of all the random variables. In the constant correlation matrix, the variances are kept the same, while the off-diagonal covariances are adjusted by $\bar{\varrho}\sqrt{\sigma_{ii}\sigma_{jj}}$, where $\bar{\varrho}$ gives the average correlation. The shrinkage intensity is calculated by minimising the Frobenius loss function between the true covariance matrix and the shrinkage estimator $\tilde{\Sigma}_t$ (Ledoit & Wolf, 2003). The shrinkage intensity is calculated by $\psi_t^* = \frac{1}{T}\frac{\pi-\rho}{\gamma} + O(\frac{1}{T^2})$, where π is the sum of asymptotic variances of the entries of the sample covariance s_{ij} scaled by $\sqrt{T}$: $\pi = \sum_{i=1}^{K}\sum_{j=1}^{K} AsyCov[\sqrt{T}s_{ij}]$ for each time period (Ledoit & Wolf, 2003). Furthermore, ρ denotes the sum of asymptotic covariances of entries of the sample covariance matrix s_{ij} with entries of the single-index covariance matrix f_{ij} scaled by $\sqrt{T}$: $\rho = \sum_{i=1}^{K}\sum_{j=1}^{K} AsyCov[\sqrt{T}f_{ij}, \sqrt{T}s_{ij}]]$ (Ledoit & Wolf, 2003). Finally, let γ measure the misspecification of the single-index model: $\gamma = \sum_{i=1}^{K}\sum_{j=1}^{K}(\phi_{ij} - \sigma_{ij})^2$ where ϕ_{ij} indicates elements of the target covariance matrix and σ_{ij} of the true covariance matrix (Ledoit & Wolf, 2003).

Moreover, it is possible to apply nonlinear shrinkage and implement shrinkage together with the DCC models. For both, we require the spectral decomposition of the covariance matrix, given by $\Sigma_t = U_t \Lambda_t U_t$, in more detail:

$$\Sigma_t = \sum_{i=1}^{K} \lambda_{t,i} \cdot u_{t,i} u'_{t,i}, \tag{20}$$

where $\Lambda_t = \text{diag}(\lambda_{t,1} \ldots \lambda_{t,K})$ is a diagonal matrix whose diagonal entries are the sample eigenvalues sorted in descending order and $U_n = (u_{t,1}, \ldots, u_{t,K})$ is an orthogonal matrix of eigenvectors (Ledoit & Wolf, 2022a).

The nonlinear shrinkage method we consider, the quadratic inverse shrinkage method (QIS), is introduced in Ledoit & Wolf (2022b). This research follows their notation.

First of all, this method takes a couple of assumptions that are outlined in Ledoit & Wolf (2022b). The specifications of these assumptions are given in Appendix A. Moreover, Ledoit & Wolf assume that the covariance matrix estimator is part of the rotation class, meaning that it adheres to $\hat{\Sigma}_t = U_t \tilde{\Delta}_t U'_t$, where $\tilde{\Delta}_t$ is a diagonal matrix whose elements, $\tilde{\Delta}_t = \text{Diag}\big(\tilde{\delta}_t(\lambda_{t,1}), \ldots \tilde{\delta}_t(\lambda_{t,K})\big)$, are a function of the eigenvalues λ_t. Moreover, $\tilde{\Delta}_t$ is a real univariate shrinkage function that can depend on Σ_t. For any covariance matrix estimator $\hat{\Sigma}_t$ in the rotation-equivariant class, the Frobenius loss $L_t^{FR}(\Sigma_t, \hat{\Sigma}_t)$ converges in probability to a nonrandom limit as t goes to infinity. This limit is minimised if $\tilde{\delta}_t(\lambda_{t,i}) = \hat{\delta}_{t,i}$ with $\hat{\delta}_{t,i}$ satisfying:

$$\hat{\delta}_{t,i}^{-1} = (1-\frac{K}{t})^2 \lambda_{t,i}^{-1} + 2\frac{K}{t}(1-\frac{K}{t})\lambda_{t,i}^{-1}\hat{\theta}_t(\lambda_{t,i}^{-1}) + (\frac{K}{t})^2 \lambda_{t,i}^{-1} A_{\hat{\theta}_t}^2(\lambda_{t,i}^{-1}) \qquad \text{where} \tag{21}$$

$$\hat{\theta}_t(x) = \frac{1}{k}\sum_{j=1}^{K} \lambda_{t,j}^{-1} \frac{\lambda_{t,j}^{-1} - x}{(\lambda_{t,j}^{-1} - x)^2 + h_t^2 \lambda_{t,j}^{-2}} \qquad \text{and} \tag{22}$$

$$A_{\hat{\theta}_t}^2(x) = \left[\frac{1}{K}\sum_{j=1}^{K} \lambda_{t,j}^{-1} \frac{\lambda_{t,j}^{-1} - x}{(\lambda_{t,j}^{-1} - x)^2 + h_t^2 \lambda_{t,j}^{-2}}\right]^2 + \left[\frac{1}{K}\sum_{j=1}^{K} \lambda_{t,j}^{-1} \frac{h_t \lambda_{t,j}^{-1}}{(\lambda_{t,j}^{-1} - x)^2 + h_t^2 \lambda_{t,j}^{-2}}\right], \tag{23}$$

where Equation (21) is the quadratic shrinkage of the inverse sample covariance matrix eigenvalues, $\lambda_{t,i}$ indicates the eigenvalue of the ith factor at time t and h_t gives the smoothing parameter, also known as the bandwidth formula and is calculated as $h_t = \min(\frac{K^2}{t^2}, \frac{t^2}{K^2})^{0.35} \times K^{-0.35}$ (Ledoit

& Wolf, 2022b). Equations (22) and (23) are used to calculate the two shrinkage targets, $\hat{\theta}_t(\lambda_{t,i}^{-1})$ and $A^2_{\hat{\theta}_t}(\lambda_{t,i}^{-1})$, respectively. These are plugged into Equation (21).

Afterwards, we calculate the shrunk covariance matrix by $\tilde{\Sigma}_t = U_t \hat{\Delta}_t U_t$ where for $\hat{\Delta}_t$ we use the following definition: $\hat{\Delta}_t = \text{diag}\big(\hat{\delta}_t(\lambda_{t,1})\big)$. We plug in the quadratic shrinkage estimator $\hat{\delta}_{t,i}$ from Equation (21) to finally get a value for the shrunk covariance $\tilde{\Sigma}_t$.

Lastly, it is possible to combine shrinkage methods with the DCC models, as discussed in Engle et al. (2019). The shrinkage is then applied to the correlation matrix of the returns S, used in Equation (10). First of all, we require the sample covariance matrix of the devolatilised returns, earlier specified as $\hat{S} = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t' \varepsilon_t$. Then, S is decomposed into eigenvalues and eigenvectors, following Equation (20). For linear shrinkage, We follow Ledoit & Wolf (2004a) where the cross-sectional average of the sample eigenvalues is determined as $\bar{\lambda} = \frac{1}{K}\sum_{i=1}^{K} \lambda_i$. Then the linear shrinkage estimator is expressed as:

$$\tilde{S} = \sum_{i=1}^{K} [\psi \bar{\lambda} + (1-\psi)\lambda_i] \cdot u_i u_i', \tag{24}$$

where ψ gives the optimal shrinkage intensity, indicating how fast the sample eigenvalues are pulled towards their cross-sectional average. Afterwards we plug the shrunk $\tilde{S}$ into Equation (10) and continue the estimation via one of the dynamic models for every t.

We follow a similar approach for nonlinear shrinkage combined with the DCC models. Again the shrinkage is applied to the unconditional correlation matrix of the returns, S. Note that the eigenvectors do not change when applying shrinkage.

We can summarise the estimation of the DCC models combined with (non)linear shrinkage by following the three steps discussed in Ledoit & Wolf (2022a):

1. Fit a univariate GARCH(1,1) model and use the fitted models to devolatise the return series f_t to obtain the estimated series of $\{\varepsilon_t\}$.
2. Estimate the unconditional correlation matrix S by applying (non)linear shrinkage to the series $\{\varepsilon_t\}$ and use the resulting $\tilde{S}$ for correlation targeting via Equation (10).
3. Maximise the MCLE, as discussed in Appendix A, to estimate the two DCC parameters (α, β).
4. Use the obtained DCC parameters combined with the shrunk $\tilde{S}$ to calculate the covariance estimate $\hat{\Sigma}_t$ via Equations (8) and (9).

3.4 Evaluation

To compare the outcomes of various models, we add benchmark models for the factor covariance matrix to the analysis. The first benchmark is the sample covariance matrix. Secondly, we consider the current estimation method by PGGM, that is the exponentially weighted moving average (EWM) model of Axioma combined with linear shrinkage towards the identity matrix. As such, we take the EWM and the shrunk EWM models as the other two benchmarks. The EWM benchmark is discussed in more detail in Appendix A.

Assessing whether the results based on real financial data are accurate out-of-sample is difficult because volatility is an unobservable variable, even ex-post. As such, a direct evaluation involves finding a proxy related to the latent variable. The indirect approach tries to overcome the nature of this variable by comparing forecasts of the minimum variance portfolio. As volatility proxies are often inaccurate, we focus on the indirect evaluation approach for the empirical backtest. For the simulations, we can directly evaluate the results via a mean-variance loss function, discussed in Engle et al. (2019) and described in Equation (A.7).

In order to perform an indirect performance analysis, we consider the global minimum variance (GMV) portfolio of Markowitz (1952). This portfolio only requires the input of the covariance matrix and is therefore a "clean" problem and a good indicator for the accuracy of the covariance matrix (Michaud, 1989). This portfolio is popular due to its simplicity and independence of the expected return Michaud (1989). Additionally, empirical research has shown that stocks with low variances tend to obtain higher (risk-adjusted) returns than more volatile equities (DeMiguel et al., 2009). Therefore, the GMV portfolio is useful for investors that want to lower their risks and those that aim for high returns. We use the out-of-sample covariance forecasts as input. For the simulation, we add the "real" covariance matrix from which we simulate. Without short-selling constraints, the minimum variance problem is defined as follows:

$$\begin{aligned} &\min_{W_t} W_t' \varSigma_t W_t \\ &\text{s.t. } W_t' \iota = 1, \end{aligned} \tag{25}$$

where for $\varSigma_t$ the covariance estimate $\hat{\varSigma}_t$ is plugged in and the portfolio weights are the $K \times 1$ vector of weights in the underlying factors. The analytical solution of this is given by:

$$W_t = \frac{\varSigma_t^{-1} \iota}{\iota' \varSigma_t^{-1} \iota}. \tag{26}$$

For both the simulation and empirical study, we consider a weekly re-balancing approach, where new GMV portfolios are constructed every week. If short-selling is not allowed, we add the constraint $W_t \in (0, 1), \forall t$.

Next to that, we consider an optimal Sharpe ratio (SR) portfolio to determine whether taking into account the expected returns in the optimisation influences the relative performance of the models.

3.5 Inference

We evaluate the characteristics of the portfolios by the mean, standard deviation (SD), information ratio (IR), Sharpe ratio (SR) and turnover. The turnover is especially important when determining the feasibility of holding a portfolio and its formula is given in Equation (A.9). We include the $1/N$ portfolio because it often gives robust and difficult-to-beat results in practice (DeMiguel et al., 2009).

Moreover, Ledoit & Wolf (2011) show that it is possible to test for the significance of the difference in variance of two portfolios. One of the methods the paper suggests is constructing a HAC inference based on the Quadratic Spectral (QS) kernel. Assume that there are two investment strategies j and n whose returns at time t are r_{tj} and r_{tn}, respectively (Ledoit & Wolf, 2011). Then it is assumed that the T return pairs $(r_{1j}, r_{1n})', \ldots, (r_{Tj}, r_{Tn})'$ are from strictly stationary time series, meaning that bivariate return distributions do not change over time. The distribution has mean μ and covariance matrix Σ, given by $\mu = \binom{\mu_j}{\mu_n}$ and covariance $\Sigma = \begin{pmatrix} \sigma_j^2 & \sigma_{jn} \\ \sigma_{jn} & \sigma_n^2 \end{pmatrix}$. The hypothesis uses the ratio of the two variances, given by $\Theta = \frac{\sigma_j^2}{\sigma_n^2}$ with estimator $\hat{\Theta} = \frac{\hat{\sigma}_j^2}{\hat{\sigma}_n^2}$. The hypothesis follows Ledoit & Wolf (2008), where the log-transformation results in better finite-sample properties,

$$\Delta = \log(\Theta) = \log(\sigma_j^2) - \log(\sigma_n^2) \tag{27}$$

$$H_0 : \Delta = 0 \qquad \text{vs.} \qquad H_1 : \Delta \neq 0. \tag{28}$$

Furthermore, in order to construct a confidence interval and calculate the p-values, we require the standard errors for $\hat{\Delta}$,

$$s(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{v}) \hat{\Psi} \nabla f(\hat{v}}{T}}, \tag{29}$$

where $\Delta = f(v)$ with $v = (\mu_j, \mu_n, \gamma_j, \gamma_n)'$, $\gamma_i = \mathrm{E}(r_{1i}^2)$ and $f(a, b, c, d) = \frac{a}{\sqrt{c-a^2}} - \frac{b}{\sqrt{d-b^2}}$ (Ledoit & Wolf, 2008). The sample counterparts are denoted by $\hat{\cdot}$. We assume $\sqrt{T}(\hat{v} - v) \rightarrow N(0; \nabla' f(v) \Psi \nabla f(v))$ where Ψ is an unknown symmetric positive semi-definite matrix (Ledoit & Wolf, 2008). This relation holds under mild regularity conditions: finite $4 + \delta$ moments where δ is a small positive constant, together with appropriate mixing conditions (Ledoit & Wolf, 2008). It is possible to consistently estimate Ψ by heteroskedasticity and autocorrelation consistent (HAC) kernel methods. Appendix A gives the details of the HAC inference.

After we estimate and obtain $\hat{\Psi}$, it is possible to construct the standard errors $s(\hat{\Delta})$ by Equation (29). A two-sided p-value for the null hypothesis $H_0 : \Delta = 0$ is given by

$$\hat{p} = 2\Phi\left(- \frac{|\hat{\Delta}|}{s(\hat{\Delta})} \right). \tag{30}$$

where $\Phi(\cdot)$ denotes the c.d.f. of the standard normal distribution. One of the drawbacks is that the HAC inference is often liberal when sample sizes are small to moderate (Ledoit & Wolf, 2011). Meaning that hypothesis tests tend to reject a true null hypothesis too often compared to the nominal significance level. Similarly to Ledoit & Péché (2011), we use a Parzen kernel instead of the Quadratic Spectral kernel because of computation time. A similar approach can be conducted to verify whether the difference between the Sharpe ratios of portfolios is significantly different from zero which is discussed in Ledoit & Wolf (2008). The hypothesis then becomes $\Delta = SR_j - SR_n = \frac{\mu_i}{\sigma_j} - \frac{\mu_n}{\sigma_n}$, whereafter a similar procedure is conducted.

As a robustness check, we investigate the impact of different estimation windows for the simulations. Also, the consequences of implementing a short-selling constraint on the performance of the different models are examined for the GMV portfolios. This constraint is applicable for PGGM and therefore relevant to incorporate into this paper. Lastly, we conduct a high- versus low-volatility period analysis to assess the performance of GMV portfolios for the considered estimators in different economic circumstances. Their outcomes are given in Appendix E.

4 Data

In this research, we consider a universe consisting of the 2171 constituents of the FTSE Developed Index. This index contains stocks from developed markets worldwide and covers most of the investable market capitalisation containing large and mid-sized companies. However, due to regulations within PGGM, stocks related to controversial weapons and tobacco are excluded or if certain criteria are not met concerning human rights, social circumstances and environment. Additionally, we remove stocks if they do not meet the liquidity requirements or if the company is involved in a merger or being acquired by another firm. Lastly, if the stock is available on different listings, the one that provides the most liquidity is used. Note that the number of assets that adhere to these rules and stay within the investment universe can change over time.

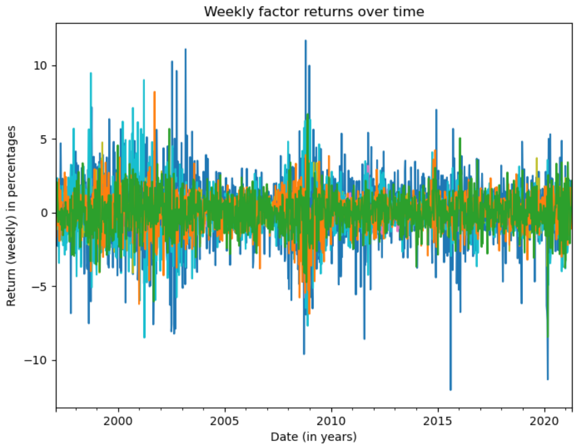
We use weekly data for the period of 4 February 1997 until 27 April 2021, consisting of 1265 datapoints. The sample period contains several events causing spikes of high volatility, namely the burst of the internet bubble (2000-2002), 9/11 (2001), the financial crisis of 2007-2008 and the COVID-19 crisis. As such, the various models are tested during both low- and high-volatility periods. We consider weekly data due to the results of Merton (1980), that argues that weekly estimation is more accurate than monthly.

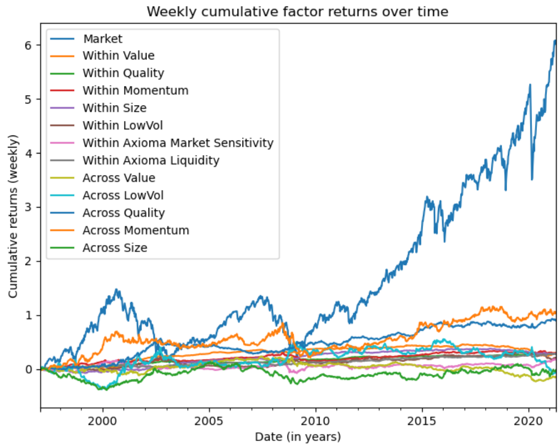
As we are interested in estimating the factor covariance matrix, we directly use factor-specific return data. This implies that the total returns of the assets are formed into returns per factor. The dataset consists of 42 factor mimicking portfolio (FMP) returns, where the factors are orthogonalised and neutral cross-sectionally towards the other factors.

We describe the characteristics of the 13 most important FMPs in this section and give the characteristics of the whole dataset of 42 FMPs, including country and industry FMPs, in Appendix B. The 13 considered FMPs are Market, Within Axioma Liquidity, Within Axioma Market Sensitivity, Within Value, Within Momentum, Within Size, Within Quality, Within LowVol, Across Value, Across Momentum, Across Size, Across Quality and Across LowVol. The Market FMP characterises the market without the influence of any factor. The next two FMPs are given by Axioma, while the others are constructed based on the characteristics in Appendix A.

The "Within" factors are asset-specific, such that their exposure to different industries and regions is neutral. The "Across" factors capture the part of the factor return that is industry-specific. Observe that especially the Market and the Within Market Sensitivity attain large positive and negative weekly return values, indicating that the variances of these FMPs are likely to be the highest. Note that the Within and Across FMPs rebalance every week, depending on the scores of the variables per factor for each asset. Therefore, the risk of an FMP depends on the asset decomposition within the FMP, which changes every week. This weekly change results in differing FMP characteristics and causes the volatility of the FMPs to be more dynamic over time. These 10 FMPs are thus not comparable to a regular asset which attains the same characteristics each period. The industry and country FMPs are constructed by dummies that indicate whether a certain asset belongs to a specific industry or country. The weights of the assets within these FMPs do not change. As such these behave more like regular assets.

Figure 1a shows the weekly returns of the 13 factors in percentages and Figure 1b depicts the cumulative returns. The first shows the spikes and clustering in the returns, indicating that they are dependent over time and that their dependence can change during certain stages in the economy. This implies that the volatilities are time-varying, making it important to apply a covariance model that considers past information, such as a DCC model. High volatility is visible during the financial crisis and internet bubble burst periods as returns alternate strongly between high and low values. The latter figure shows that especially the Market FMP has had high positive returns over the last few years. Moreover, observe that the Across Momentum FMP performs badly during the financial crisis while performing well in times of a stable economy. This shows that the Across Momentum FMP follows the sentiment of the market.





(a) Weekly FMP returns in percentages.

(b) Geometric cumulative FMP returns.

Figure 1: Weekly FMP returns over time period of 4 February 1997 up until 27 April 2021 for 13 FMPs.

The characteristics of the 13 FMP returns are shown in Table 1. First of all, the mean returns are all relatively close to zero. We observe some small differences between the factors. For example, the Across Value FMP attains a small negative average weekly return, while the Across Momentum has a small positive average weekly return. Overall, the differences between the average weekly FMP returns are not significant and close to zero. However, the standard deviations differ quite a lot for the various FMPs. Note that the most volatile FMPs are the Across FMPs, particularly Across Momentum, Across LowVol and Across Size. The kurtosis is larger than three for all FMPs, especially for the Within Axioma market Sensitivity, with a kurtosis value of 14.204. This shows that the FMPs are not normally distributed but follow a heavy-tailed distribution, signalling the need for the DCC models. Furthermore, the skewness is negative for most factors, indicating that declines more often occur than increases. Table B2 shows the statistics of the correlations of the FMP returns. The mean and median are relatively low, as high minimum and maximum values appear to cancel each other out. Do note that the mean and median are slightly negative, indicating that the FMPs are more often negatively correlated. Additionally, the high min and max values imply that correlations between FMPs can not be ignored and shows the need for multivariate models.

Table 1: Descriptive statistics of the FMP return for period 4 February 1997 until 27 April 2021.

	Mean	SD	SR	Kurtosis	Skewness
Market	0.183	2.365	0.077	5.914	-0.371
Within Value	0.021	0.272	0.077	5.849	0.418
Within Quality	0.020	0.180	0.112	7.504	0.312
Within Momentum	0.021	0.360	0.058	8.389	-1.108
Within Size	0.020	0.326	0.063	3.366	-0.268
Within LowVol	0.016	0.520	0.031	5.253	-0.082
Within Axioma Market Sensitivity	0.015	0.649	0.024	14.161	-0.038
Within Axioma Liquidity	0.019	0.276	0.070	5.273	-0.020
Across Value	-0.009	0.868	-0.010	6.884	0.071
Across LowVol	0.010	1.697	0.006	6.460	0.283
Across Quality	0.053	0.675	0.078	6.765	-0.220
Across Momentum	0.065	1.280	0.051	6.918	-0.338
Across Size	0.007	1.294	0.005	5.689	-0.144

Note. This table gives the statistical characteristics of the 13 most important FMPs in the dataset in percentages (Mean and SD).

5 Simulation

We follow a simulation approach to evaluate and test the covariance models in a controlled setting. The main goal is to examine the statistical behaviour of the estimators, assess the performance of various methods and test robustness by the impact of different covariance distributions on the accuracy of the methods. We first discuss the simulation setup in more detail. The results of the simulations are given in the following two subsections where S indicates the linear shrinkage towards the identity matrix, the CC model shrinks the estimate towards the constant correlation matrix and the 1F linearly shrinks the covariance estimate towards the one-factor model. Lastly, NL depicts the QIS method of Ledoit & Wolf (2022b). For shrinkage without the DCC models, we use the EWM as the covariance estimate, in line with PGGM. The parameters of the DCC models in the estimation of the simulated data are stated in Appendix C.

5.1 Simulation setup

In this simulation setup, we use Monte-Carlo simulations and generate paths for factor returns with a covariance defined beforehand such that the characteristics of the returns are known ex-ante. As simulation models, we consider the EWM and DCC models combined with a normal or t-distribution to obtain, what we assume to be, the real covariance of the simulation. Afterwards, we apply the estimation models discussed in Section 3: the sample covariance, EWM, DCC, DECO, ADCC, GOGARCH, and the linearly and nonlinearly shrunk versions, to fit the simulated factor returns. This verifies whether models perform well under a "false" distribution for the covariance matrix. The simulation is based on the factor return data of PGGM discussed in Section 4. We follow the approach of Fan et al. (2013) that split the procedure into a calibration and simulation part.

Calibration part

1. We consider the first 550 weekly observations $\{f_t\}_{t=1}^{550}$ of the K factors. This spans the period 4 February 1997 until 14 August 2007.

2. The calibrations are slightly different for the EWM and DCC models. For the EWM simulation model:

   (a) Estimate the mean μ_F of the factor returns over time. This results in a $K \times 1$ vector.

   (b) Estimate the covariance of the weekly factor return data $\Sigma_{0,F}$. This $K \times K$ covariance matrix is the starting point of the iterative procedure we conduct for the EWM covariances.

3. For the DCC simulations, we follow the structure outlined in Equations (8), (9) and (10). We require estimating some parameters first:

   (a) Estimate the covariance of the weekly return data. This $K \times K$ covariance matrix $\Sigma_{0,F}$ is the starting point of the iterative procedure of the covariance matrix $\Sigma_{t,F}$.

   (b) Fit the DCC model for the 550 observations to obtain an estimate for the DCC parameters (α, β).

   (c) We estimate the mean μ_ε, covariance $\Sigma_{0,\varepsilon}$ and unconditional correlation matrix $S = \frac{1}{T}\sum_{t-1}^{T} \varepsilon_t' \varepsilon_t$ of the standardised returns ε_t over the 550 observations.

Simulation part

For each factor $k = 1, \ldots K$, we generate T^{sim} returns per factor by applying the iterative nature of both models. We rerun the simulation 100 times to provide reliable simulation results.

1. If simulating from an EWM model, we assume that the covariance follows $\Sigma_{t,F} = \lambda \Sigma_{t-1} + (1-\lambda) f_{t-1} f_{t-1}'$:

   (a) Generate the first return series f_0 by the normal distribution $N_K(\mu_F, \Sigma_{0,F})$ or student's t-distribution $t_K(\mu_F, \Sigma_{0,F}, 5)$.

   (b) Initialise the iterative procedure by Σ_0 and f_0 to obtain the covariance Σ_t for $t = 1$. We set λ to obtain a half-life of 52 weeks for the covariances.

   (c) Simulate the factor returns f_t independently from $N_K(\mu, \Sigma_{t,F})$ or $t_K(\mu, \Sigma_{t,F}, 5)$ for $t = 1$.

   (d) Repeat the iterative procedure for $t = 1, ..., T^{sim}$.

2. If simulating from a DCC model, we utilise its iterative nature, described in Section 3.2:

   (a) Generate the standardised returns ε_t by the normal distribution as $N_K(\mu_\varepsilon, \Sigma_{0,\varepsilon})$.

   (b) Take the calibrated values of S, α and β together with the standardised returns to initialise Equation (10) and obtain Q_t for $t = 1$. Then calculate $\Sigma_{t,F}$ via Equations (8) and (9). For Q_0 we use the covariance matrix $\Sigma_{0,F}$ as initialisation.

   (c) Repeat steps a) and b) to obtain covariance estimates $\Sigma_{t,F}$ for $t = 1, \ldots, T^{sim}$.

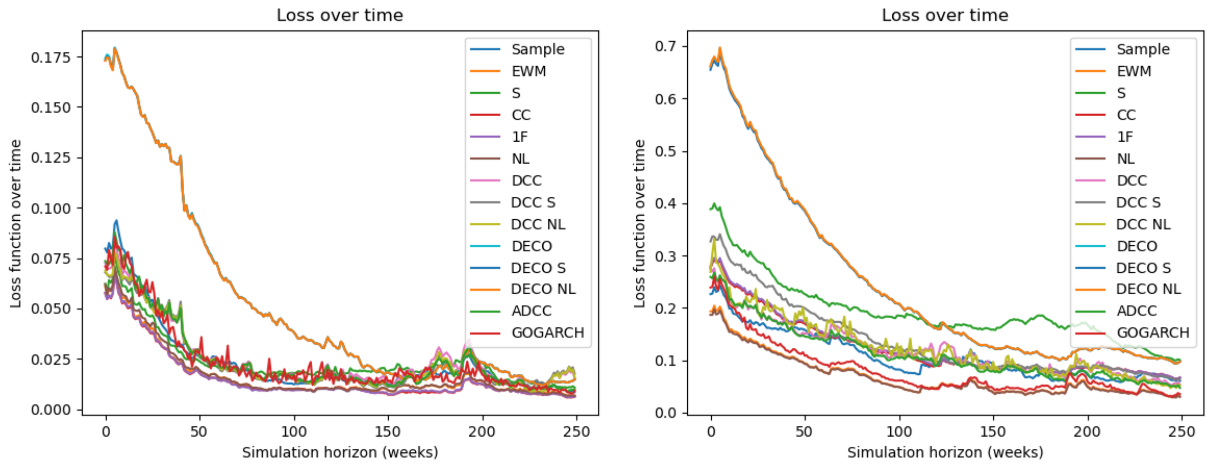
(d) Simulate the factor returns f_t independently from $N_K(\mu, \Sigma_{t,F})$ or $t_K(\mu, \Sigma_{t,F}, 5)$ for $t = 1, \ldots, T^{sim}$.

Afterwards, we estimate the simulated data with the considered estimation models. To evaluate small and large sample properties and check for robustness, we evaluate the performance for different rolling window sizes, namely 150, 300 and 2000 datapoints. We aim to obtain 250 covariance estimates for each rolling window, so when taking into account the different estimation windows, we simulate $T^{sim} = \{300, 550, 2250\}$ weekly returns. In the remaining sections, we consider a rolling window of 150 weeks. Appendix E discusses the other two windows.

5.2 Direct evaluation

We conduct a direct evaluation by calculating the mean-variance loss function of Engle et al. (2019), described in Equation (A.7). The results for the normal simulation with EMW covariance for $K = 13$ and $K = 42$ are shown in Figures 2a and 2b for a loss averaged over 100 iterations. Both figures show that the DECO estimators perform the worst, while the shrinkage methods for the EWM model perform the best, with all average losses relatively close together. Moreover, it is striking that the DCC model also performs well, while the related DECO model does not. The loss over time is larger for $K = 42$, confirming that the accuracy of the models diminishes and estimation becomes more difficult and noisy as K increases.

For $K = 42$, we observe more spikes in the loss over time and alternating optimal estimators. This indicates that capturing the covariance of 42 FMPs is not straightforward. The large average loss during most of the simulation period for the DECO models is potentially caused by the estimation window being too short and therefore not capturing enough information. Another reason could be model misspecification. Specifically, that the assumption of constant correlations over time that the DECO model imposes is not fitting for this dataset. We have conducted similar analyses for the t-distribution with EWM covariance and the normal and t-distributions with DCC covariance. The figures are stated in Appendix C and show similar outcomes.



(a) 13 FMPs

(b) 42 FMPs

Figure 2: Average mean-variance loss function of Engle et al. (2019) of the 250 simulated covariances for the estimation models for a rolling window of 150 weekly observations with simulation data by the normal distribution of the EWM covariance for 100 iterations.

We compare the loss averaged over the number of simulations and over time. Table 2 shows the outcomes for $K = 13$ where the best-performing model, which gives the lowest loss value, is indicated in bold. We observe that the CC model performs best for all simulations, indicating its robustness to misspecification. The DCC models perform worse than the static shrinkage estimators, especially the DECO models perform poorly. The best dynamic model is the DCC S, being the most accurate in three out of four simulations. Note that the DCC estimators perform better for the DCC than the EWM simulations, but still do not outperform most EWM shrinkage methods.

Moreover, Table 2 shows that the shrinkage methods are effective and reduce the loss compared to their non-shrunk counterparts. The only exception is the S model, which performs worse than the EWM model for all simulations. This signals that for shrinking to the identity matrix, the increase in bias outweighs the decrease in variance. Strikingly, shrinking linearly yields slightly better outcomes than nonlinearly in all cases. A reason could be that the dimension of the covariances is not large enough and the differences between eigenvalues are not dispersed enough such that assigning different weights to each eigenvalue does not add value. A nonlinear structure would in that case only increase complexity and noise.

Next to that, Table C3 in Appendix C shows the improvement in percentages of the loss function compared to the current method of PGGM, the EWM with linear shrinkage (S). This table confirms earlier observations and indicates room for improvement in the covariance estimation compared to the current method.

Table 2: Average loss for four simulation setups for a rolling window of 150 weeks and 13 FMPs.

	Benchmarks		EWM shrinkage				DCC models							
	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
EWM normal	0.024	0.017	0.023	**0.015**	0.016	0.017	0.025	0.022	0.023	0.052	0.052	0.051	0.024	0.025
EWM t-dist	0.030	0.022	0.035	**0.019**	0.019	0.021	0.032	0.026	0.029	0.055	0.054	0.054	0.030	0.030
DCC normal	0.029	0.024	0.030	**0.022**	0.022	0.023	0.029	0.028	0.028	0.053	0.053	0.052	0.028	0.031
DCC t-dist	0.034	0.028	0.040	**0.025**	0.025	0.027	0.034	0.029	0.032	0.054	0.054	0.054	0.033	0.035

Note. Lowest average loss defined by the loss function of Engle et al. (2019) is indicated in bold.
The loss is averaged over the 250 losses from the covariance estimates and 100 simulations for each estimator. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns.

Table 3 gives the average losses for the covariance estimates for 42 FMPs. We observe a shift in the best-performing model from CC to NL consistent over all simulations. The bad performance of S, especially its underperformance compared to the EWM and the sample covariance, is in line with $K = 13$. Similarly to $K = 13$, the best dynamic model depends on the simulation; the GOGARCH estimator performs the most accurately for the EWM simulations, while the DCC NL is the best when considering a DCC covariance simulation.

Compared to $K = 13$, we now find that nonlinear shrinkage leads to more accurate results than its linear counterparts for both EWM and DCC. This is in line with literature, where Ledoit & Wolf (2022b) state that nonlinear shrinkage outperforms linear shrinkage when K becomes large compared to the sample size. The PRIAL results are given in Table C4 and are again in line with the average loss discussed in Table 3. It shows that even bigger improvements can be made compared to $K = 13$.

Table 3: Average loss for four simulation setups for a rolling window of 150 weeks and 42 FMPs.

	Benchmarks		EWM shrinkage				DCC models							
	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
EWM normal	0.127	0.067	0.191	0.121	0.122	**0.065**	0.121	0.137	0.124	0.240	0.240	0.240	0.109	0.071
EWM t-dist	0.131	0.087	0.262	0.136	0.138	**0.084**	0.128	0.155	0.132	0.246	0.244	0.246	0.120	0.104
DCC normal	0.123	0.106	0.198	0.111	0.114	**0.102**	0.124	0.117	0.117	0.203	0.203	0.203	0.120	0.127
DCC t-dist	0.146	0.119	0.269	0.122	0.126	**0.114**	0.143	0.134	0.130	0.213	0.212	0.213	0.138	0.143

Note. Lowest average loss defined by the loss function of Engle et al. (2019) is indicated in bold.
The loss is averaged over the 250 losses from the covariance estimates and 100 simulations for each estimator. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns.

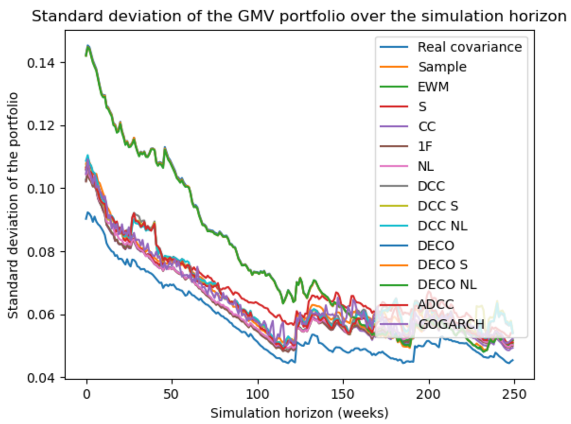
5.3 Indirect evaluation

5.3.1 Global minimum variance portfolio

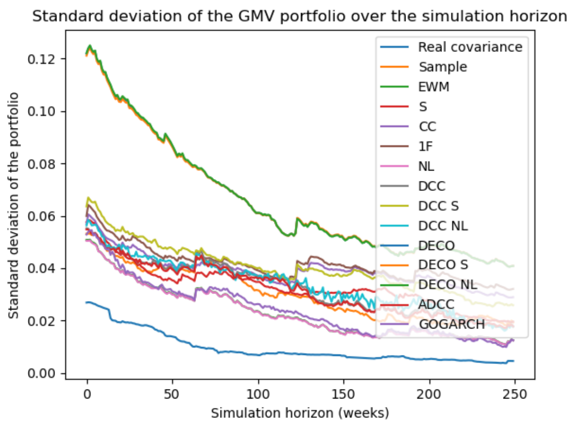
Next to the direct evaluation, we perform an indirect evaluation of the conditional covariance estimates by forming global minimum variance (GMV) portfolios. Firstly, we determine the optimal weights for each estimation method that minimises the risk of the portfolio for each time period. We show the realised volatility over time averaged over the simulations and evaluate the performance of the different GMV portfolios based on their average annualised standard deviation (SD), Sharpe ratio (SR), information ratio (IR) and turnover. The SDs of the GMV portfolios over time are calculated as $SD = \sqrt{W_t' \Sigma_t W_t}$, and the SR is defined as $SR = \frac{W_t' \mu}{\sqrt{W_t' \Sigma_t W_t}}$. We plug in the population covariance of the simulation for Σ_t, for μ we use the mean we have simulated from, and we consider the portfolio weights obtained by the portfolio optimisation for the different covariance estimators. Note that the turnover is defined as the weight changes in the different FMPs within the portfolios and not of the assets used to construct the FMPs.

Figures 3a and 3b show the average realised volatility of the GMV portfolios over time for the normal distribution with EWM covariance simulation for $K = 13$ and $K = 42$, respectively. The results are in line with the direct evaluation. Logically, the GMV portfolio based on the real covariance gives the best results, whereafter the linear shrinkage models appear to perform the best for $K = 13$ and the static nonlinear shrinkage and GOGARCH for $K = 42$. The DECO models perform the worst in both dimensions.

The differences between the real standard deviation and those produced by the estimation models are larger for $K = 42$ than for $K = 13$. This higher inaccuracy in larger dimensions is consistent with the direct evaluation method and literature. Note that the standard deviation is higher for $K = 13$, on average in the range of 0.09 and 0.12, while the standard deviation for $K = 42$ is between 0.04 and 0.07. This can be explained by the fact that more FMPs allow the implementation of more diversification by choosing between more investment options. This diversification leads to a risk reduction, causing the lower standard deviation. The results for the normal and t-distribution with the DCC covariance and t-distribution with EWM covariance are stated in Appendix C. We find similar patterns for all of them; the DECO models perform poorly, and the EWM shrinkage models work relatively well.



(a) 13 FMPs



(b) 42 FMPs

Figure 3: Average volatility of the GMV portfolios based on simulation data of 250 weekly returns and 100 iterations by the normal distribution simulation of the EWM covariance for a rolling window of 150 weekly observations.

The results for the average annualised portfolio performance are given in Table 4 for $K = 13$ and Table 5 for $K = 42$. The lowest SD is indicated in bold.

Overall, the results are consistent with the direct evaluation in both dimensions. The results indicate that the simpler static models perform better compared to the dynamic models, for both $K = 13$ and $K = 42$.

Next to that, we note that the 1/N portfolio performs significantly worse compared to the other portfolios, attaining an SD value at least three times as large as the real GMV portfolio and a lower SR. These observations are in line with Engle et al. (2019) and Jagannathan & Ma (2003). The latter argue that GMV portfolios have desirable properties, not only in terms of risk but also reward-to-risk (SR).

In more detail, we find that the CC performs the best in most cases for 13 FMPs, in line with Table 2. Note that for the t-dist EWM simulation, the 1F model performs the best in the indirect evaluation, while the CC model attains the lowest loss in the direct approach. This suggests that the accuracy of the covariance estimator is not the only factor that influences portfolio performance.

Moreover, Table 4 shows that the static models widely outperform the best dynamic estimator, the DCC S. We also observe that the DECO model performs the worst in all simulations. This suggests that the assumption of equal pairwise correlations used to convert the DCC model into the DECO is not beneficial (Engle & Kelly, 2012). In terms of turnover, the static models attain lower turnover than the dynamic models. This indicates that the positions are adjusted much smoother.

Table 5 shows the performance of the GMV portfolios in a 42 FMP dimension. The results are again in line with the direct evaluation; the NL model performs the best and DECO the worst in terms of SD.

The performance of the dynamic models attains a similar behaviour as for the direct analysis. The best-performing dynamic estimator in all simulations is the GOGARCH. However, the dynamic models still underperform compared to the static estimators similarly to the results for 13 FMPs. Note that the turnover of the GOGARCH estimator is relatively high, indicating that the weights between time periods change a lot. This is unattractive, as most portfolio managers are restricted to a certain percentage of turnover per year.

Additionally, the results in Tables 4 and 5 suggest that the shrinkage methods, especially linear shrinkage, help lower the standard deviation and often lead to lower turnover compared to the versions without shrinkage.

We find that by increasing the dimension K, the optimal standard deviation lowers. However, the estimates are further away from the optimum, indicating the difficulty of forecasting the covariance for large K. Moreover, the results show that in larger dimensions, more complex models such as the NL, ADCC and GOGARCH become necessary to capture relevant covariance dynamics and perform better than linear shrinkage and the sample covariance. The latter performs poorly because the quality of the sample covariance matrix rapidly deteriorates as the concentration ratio K/T rises (Ledoit & Wolf, 2004a). Note that the NL model also performs relatively well in the smaller dimension.

Table 4: Average annualised performance of the GMV portfolios for rolling window of 150 weeks and 13 FMPs.

		Benchmarks				EWM shrinkage					DCC models				
Normal EWM	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.411	1.642	0.487	0.464	0.482	**0.455**	0.459	0.464	0.489	0.486	0.486	0.571	0.570	0.568	0.486	0.477
SR	4.053	1.410	3.430	3.585	3.423	3.675	3.628	3.565	3.323	3.299	3.307	2.837	2.833	2.850	3.309	3.475
IR	-	5.205	-6.573	-7.675	-4.660	-9.588	-8.947	-7.623	-12.652	-13.435	-13.975	-9.902	-10.120	-10.079	-12.760	-6.983
Turnover	1.424	0.312	1.530	1.396	0.958	1.170	1.194	1.446	2.739	2.920	2.793	2.196	2.226	2.209	2.781	9.952
***t*-dist EWM**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.411	1.642	0.480	0.458	0.497	0.456	**0.454**	0.459	0.475	0.475	0.473	0.591	0.589	0.589	0.472	0.477
SR	4.053	1.410	3.690	3.790	3.421	3.783	3.771	3.755	3.697	3.643	3.678	2.923	2.933	2.931	3.688	3.615
IR	-	-0.124	1.526	-1.182	-1.600	-3.220	-3.793	-1.671	2.304	-0.252	1.094	-4.283	-4.181	-4.211	2.919	-1.862
Turnover	1.424	0.312	1.567	1.274	0.901	1.077	1.028	1.314	2.506	2.289	2.515	2.141	2.136	2.140	2.944	8.821
Normal DCC	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.357	1.549	0.453	0.441	0.447	**0.428**	0.434	0.437	0.460	0.458	0.457	0.517	0.515	0.515	0.449	0.459
SR	4.618	1.494	3.921	3.992	3.782	4.087	4.029	3.996	3.630	3.503	3.574	3.077	3.082	3.084	3.778	3.888
IR	-	3.217	2.353	1.220	-1.608	1.433	1.003	0.342	-2.014	-4.118	-3.602	-3.809	-3.763	-3.645	-2.854	1.016
Turnover	6.136	0.416	1.568	1.353	0.943	1.134	1.146	1.425	5.057	4.933	5.037	2.786	2.756	2.784	3.423	9.595
***t*-dist DCC**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.357	1.549	0.468	0.453	0.482	**0.445**	0.448	0.452	0.470	0.462	0.468	0.556	0.555	0.556	0.472	0.469
SR	4.618	1.494	3.558	3.688	3.412	3.772	3.729	3.705	3.526	3.587	3.549	2.998	3.001	2.996	3.490	3.556
IR	-	7.717	-1.882	-2.307	1.908	-2.623	-3.443	-1.537	-0.288	0.364	0.139	-0.297	-0.346	-0.300	-1.020	-2.692
Turnover	6.136	0.416	1.684	1.351	0.922	1.138	1.106	1.433	3.251	3.089	3.241	2.791	2.796	2.781	3.603	9.925

Note. The SDs of the GMV portfolios over time are calculated as $SD = \sqrt{W_t' \Sigma_t W_t}$, and the SR is defined as $SR = \frac{W_t' \mu}{\sqrt{W_t' \Sigma_t W_t}}$. The SD is given in percentages.

The GMV portfolio performance is based on 250 weekly simulated observations and results are annualised by multiplying by 52.

The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns.

Table 5: Average annualised performance of the GMV portfolios for $RW = 150$ and $K = 42$.

		Benchmarks			EWM shrinkage				DCC models							
Normal EWM	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.068	0.738	0.262	0.185	0.231	0.288	0.306	**0.183**	0.254	0.298	0.266	0.472	0.472	0.472	0.243	0.186
SR	6.934	1.098	1.789	2.468	2.926	4.011	3.942	2.511	1.878	3.519	2.137	2.585	2.611	2.555	2.330	2.386
IR	-	6.967	2.501	-0.213	8.554	22.399	21.558	-0.145	4.076	20.410	7.351	13.002	13.184	12.775	6.601	-0.798
Turnover	1.465	0.364	2.058	1.704	0.639	1.182	1.116	1.727	2.792	2.341	16.494	2.003	2.038	1.970	4.154	4.781
***t*-dist EWM**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.068	0.738	0.235	0.180	0.295	0.306	0.323	**0.179**	0.238	0.340	0.253	0.495	0.496	0.495	0.230	0.190
SR	6.934	1.098	2.669	3.383	2.482	3.905	3.960	3.376	2.500	3.455	2.842	2.431	2.488	2.426	2.752	3.137
IR	-	2.139	2.572	3.206	2.511	10.764	11.754	3.131	0.418	7.837	1.145	5.184	5.482	5.114	0.936	3.920
Turnover	1.465	0.364	2.214	1.807	0.637	1.162	1.075	1.784	2.901	2.436	13.570	1.968	2.060	1.995	3.577	4.969
Normal DCC	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.089	0.727	0.265	0.230	0.255	0.289	0.308	**0.225**	0.264	0.279	0.261	0.437	0.436	0.437	0.258	0.250
SR	5.575	1.116	2.800	3.074	2.742	4.177	4.048	3.095	2.675	3.874	3.052	2.817	2.848	2.793	2.810	2.734
IR	-	5.791	9.839	9.415	8.239	19.886	20.824	9.437	10.466	19.038	12.938	13.741	13.992	13.537	11.868	8.495
Turnover	3.783	0.468	2.338	2.039	0.709	1.267	1.164	2.048	3.517	3.602	9.895	2.632	2.660	2.599	4.240	6.238
***t*-dist DCC**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.089	0.727	0.277	0.244	0.319	0.315	0.330	**0.242**	0.273	0.324	0.273	0.459	0.459	0.459	0.267	0.262
SR	5.575	1.116	2.250	2.503	2.252	3.714	3.725	2.516	2.207	3.371	2.639	2.522	2.572	2.509	2.420	2.334
IR	-	1.692	-2.641	-3.558	2.593	8.960	10.406	-3.060	-6.077	5.172	-3.977	3.537	3.757	3.456	-5.138	-1.295
Turnover	3.783	0.468	2.284	1.925	0.698	1.254	1.139	1.942	3.614	3.342	12.364	2.800	2.888	2.807	4.388	5.348

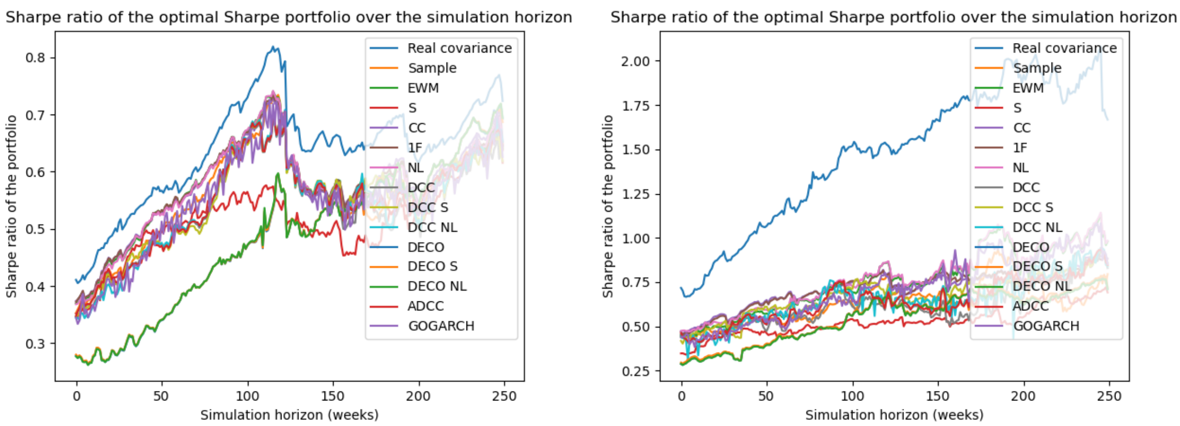
Note. The SDs of the GMV portfolios over time are calculated as $SD = \sqrt{W_t' \Sigma_t W_t}$, and the SR is defined as $SR = \frac{W_t' \mu}{\sqrt{W_t' \Sigma_t W_t}}$. The SD is given in percentages.

The GMV portfolio performance is based on 250 weekly simulated observations and results are annualised by multiplying by 52.
The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns.

5.3.2 Sharpe ratio portfolios

Next to the performance of the GMV portfolios, we evaluate how the models perform when the aim is to maximise the Sharpe ratio (SR). The SR portfolios try to find the highest expected return combined with the lowest standard deviation possible. Such application is important to consider as portfolio managers aim to find the maximum return for a predetermined amount of risk. We calculate the performance measures similarly as for the GMV portfolios. The mean is calculated by $W_t'\mu$, SDs as $SD = \sqrt{W_t'\Sigma_t W_t}$, and the SR is defined as $SR = \frac{W_t'\mu}{\sqrt{W_t'\Sigma_t W_t}}$. We plug in the population covariance of the simulation for Σ_t, for μ we use the mean we have simulated from, and we consider the portfolio weights obtained by the portfolio optimisation using the different covariance estimators. We use the population mean in the optimisation to avoid optimising over an inaccurate sample mean. Especially Jagannathan & Ma (2003) notes that as dimensions grow, the sample mean is often estimated with much error, such that these outcomes can not be trusted. This is also why we only consider the SR portfolios in a simulation setting where the sample mean is known and not an empirical backtest.

Figures 4a and 4b show the SR over time for $K = 13$ and $K = 42$, respectively. First of all, one can observe that the DECO models attain the lowest SRs for both dimensions. Additionally, the S estimator does not seem to perform very well for $K = 42$, showing low SRs. Many models move closely together, such that we require further analysis to determine the best-performing model. Do note that the difference between the optimal SR, derived by the real covariance, and those from the models are significantly closer for $K = 13$ than $K = 42$. This again signals the difficulty of estimating the covariance matrix in larger dimensions. Moreover, the SR values are almost twice as high when considering more FMPs, showing the diversification opportunities that lower the standard deviation when enlarging the number of FMPs.



(a) 13 FMPs

(b) 42 FMPs

Figure 4: Average Sharpe ratio of the SR portfolios based on simulation data of 250 weekly returns and 100 iterations by the normal distribution simulation and the EWM covariance for a rolling window of 150 weekly observations.

Tables 6 and 7 show the SR portfolio performance for both covariance dimensions where the lowest Sharpe ratios are made bold. The tables aim to give a clearer image of the best- and worst-performing models concerning the figures above.

We find that the best-performing model in both dimensions is the CC model. Worth to consider is that the SR portfolios do not necessarily produce this kind of performance in practice due to the additional estimation error in the expected returns that is not taken into account in this simulation approach.

Table 6 considers the evaluation of the SR portfolios for 13 FMPs. The top performance of the CC estimator is similar to the GMV results for $K = 13$, where the CC model performs best in three out of four simulations. The best-performing dynamic estimator depends on the simulation. For example, the GOGARCH performs well for the normal distribution simulations. This is unexpected, as the GOGARCH was not among the top performers for the direct and GMV evaluation.

We find that the SRs of the SR portfolios are higher than for the GMV portfolios, indicating that adding the mean in the optimisation adds value. The mean should thus be considered in theory when finding the best risk-return trade-off. Observe that the improvements in SR are often small.

Table 7 gives the results for the dimension of $K = 42$. The top estimators differ greatly when comparing them to the GMV outcomes. Similarly to $K = 13$, the best model in terms of SR is the CC in three of the four simulations. Strikingly, the NL method that is the most accurate in the GMV portfolio and direct evaluation is never the best in this setup. Moreover, the linear shrinkage methods now perform better than the nonlinear and dynamic models, in contrast to the GMV portfolio results. The best dynamic model is the DCC S in all simulations, while the GOGARCH had the best results for the GMV evaluation.

A possible explanation for the deviations when comparing GMV to SR optimal portfolios is that the SR portfolios not only consider low risk measured by the covariance matrix but also want high returns expressed by the expected returns. As such, more weight is put on the assets with higher expected returns, lowering the focus on the risks involved. This reduces the sensitivity to the accuracy of the covariance matrix and could lead to different optimal estimators.

Lastly, we again observe that estimators are further away from the optimal SR when moving to the larger dimension of 42 FMPs. This observation again signals the difficulties that arise when moving to a high-dimensional dataset.

Table 6: Average annualised performance of the SR portfolios for a rolling window of 150 weeks and 13 FMPs.

		Benchmarks			EWM shrinkage				DCC models						
Normal EWM	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Mean	2.023	2.315	2.041	2.037	2.181	2.034	2.030	2.047	2.012	2.018	2.016	2.029	2.036	2.036	2.011	2.036
SD	0.452	1.642	0.527	0.506	0.565	0.499	0.502	0.507	0.530	0.526	0.525	0.633	0.633	0.632	0.526	0.520
SR	4.475	1.410	3.874	4.026	3.862	**4.076**	4.046	4.038	3.800	3.835	3.839	3.208	3.216	3.222	3.826	3.914
IR	-	4.254	-4.341	-3.743	1.727	-2.646	-3.271	-2.837	-8.787	-8.815	-9.032	-4.811	-4.835	-4.789	-8.122	-3.270
Turnover	1.637	0.312	1.644	1.536	1.133	1.250	1.299	1.570	3.006	3.228	3.103	2.360	2.402	2.378	3.109	10.889
***t*-dist EWM**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Mean	2.023	2.315	2.102	2.061	2.247	2.049	2.051	2.065	2.087	2.104	2.093	2.137	2.142	2.140	2.094	2.067
SD	0.452	1.642	0.543	0.517	0.624	0.504	0.505	0.518	0.538	0.533	0.533	0.667	0.666	0.666	0.536	0.539
SR	4.475	1.410	3.871	3.987	3.599	**4.064**	4.063	3.984	3.880	3.946	3.928	3.203	3.217	3.214	3.903	3.833
IR	-	-0.978	5.516	1.787	5.671	0.340	0.528	2.060	5.415	4.355	4.847	-1.894	-1.627	-1.662	6.484	2.072
Turnover	1.637	0.312	1.620	1.414	1.125	1.160	1.139	1.426	2.675	2.473	2.755	2.209	2.201	2.216	3.153	9.786
Normal DCC	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Mean	1.999	2.315	2.094	2.063	2.230	2.055	2.055	2.075	2.020	1.989	2.006	2.016	2.022	2.018	2.056	2.096
SD	0.392	1.549	0.513	0.494	0.546	0.480	0.487	0.492	0.516	0.504	0.508	0.576	0.577	0.577	0.513	0.516
SR	5.095	1.494	4.084	4.173	4.081	**4.280**	4.224	4.217	3.912	3.949	3.944	3.498	3.506	3.500	4.010	4.059
IR	-	1.512	0.652	-1.869	0.208	-0.985	-1.501	-2.091	-2.776	-4.153	-3.607	-2.871	-2.787	-2.676	-3.466	-1.467
Turnover	6.708	0.416	1.646	1.527	1.122	1.239	1.279	1.585	5.138	5.093	5.153	2.689	2.659	2.701	3.321	10.00
***t*-dist DCC**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Mean	1.999	2.315	2.035	2.041	2.221	2.031	2.037	2.064	2.021	2.050	2.031	2.089	2.091	2.087	2.025	2.056
SD	0.392	1.549	0.508	0.488	0.575	0.478	0.483	0.491	0.507	0.503	0.504	0.617	0.617	0.617	0.509	0.518
SR	5.095	1.494	4.008	4.184	3.865	**4.252**	4.218	4.200	3.990	4.077	4.027	3.386	3.391	3.383	3.982	3.969
IR	-	6.640	-0.805	-1.357	8.443	-0.137	-0.868	0.112	0.769	3.460	1.874	3.866	3.893	3.808	0.250	0.514
Turnover	6.708	0.416	1.796	1.476	1.104	1.187	1.178	1.573	3.355	3.140	3.330	2.802	2.805	2.795	3.785	10.739

Note. The mean of the SR portfolios over time is calculated as Mean=$W_t'\mu$, the SD as $SD = \sqrt{W_t' \Sigma_t W_t}$ and the SR is defined as $SR = \frac{W_t'\mu}{\sqrt{W_t'\Sigma_t W_t}}$. The Mean and SD are given in percentages.

The SR portfolio performance is based on 250 weekly simulated observations and results are annualised by multiplying by 52.
The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns.

Table 7: Average annualised performance of the SR portfolios for a rolling window of 150 weeks and 42 FMPs.

		Benchmarks			EWM shrinkage				DCC models							
Normal EWM	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Mean	0.811	0.811	1.447	1.350	3.439	2.260	2.481	1.458	1.624	2.523	1.839	2.313	2.360	2.261	1.909	1.435
SD	0.094	0.738	0.367	0.297	0.925	0.473	0.523	0.313	0.390	0.557	0.440	0.637	0.646	0.624	0.438	0.333
SR	8.610	1.098	3.940	4.553	3.717	**4.780**	4.747	4.653	4.161	4.529	4.184	3.630	3.652	3.626	4.363	4.312
IR	-	2.915	20.097	19.497	13.163	19.618	18.872	19.155	4.756	19.149	5.675	11.236	11.394	11.221	16.983	17.923
Turnover	2.148	0.364	3.869	3.106	2.655	1.720	1.769	3.261	5.531	3.493	27.173	2.383	2.426	2.366	6.776	9.400
***t*-dist EWM**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Mean	0.811	0.811	1.406	1.301	4.400	2.195	2.558	1.359	1.459	2.568	1.774	2.084	2.204	2.075	1.776	1.332
SD	0.094	0.738	0.360	0.289	1.348	0.462	0.538	0.298	0.367	0.586	0.422	0.588	0.611	0.585	0.430	0.316
SR	8.610	1.098	3.902	4.497	3.264	4.752	**4.753**	4.567	3.975	4.382	4.205	3.545	3.609	3.544	4.129	4.217
IR	-	-0.840	7.067	6.867	8.112	9.531	10.828	6.620	6.652	7.196	7.951	4.587	5.122	4.547	9.527	7.317
Turnover	2.148	0.364	3.367	2.826	2.768	1.516	1.517	2.857	4.575	3.529	21.757	2.428	2.567	2.474	6.325	8.528
Normal DCC	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Mean	1.136	0.811	1.919	1.888	3.750	2.284	2.492	2.000	1.982	2.522	2.128	2.316	2.367	2.274	2.219	1.875
SD	0.146	0.727	0.441	0.410	1.020	0.461	0.517	0.429	0.458	0.547	0.477	0.591	0.604	0.583	0.507	0.454
SR	7.804	1.116	4.354	4.610	3.678	**4.952**	4.819	4.663	4.324	4.614	4.460	3.919	3.918	3.903	4.375	4.127
IR	-	-1.663	5.474	4.182	6.805	9.383	9.993	4.595	5.022	11.364	7.998	8.337	8.520	8.047	6.403	1.136
Turnover	5.512	0.468	4.052	3.696	2.811	1.792	1.781	3.722	6.652	5.256	12.920	3.178	3.231	3.151	7.799	11.290
***t*-dist DCC**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Mean	1.136	0.811	1.697	1.612	4.439	2.218	2.547	1.708	1.749	2.569	1.939	2.166	2.264	2.141	1.934	1.686
SD	0.146	0.727	0.423	0.376	1.383	0.456	0.542	0.391	0.433	0.581	0.454	0.564	0.585	0.559	0.466	0.424
SR	7.804	1.116	4.008	4.291	3.209	**4.864**	4.700	4.372	4.042	4.421	4.271	3.840	3.873	3.828	4.147	3.972
IR	-	-7.556	7.278	9.415	14.065	4.558	8.495	10.903	6.372	3.725	4.887	-1.982	-1.329	-2.209	6.982	8.832
Turnover	5.512	0.468	3.937	3.350	2.845	1.676	1.721	3.345	6.506	4.896	14.829	3.436	3.572	3.472	7.646	9.629

Note. The mean of the SR portfolios over time is calculated as Mean=$W_t'\mu$, the SD as $SD = \sqrt{W_t'\Sigma_t W_t}$ and the SR is defined as $SR = \frac{W_t'\mu}{\sqrt{W_t'\Sigma_t W_t}}$. The Mean and SD are given in percentages.

The SR portfolio performance is based on 250 weekly simulated observations and results are annualised by multiplying by 52.
The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns.

6 Empirical results

This section discusses the performance and the summary statistics of the weights of the global minimum variance (GMV) portfolio and the GMV portfolio including short-selling constraints. The weight statistics indicate how realistic constructed portfolios are in practice. We do not consider the maximum Sharpe ratio (SR) portfolios because the literature shows that the results in practice are not accurate and reliable due estimation error in the average returns (Jagannathan & Ma, 2003).

We consider a rolling window of 150 weekly observations and use a weekly rebalancing, consistent with the simulation study. We assess the performance of the GMV portfolios by the mean, standard deviation (SD), Sharpe ratio (SR) and turnover of the portfolio returns. Note that for the empirical analysis, we can only calculate the mean, SD and SR in practice by taking the statistics of the obtained portfolio returns rather than in theory like the formulations described in Section 5. The parameters of the estimation models are given in Appendix D.

Next to the GMV portfolio analyses, we conduct a high- versus low-volatility period comparison of which Appendix E gives the setup and outcomes. We find that the best estimators of the high- and low-volatility period GMV portfolios are consistent with the results for the complete dataset and that short-selling constraints are particularly beneficial in low-volatility periods. Returns are often positive and stable in these periods, such that positive weights are favourable. In high-volatility periods, which indicate periods of crisis, short-selling would be helpful to hedge against the risk of negative returns.

6.1 Global minimum variance portfolio

The performance metrics for the GMV portfolios are given in Table 8 for 13 and 42 FMPs, respectively. The results are consistent with the simulations; for $K = 13$, the CC model shows the best results, while the NL estimator attains the lowest SD for the dimension of 42 FMPs.

Table 8 shows that for $K = 13$, the best-performing model in terms of SD is the CC. The outperformance of the CC compared to S is significant. Strikingly, the S estimator performs worse than the EWM and the sample covariance models. Thus, applying shrinkage to the identity matrix only adds bias without resulting in lower variance.

In terms of dynamic model performance, we find that the ADCC model is the best-performing dynamic estimator, in contradiction to the simulation results where the DCC S model is the most accurate. This shows that the simulations can indicate the performance of the models, but can only partly imitate the real data and thus predict their results. The GOGARCH model is the worst performer in the 13-FMP dimension. This could be because the GOGARCH model creates a factor structure of the dataset, but in this case, we are already considering an FMP dataset instead of a large dataset of assets. As such, there is no room for additional factorisation of the FMPs and imposing a factor structure on the FMP return dataset only results in more noise, loss of important information and a higher SD.

When considering the full dimension of the dataset ($K = 42$), Table 8 shows that the NL model gives the lowest SD. Surprisingly, the second-best model is the EWM model. Moreover, we observe that the GOGARCH has the lowest SD of all the dynamic models, consistent with the outcomes for the simulations. Similarly to the simulations, the dynamic models attain higher turnover values in both covariance dimensions.

Table D2 shows the characteristics (max, min, median, SD) of the average weights of the GMV portfolios for both 13 and 42 dimensions. These characteristics can be used to explain turnover values in more detail. In the dimension of 13 FMPs, the high maximum and low minimum weights of the sample covariance and dynamic models indicate instability and imply that investors have to take large long and short positions in the 13 FMPs. This is undesirable because of transaction costs. The SD indicates how much the weights differentiate over the considered time horizon. We find that the dynamic models, such as the GOGARCH, and the sample covariance have the highest SD. This is again undesirable as it leads to much turnover and thus high transaction costs to attain the optimal GMV portfolio weights over time. The high min and max values, combined with the high SD, result in the high turnover values in Table 8 for the dynamic models.

For 42 FMPs, Table D2 shows lower max values and SDs, and higher min values. This because increasing the dimension causes the weights to be spread among more factors, resulting in a lower weight per factor on average. These observations can also be deducted from the turnover, as for most estimators, the dimension of 13 FMPs attains higher turnover compared to the 42-FMP dimension. Altogether, it signals the increase in diversification possibilities when the factor space grows. Similarly to the smaller dimension, we observe higher SD values for the dynamic models than the static methods, consistent with the higher turnover values. Strikingly, the DECO models attain the highest SDs, but do not show the highest turnover in Table 8.

6.2 Short-selling constraints

Since negative portfolio weights are difficult to implement in practice, most investors, including PGGM, impose the constraint that portfolio weights should be non-negative when constructing portfolios (Jagannathan & Ma, 2003). We investigate whether the short-selling constraint alters the performance of the GMV portfolios dramatically and if it results in different relative performance amongst models. Table 8 shows the constrained GMV portfolio results of the empirical backtest. The simulation results are given in Appendix E and are similar to the outcomes discussed in this subsection.

We find that the optimal estimator for $K = 13$ is consistent with the unconstrained results. However, in the 42-factor dimension, the EWM is the best model in the constrained case, while the NL estimator attains the lowest SD for the unconstrained portfolios. The best dynamic model is consistent with the unconstrained outcomes in both dimensions. All estimators show lower SDs for the constrained GMV portfolios compared to the unconstrained ones. This shows that the additional shrinkage due to the short-selling constraint is effective and could be an attractive instrument for investors.

Table 8 shows that models that have not implemented some form of shrinkage yet improve the most in terms of SD, for both dimensions. Models such as CC and 1F do not show much better outcomes by adding the short-selling constraint, while the EWM and DCC models show improved SDs. This is especially the case for $K = 42$, where the EWM model overtakes the NL model as the best-performing model due to the additional shrinkage imposed. Note that the difference between the EWM and NL is minimal.

Comparing the constrained versus the unconstrained portfolios for the simulations, Tables E5 and E6 show that the real SD is higher in the constrained case. A reason for this is that the constrained portfolio has a smaller range of weights, so there are fewer hedging possibilities to lower the SD in recession periods. However, similarly to the empirical results, we obtain that the SDs in the restricted case are equal to or lower than their unconstrained counterparts. This illustrates again the additional shrinkage effect of imposing weight constraints. The constraints reduce sampling error while specification error increases relatively less, as illustrated by Jagannathan & Ma (2003).

Lastly, we observe in Table 8 that the improvement in the GMV portfolios in terms of SD is larger in a higher dimension ($K = 42$). This again illustrates the difficulty of high-dimensional estimation and the usefulness of shrinkage methods when there is a lot of estimation uncertainty. Additionally, the turnover is lower than for the unconstrained portfolios in both dimensions. This is due to the smaller range of values the weights can take on.

The weights of the constrained GMV portfolios are shown in Table D2. The minimum weights are equal to zero, resulting from the short-selling constraint that does not allow for negative weights. The maximum weights are also often lower compared to the unconstrained GMV portfolios. These two observations together cause the lower SD in the constrained case. This supports the observation that the turnover in Table 8 decreases when switching from unconstrained to constrained portfolios. These outcomes are beneficial in practice and incentivize implementing short-selling constraints. The results hold for both the 13- and 42-FMP dimension.

From the aforementioned observations, we can deduce that imposing a short-selling constraint is beneficial for the performance of GMV portfolios, especially when the dimension increases. Lastly, note that especially the improved performance of the GOGARCH for $K = 42$ gives an incentive to investigate the performance of the GOGARCH combined with shrinkage methods in future research.

Table 8: Average annualised performance of the GMV portfolios for 4 February 1997 until 27 April 2021.

	Benchmarks			EWM shrinkage				DCC models							
$K = 13$	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	1.830	0.648	0.637	0.669	**0.630*****	0.631	0.637	0.653	0.651	0.652	0.685	0.685	0.685	0.645	0.946
SR	0.988	1.546	1.562	1.469	1.603	1.589	1.581	1.464	1.512	1.479	1.468	1.472	1.472	1.423	1.438
Turnover	0.312	1.683	1.904	1.373	1.644	1.646	2.015	7.121	6.811	7.010	4.135	4.129	4.144	4.887	10.557
$K = 42$	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.825	0.214	0.175	0.285	0.343	0.355	**0.175*****	0.298	0.362	0.298	0.553	0.556	0.553	0.314	0.193
SR	0.719	0.171	0.271	0.781	1.461	1.564	0.294	0.085	1.248	0.232	1.254	1.275	1.251	0.296	0.326
Turnover	0.416	1.772	1.707	0.743	1.479	1.517	1.738	4.410	4.261	11.506	3.309	3.458	3.294	5.713	5.437
$K = 13$ **Cons**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	1.830	0.651	0.634	0.672	**0.629*****	0.630	0.637	0.649	0.647	0.651	0.684	0.684	0.684	0.643	0.653
SR	0.988	1.574	1.578	1.486	1.614	1.596	1.599	1.458	1.496	1.459	1.474	1.481	1.479	1.446	1.509
Turnover	0.364	1.398	1.526	1.241	1.428	1.386	1.650	6.131	5.936	6.109	4.116	4.110	4.124	4.254	8.643
$K = 42$ **Cons**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.825	0.209	**0.168*****	0.284	0.342	0.353	0.169	0.291	0.362	0.289	0.553	0.556	0.553	0.305	0.178
SR	0.719	0.255	0.360	0.774	1.456	1.563	0.378	0.133	1.250	0.267	1.254	1.274	1.251	0.323	0.514
Turnover	0.416	1.572	1.526	0.748	1.466	1.481	1.538	4.028	4.210	9.801	3.316	3.462	3.300	5.019	4.216

Note. A significant decrease of the SD of best-performing GMV portfolio compared to the SD of the GMV of the real covariance is indicated with a *,** and *** for a p-value below 0.1, 0.05 and 0.01, respectively. Using a two-sided test by Ledoit & Wolf (2011) and HAC standard errors.
K indicates the dimension of the FMP returns. The weekly performance is annualised by multiplying by 52 and a rolling window of 150 observations is used. The portfolios are rebalanced for each new observations, thus for each week. The SD is given in percentages.

7 Conclusion

We consider the large-dimension covariance estimation problem and deviate from the approach in literature such as Engle et al. (2019) by not considering an individual asset-based covariance matrix but instead consider the covariance matrix of FMP returns. An FMP entails a periodically rebalanced basket of assets, weighted based on their exposure to a certain factor. The main difference between FMP and asset returns is that FMP returns have different characteristics than stock returns, as discussed in Section 4. FMP returns attain a dynamic structure with time-varying characteristics due to the rebalancing within the FMPs over time, while stock returns keep the same interpretation.

Also, we consider a factor matrix of a size that falls between the sizes often discussed in literature for factors (fairly small dimensions) and those for assets (often a few hundred assets). We are therefore interested in finding if the considered numbers of variables exhibit the high-dimensional estimation problem as well. In particular, this research aims to find the best model in terms of covariance estimation accuracy and portfolio performance for the dimensions of 13 and 42 FMPs. We assess and compare the estimation performance of static shrinkage methods, dynamic models, and a combination of the two. The combination of the dynamic models with the relatively new QIS method of Ledoit & Wolf (2022b) also adds to the current literature.

In sum, we recommend implementing the CC and NL estimator for a dimension of 13 and 42 FMPs, respectively. The superior performance of these two models compared to PGGM's current estimator (S) is consistent over multiple analyses. They attain the lowest SD values for the GMV and constrained GMV portfolios for the empirical backtest and simulation study. The outcomes are supported by the direct analysis of the simulations, where the CC and NL estimators give the highest accuracy when comparing the real covariance and the estimator via the mean-variance loss function of Engle et al. (2019). Note that the only contradicting result is the outperformance of the EWM model compared to NL for the constrained GMV portfolio in a dimension of 42 FMPs. However, as the differences are minimal, we would still recommend the NL model. Lastly, we want to stress the importance of short-selling constraints because they improve the estimators in terms of SD in both dimensions.

It is also important to note that the SR portfolios are not reliable in practice because they are extremely cumbersome to obtain accurately due to inaccuracy in estimating the expected returns, leading to much estimation error in the mean estimates (Jagannathan & Ma, 2003). We therefore ignore these results when making this recommendation. Further research is required to determine which models perform best for mean-variance optimisation.

In more detail, we find for the dynamic models that the ADCC estimator is the best for the empirical analysis of the smaller dimension, showing that some additional asymmetric information can be captured. The best dynamic estimator for the dimension of 42 FMPs is the GOGARCH, indicating that imposing a factor structure on the FMP dataset leads to beneficial outcomes in the setting as presented in this paper. Nevertheless, both dynamic top performers still underperform compared to their static counterparts.

For the simulations, the direct and indirect results confirm the outcomes of the empirical analysis. We conclude that the best-performing models, CC and NL, are robust against fat-tailedness and misspecification of the covariance dynamics. In detail, the deviations between the simulations are minimal and there is no discernible pattern in the rank of best- and worst-performing models. It is important to consider that the best dynamic model is the DCC S for $K = 13$, in contrast to the empirical GMV portfolios where the ADCC estimator is the top performer. This is due to the additional asymmetry in the market that the simulations do not account for.

Next, we consider the performance of GMV portfolios combined with a short-selling constraint. We conclude that the best estimators are consistent with the empirical and simulation results without restriction for $K = 13$. For $K = 42$, the best model is the EWM for the empirical dataset, but the difference with the NL model, the top performer in all other evaluations, is minimal. More importantly, we find that portfolios with short-selling constraints generally perform much better. Especially the improvement in performance for the larger dimension is substantial and is most effective for models that have not yet implemented some form of shrinkage. This confirms the hypothesis that large dimensions are more difficult to estimate accurately and therefore require more shrinkage. The additional shrinkage by the portfolio constraints alleviates the extreme portfolio weights, which would otherwise have caused extreme portfolio returns.

8 Discussion

In this section, we discuss some striking results that deviate from existing literature and acknowledge where this paper is in line with earlier research.

Firstly, in contrast to Engle et al. (2019), we find that the DCC NL model is not always the top performer. Specifically, this research finds that in smaller dimensions, the DCC S outperforms the DCC NL estimator. A possible explanation for the difference between the two could be the size and type of the considered dataset. Engle et al. (2019) analyse a dataset of 100 assets, whereas we use an FMP dataset of dimension 13. Especially the characteristics of the FMPs, which rebalance weekly, differ from those of the assets that do not change in composition over time. Moreover, Ledoit & Wolf (2022b) argue that nonlinear shrinkage specifically yields significant improvement compared to its linear counterparts when the dimension size is very large compared to the sample size. Thus, we might not consider a large enough dimension to benefit from the nonlinear structure of the shrinkage.

Secondly, we observe that the DECO models perform poorly for the simulations and the empirical study. It is never among the top-performing models, while the related DCC models perform significantly better. This contrasts the results of Engle & Kelly (2012), who argue that the DECO model outperforms the DCC model regarding portfolio selection when considering out-of-sample forecasts. Engle & Kelly (2012) evaluate a dataset of daily asset returns, whereas we consider weekly factor returns. A possible explanation would be that the constant correlation assumption of the DECO model is not realistic for this dataset.

Thirdly, we can deduce that the diversification benefits grow larger moving from a $K = 13$ towards $K = 42$ dimension, resulting in a lower SD of the GMV portfolios. However, this reduction in SD is paired with an increase in estimation inaccuracy, confirming the high-dimensionality problem discussed in Ledoit & Wolf (2003), among others.

Lastly, the outperformance of the constrained GMV portfolio estimators compared to its unconstrained counterparts is confirmed by the findings of Jagannathan & Ma (2003) and in line with Brandt (2010). The latter states that imposing portfolio constraints is effective. This argument is based on Frost & Savarino (1988), who discuss that portfolio restrictions truncate the most extreme portfolio weights and reduce the estimation error of the covariance matrix based on sample moments. This is consistent with Michaud (1989), who states that the largest estimation errors cause the most extreme weights.

The inaccuracy in the mean of the SR portfolios makes it difficult to apply these portfolios in an empirical setting. Jobson & Korkie (1980) also discuss this, stating that the sample moments result in extreme and unstable portfolios. Also, Jagannathan & Ma (2003) mention that the estimation error in the sample mean is often so large that little is lost in ignoring the mean altogether. Chopra & Ziemba (2013) support this, stating that the estimation error in the mean affects the portfolio weights ten times more than errors in the covariances and variances combined.

This is an important result for portfolio managers who not only consider a certain level of risk, but also want to obtain the highest return paired with it. In this research, we focus solely on estimating the covariance matrix, while estimating the average return accurately is also vital when considering mean-variance portfolios. The latter is still a relatively underrepresented topic in finance and beyond and allows further investigation. Some examples of research that could be considered when assessing the accuracy of the sample mean are the ones by James & Stein (1992) and Black & Litterman (1992). The first applies shrinkage to the sample mean, while the latter uses a mixed estimation approach, starting from an equilibrium model.

The following additional recommended directions can be pursued to improve and expand this research. First of all, we suggest exploring different dynamic models. The most obvious ones are the ADCC and GOGARCH models combined with the shrinkage methods discussed in this paper. Especially the GOGARCH model performs quite well in high dimensions, particularly with the short-selling constraint. This shows that adding shrinkage to the GOGARCH model has beneficial consequences. We are thus curious how linear and nonlinear shrinkage methods would influence the outcomes of the GOGARCH model.

Other models we would recommend looking into are the Rotated ARCH (RARCH) and Rotated DCC (RDCC) of Noureldin et al. (2014). Noureldin et al. (2014) rotate the raw returns, extending the idea of variance targeting to covariance targeting in multivariate models in any dimension. Compared to the GOGARCH, the RARCH also captures the dynamics of the covariances of the rotated returns which improves the prediction of the conditional correlations of the unrotated returns (Noureldin et al., 2014). By applying such rotation to the devolatilised returns of the DCC, Noureldin et al. (2014) obtain the RDCC model. They conclude that the RDCC model performs better than the regular DCC, given its flexibility.

Lastly, one could consider high-frequency data. Andersen et al. (2003) show that high-frequency-based covariance predictions yield significantly lower portfolio volatility than methods employing daily returns, especially during periods of turbulence. Moreover, Bollerslev et al. (2020) propose a new asymmetric multivariate volatility (GARCH) model that exploits the estimates of variances and covariances based on the signs of high-frequency returns. The resulting realised semivariances allow for more nuanced responses to positive and negative shocks than threshold terms, which are modelled in the ADCC model. Another option is to apply the

DCC and DECO-HEAVY models of Bauwens & Xu (2023), which are also based on measures of realised variances and correlations built from intra-day data. Bauwens & Xu (2023) find that the scalar HEAVY models outperform the DCC and DECO models, which are based on a lower frequency, such as daily or weekly data.

References

Aielli, G. P. (2013). Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics*, *31*(3), 282–299.

Alexander, C. (2002). Principal component models for generating large GARCH covariance matrices. *Economic Notes*, *31*(2), 337–359.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579–625.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, 817–858.

Arnott, R., Harvey, C. R., Kalesnik, V., & Linnainmaa, J. (2019). Alice's adventures in factorland: Three blunders that plague factor investing. *Journal of Portfolio Management*, *45*(4), 18–36.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3–18.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*(3), 307–343.

Bauwens, L., Laurent, S., & Rombouts, J. V. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, *21*(1), 79–109.

Bauwens, L., & Xu, Y. (2023). DCC-and DECO-HEAVY: Multivariate GARCH models based on realized variances and correlations. *International Journal of Forecasting*, *39*(2), 938–955.

Berk, J. B. (1995). A Critique of Size-Related Anomalies. *Review of Financial Studies*, 275–286.

Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial analysts journal*, *48*(5), 28–43.

Blitz, D., & Vidojevic, M. (2019). The characteristics of factor investing. *The Journal of Portfolio Management*, *45*(3), 69–86.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics*, 498–505.

Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of political Economy*, *96*(1), 116–131.

Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2020). Multivariate leverage effects and realized semicovariance GARCH models. *Journal of Econometrics*, *217*(2), 411–430.

Brandt, M. W. (2010). Portfolio choice problems. In *Handbook of financial econometrics: Tools and techniques* (pp. 269–336). Elsevier.

Caporin, M., & McAleer, M. (2012). Do we really need both BEKK and DCC? A tale of two multivariate GARCH models. *Journal of Economic Surveys*, *26*(4), 736–751.

Caporin, M., & McAleer, M. (2014). Robust ranking of multivariate GARCH models by problem dimension. *Computational Statistics & Data Analysis*, *76*, 172–185.

Cappiello, L., Engle, R. F., & Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*, *4*(4), 537–572.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*(1), 57–82.

Chan, K. C., & Chen, N.-F. (1991). Structural and return characteristics of small and large firms. *Journal of Finance*, *46*(4), 1467–1484.

Chopra, V. K., & Ziemba, W. T. (2013). The effect of errors in means, variances, and covariances on optimal portfolio choice. In *Handbook of the fundamentals of financial decision making: Part i* (pp. 365–373). World Scientific.

Coqueret, G., & Milhau, V. (2014). Estimating covariance matrices for portfolio optimization. *ERI Scientific Beta White Paper*.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *Journal of Finance*, *53*(6), 1839–1885.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, *22*(5), 1915–1953.

Engle, R. F. (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., & Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, *30*(2), 212–228.

Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric theory*, *11*(1), 122–150.

Engle, R. F., Ledoit, O., & Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, *37*(2), 363–375.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fan, J., Liao, Y., & Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, *75*(4).

Frost, P. A., & Savarino, J. E. (1988). For better performance: Constrain portfolio weights. *Journal of Portfolio Management*, *15*(1), 29.

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, *58*(4), 1651–1683.

James, W., & Stein, C. (1992). Estimation with quadratic loss. *Breakthroughs in statistics: Foundations and basic theory*, 443–460.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, *48*(1), 65–91.

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance*, *56*(2), 699–720.

Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, *75*(371), 544–554.

Johnson, T. C. (2002). Rational momentum effects. *Journal of Finance*, *57*(2), 585–608.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, *49*(5), 1541–1578.

Ledoit, O., & Péché, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, *151*(1-2), 233–264.

Ledoit, O., & 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., & Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, *30*(4), 110–119.

Ledoit, O., & Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, *88*(2), 365–411.

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859.

Ledoit, O., & Wolf, M. (2011). Robust performances hypothesis testing with the variance. *Wilmott*, *2011*(55), 86–89.

Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, *40*(2), 1024–1060.

Ledoit, O., & Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, *48*(5), 3043–3065.

Ledoit, O., & 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., & Wolf, M. (2022b). Quadratic shrinkage for large covariance matrices. *Bernoulli*, *28*(3), 1519–1547.

Ma, C., Cheng, E., & Lee, W. (2022). The Lost Decade: Have Macro Factor Risk Premia Become Irrelevant? *Journal of Portfolio Management*, *49*(2), 95–110.

Markowitz, H. M. (1952). The utility of wealth. *Journal of political Economy*, *60*(2), 151–158.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, *8*(4), 323–361.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized'optimal? *Financial analysts journal*, *45*(1), 31–42.

Noureldin, D., Shephard, N., & Sheppard, K. (2014). Multivariate rotated ARCH models. *Journal of Econometrics*, *179*(1), 16–30.

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (scopes)* (pp. 1345–1350).

Pakel, C., Shephard, N., Sheppard, K., & Engle, R. F. (2021). Fitting vast dimensional time-varying covariance models. *Journal of Business & Economic Statistics*, *39*(3), 652–668.

Sagi, J. S., & Seasholes, M. S. (2007). Firm-specific attributes and the cross-section of momentum. *Journal of Financial Economics*, *84*(2), 389–434.

Santos, A. A., & Moura, G. V. (2014). Dynamic factor multivariate GARCH model. *Computational Statistics & Data Analysis*, *76*, 606–617.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425–442.

Silvennoinen, A., & Teräsvirta, T. (2009). Multivariate GARCH models. In *Handbook of financial time series* (pp. 201–229). Springer.

Stoll, H. R., & Whaley, R. E. (1983). Transaction costs and the small firm effect. *Journal of Financial Economics*, *12*(1), 57–79.

Van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, *17*(5), 549–564.

Zhang, X. F. (2006). Information Uncertainty and Stock Returns. *Journal of Finance*, *61*(1), 105–137.

A Appendix: Methodology

A.1 FMP construction

First of all, to determine the exposures per factor, it is important to note that the five different factors consist of the following characteristics as variables. The variables are obtained using various data sources. For value:

1. Book-to-Price ratio
2. Expected Profit-to-Price ratio
3. Free Cashflow-to-Price ratio
4. EBITDA-to-EV ratio

For Quality:

1. Profit-to-Assets ratio
2. Return on investments
3. Growth return on investments
4. Growth on assets

For Low Volatility only 1-year volatility is used, while for Size the investible market capitalisation is used. Lastly, for Momentum the following variables are considered:

1. 1-year return
2. 6-month return
3. Profit adjustments by analysts

A.2 GARCH model

For the DCC, DECO and ADCC models, we require a univariate GARCH model to describe the covariance of the factors. This paper follows the notation in Cappiello et al. (2006) and considers the GARCH model,

$$h_{i,t} = (1 - a_i - b_i) + a_i \varepsilon_{i,t-1}^2 + b_k h_{i,t-1}, \tag{A.1}$$

where $h_{i,t}$ is the conditional variance for factor i at time t, $\varepsilon_{i,t-1}$ gives the standardised residuals and the parameters (a_i, b_i) are calculated by maximum likelihood and required that $a_i + b_i < 1$ to ensure consistency and stationarity. For the GOGARCH model, we use the returns of the factor components $y_{i,t}$ instead of the standardized residuals. We re-estimate the parameters in the GARCH model for each new rolling window.

A.3 Estimation

For the DCC models, we assume the following specifications, as discussed in more detail in the methodology:

$$\begin{aligned}
&f_{t|t-1} \sim N(0, \Sigma_t),\\
&h_{i,t} = (1 - a_i - b_i) + a_i \varepsilon_{i,t-1}^2 + b_i h_{i,t-1},\\
&\varepsilon_{i,t} = f_{i,t}/\sqrt{h_{i,t}}\\
&S = \sum_{t=1}^{T} \frac{1}{T} \varepsilon_t' \varepsilon_t\\
&Q_t = (1 - \alpha - \beta)S + \alpha\{\varepsilon_{t-1}\varepsilon_{t-1}'\} + \beta Q_{t-1}\\
&\Sigma_t = D_t R_t D_t
\end{aligned}$$

where for $h_{i,t}$ the regular univariate GARCH is considered. The same holds for the specifications for the DECO and ADCC models.

For these models, the Composite Likelihood (MCLE) method of Pakel et al. (2021) is implemented. It approximates the full log-likelihood by the sum of marginal univariate likelihoods. The MCLE is often used when the MLE is too difficult to estimate, which can be the case for the complex models considered in this research. It assumes the DCC model pairs are used to obtain pair-wise log-likelihoods that finally result in the full log-likelihood. This is summarised by the following assumptions defined in Pakel et al. (2021): let $h_{j_1,t}(\eta_{j_1})$ and $h_{j_2,t}(\eta_{j_2})$ be the univariate volatility models for assets j_1 and j_2, where η_{j_1} and η_{j_2} are the model parameters. Let $\varepsilon_{jt}(\theta_j) = (\varepsilon_{j_1,t}(\eta_{j_1}), \varepsilon_{j_2,t}(\eta_{j_2}))'$ where $\theta_j = (\eta_{j_1}', \eta_{j_2}')'$, $\varepsilon_{j_1,t}(\eta_{j_1}) = r_{j_1,t} h_{j_1,t}^{-1/2}(\eta_{j_1})$ and similarly for $\varepsilon_{j_2,t}(\eta_{j_2})$. Let $\phi = (\alpha, \beta)'$, then the DCC model for the jth pair is given by:

$$\begin{aligned}
&H_{jt}(\theta_j, S_j(\theta_j, \phi), \phi) = D_{jt}(\theta_j) R_{jt}(\theta_j, S_j(\theta_j, \phi), \phi) D_{jt}(\theta_j),\\
&R_{jt}(\theta_j, \bar{R}_j(\theta_j, \phi), \phi) = (Q_{jt}^{*-1/2}(\theta_j, \phi) Q_{jt})(\theta_j, \bar{R}_j(\theta_j, \phi), \phi)(Q_{jt}^{*-1/2}(\theta_j, \phi)),\\
&Q_{jt}(\theta_j, \bar{R}_j(\theta_j, \phi), \phi) = (1 - \alpha - \beta) S(\theta_j, \phi) + \alpha\{Q_{j,t-1}^{*-1/2}(\theta_j, \phi)\varepsilon_{j,t-1}(\theta_j)\varepsilon_{j,t-1}'(\theta_j) Q_{j,t-1}^{*-1/2}(\theta_j, \phi)\}\\
&\qquad\qquad\qquad\qquad + \beta Q_{j,t-1}(\theta_j, \phi),\\
&S_j(\theta_j, \phi) = \mathrm{E}[\varepsilon_{jt}(\theta_j)' \varepsilon_{jt}(\theta_j)]
\end{aligned}$$

where $D_{jt}(\theta_j)$ is a (2×2) diagonal matrix with the diagonal entries $\sqrt{h_{j_1,t}(\eta_{j_1})}$ and $\sqrt{h_{j_2,t}(\eta_{j_2})}$ and $Q_{jt}^*(\theta_j, \phi)$ is a (2×2) matrix consisting of the diagonal entries of $Q_{jt}^*(\theta_j, S_j, \phi)$. Then, we follow Pakel et al. (2021), who first target the intercept, conditional on the correlation dynamics. Firstly, we obtain the estimator for the volatility part:

$$\hat{\theta}_j = \mathrm{argmax}_{\eta_{j_1}\eta_{j_2}} (l_{j_1 T}(\eta_{j_1}), l_{j_2 T}(\eta_{j_2}))', \tag{A.2}$$

where $l_{iT}(\cdot)$ is the log-likelihood for asset i and defined as $l_{jT}(\theta_j, S_j, \phi) = T^{-1} \sum_{t=1}^{T} l_{jt}(\theta_j, S_j, \phi)$, where $l_{jt}(\theta_j, S_j, \phi) = -\log|H_{jt}(\theta_j, S_j, \theta)| - X_{jt}' H_{jt}^{-1}(\theta_j, S_j, \phi) X_{jt}$. The composite likelihood estimator or ϕ is then given by

$$\hat{\phi} = \text{argmax}_{\phi} \frac{1}{K} \sum_{j=1}^{K} l_{jT}(\hat{\theta}_j, \hat{S}_j(\hat{\theta}_j, \phi), \phi), \tag{A.3}$$

where $\hat{S}_j(\hat{\theta}_j, \phi) = T^{-1} \sum_{t=1}^{T} Q_{jt}^{*1/2}(\theta_j, \phi)\varepsilon_{jt}(\theta_j)\varepsilon'_{jt}(\theta_j)Q_{jt}^{*1/2}(\theta_j, \phi)$. Note that we have used the earlier obtained estimator $\hat{\theta}$.

For the GOGARCH model we apply an MLE, as defined in Van der Weide (2002). The log-likelihood is expressed as

$$L = -\frac{1}{2} \sum_{t=1}^{T} (K \log(2\pi) + \log |Z_\theta H_t Z'_\theta| + y'_t Z'_\theta (Z_\theta H_t Z'_\theta)^{-1} Z_\theta y_t, \tag{A.4}$$

$$L = -\frac{1}{2} \sum_{t=1}^{T} (K \log(2\pi) + \log |Z_\theta Z'_\theta| + \log |H_t| + y'_t H_t^{-1} y_t, \tag{A.5}$$

where $Z_\theta Z'_\theta = P \Lambda P'$ is independent of θ. Thus, in the first step P and Λ matrices are estimated using unconditional information. In the second step, the rotation coefficients U and the parameters of the component univariate GARCH models are estimated using conditional information.

A.4 Nonlinear shrinkage assumptions

The nonlinear shrinkage method of Ledoit & Wolf (2022b) uses the following assumptions:

1. (Assumption 3.1) Let T denote the sample size and K the number of variables. It is assumed that the concentration ratio ($c = K/T$) converges to a limit $c \in (0, 1)$. There also exists a compact interval included in $(0, 1)$ that contains K/T for all T large enough.

2. (Assumption 3.2) Assumptions on the covariance matrix

   (a) The population covariance matrix Σ is a nonrandom symmetric positive-definite matrix of dimension $K \times K$.

   (b) X_t is an $T \times K$ matrix of i.i.d. random variables with mean zero, variance one and finite 16th moment. The matrix of observations is $Y_t = X_t \times \sqrt{\Sigma_t}$ where only Y is observed.

   (c) Let $\tau_t = (\tau_{t,1}, \ldots, \tau_{t,k})'$ denote the system of eigenvalues of Σ_t and H_t the c.d.f of population eigenvalues. It is assumed that H_t converges weakly to a limit law H called the limiting spectral distribution.

   (d) $\text{Supp}(H)$, the support of H, is the union of a finite number of closed intervals bounded away from zero and infinity. Also, there exists a compact interval $[\underline{T}, \overline{T}] \subset (0, \inf)$ that contains $\{\tau_{t,1}, \ldots \tau_{t,K}\}$ for all t large enough.

3. (Assumption 3.3) There exists a nonrandom real univariate function $\tilde{\delta}$ defined on $\text{Supp}(F)$ and continuously differentiable such that $\tilde{\delta}_t(x) \to \tilde{\delta}(x)$, for all $x \in \text{Supp}(F)$. Furthermore, this convergence is uniform over $x \in \cup_{j=1}^{\nu} [a_j + \eta, b_j - \eta]$, for any small $\eta > 0$. Finally for

any small $\eta > 0$, there exists a finite nonrandom constant $\hat{J}$ such that almost surely, over the set $x \in \cup_{j=1}^{\nu}[a_j - \eta, b_j + \eta]$, $\tilde{\delta}_t(x)$ is uniformly bounded by $\hat{J}$ from above and by $1/\hat{J}$ from below, for t large enough.

A.4 EWM benchmark

We consider the exponentially weighted moving average (EWM) model as one of our benchmarks. The exponential decay of the covariance matrix uses a half-life, τ equal to 52 weeks and the idiosyncratic volatilities have a half-life of 26 weeks, following Axioma. The EWM model is defined as follows:

$$\Sigma_t = \lambda \Sigma_{t-1} + (1-\lambda) f_{t-1} f'_{t-1}, \tag{A.6}$$

where Σ_t is the covariance at day t, λ is the decay factor set such that $\lambda = 1 - e^{\frac{-\ln(2)}{\tau}}$ which differs for volatilities and covariances. For Σ_{t-1}, we use the estimate obtained for the previous iteration because the "real" covariance can not even be determined ex-post.

A.5 Covariance evaluation

The loss function described by Engle et al. (2019) investigates the efficiency of the covariance around different alternatives. The loss function is called the minimum variance loss function and is defined as follows:

$$L(\hat{\Sigma}_t, \Sigma_t) = \frac{\mathrm{Tr}(\hat{\Sigma}_t^{-1} \Sigma_t \hat{\Sigma}_t^{-1})/K}{[\mathrm{Tr}(\hat{\Sigma}_t^{-1})/K]^2} - \frac{1}{\mathrm{Tr}(\Sigma_t^{-1})/K}, \tag{A.7}$$

where K gives the dimension of the covariance matrix, Σ_t gives the real covariance at time t measured by the covariance value used to simulate or an unbiased proxy and $\hat{\Sigma}_t$ gives the estimated covariance. After defining the loss for each time period t, we determine the average loss by averaging over the total sample size T.

Afterwards, it is possible to derive the improvement in accuracy of the covariance estimation compared to a benchmark or the current method, the EWM with linear shrinkage. We follow Engle et al. (2019) that proposes a percentage relative improvement in average loss (PRIAL). For example, the PRIAL of the DCC NL with respect to the S estimator is defined as:

$$100 \times \left\{ 1 - \frac{\mathrm{E}[\hat{L}^{\mathrm{DCCNLNL}}]}{\mathrm{E}[\hat{L}^{\mathrm{S}}]} \right\} \%, \tag{A.8}$$

where the loss $\hat{L}$ for each estimator is calculated by the aforementioned loss function and the expectation is taken over time and across Monte Carlo simulations. By construction, the PRIAL of the true conditional covariance matrix with respect to any other estimator is 100%, which is the maximum attainable, while 0% means no improvement (Engle et al., 2019). A large negative number indicates models that show worse performance than the EWM S estimator.

A.6 Inference

We compute the turnover as the difference in the weights of the factors that are implemented in the portfolio at rebalancing dates. Note that we are not considering the rebalancing of the assets within the factors, but only of the factors themselves. Turnover is calculated as

$$\text{Turnover} = \frac{1}{T - RW - 1} \sum_{t=1}^{T-1} \sum_{k=1}^{K} (|w_{j,k,t+1} - w_{j,k,t+}|), \tag{A.9}$$

where T is the total sample size, RW is the rolling window used for construction and $w_{j,k,t+1}$ is the desired factor weight for strategy j, factor k at $t+1$. The weight $w_{j,k,t+}$ is the weight of the factor k before rebalancing at the beginning of the new period. Turnover proxies transaction costs and indicate whether the weights in factors deviate a lot over the investment period.

For the HAC inference we follow Ledoit & Wolf (2008). The limiting covariance matrix , Ψ is given by

$$\Psi = \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathrm{E}[y_s y_t] \qquad \text{with} \qquad y_t' = (r_{ti} - \mu_1, r_{tn} - \mu_n, r_{ti}^2 - \gamma_i, r_{tn}^2 - \gamma_n). \tag{A.10}$$

By change of variables, the limit can be alternatively expressed as

$$\Psi = \lim_{T\to\infty} \Psi_T, \qquad \text{with} \qquad \Psi_T = \sum_{j=-T+1}^{T-1} \Gamma_T(j), \qquad \text{where} \tag{A.11}$$

$$\Gamma_T(j) = \begin{cases} \frac{1}{T} \sum_{t=j+1}^{T} \mathrm{E}[y_t y_{t-j}'] & \text{for} \quad j \geq 0 \\ \frac{1}{T} \sum_{t=-j+1}^{T} \mathrm{E}[y_{t+j} y'] & \text{for} \quad j < 0. \end{cases} \tag{A.12}$$

The standard method uses heteroskedasticity and autocorrelation robust (HAC) kernel estimation to obtain a consistent estimator $\hat{\Psi} = \hat{\Psi}_T$, for example as in Andrews (1991). This requires choosing a kernel function $k(\cdot)$. The kernel function needs to adhere to three conditions: $k(0) = 1$, $k(\cdot)$ is continuous at 0 and $\lim_{x\to\pm\infty} k(x) = 0$ (Ledoit & Wolf, 2008). The kernel estimate for Ψ is then given by

$$\hat{\Psi} = \hat{\Psi}_T \frac{T}{T-4} \sum_{l=-T+1}^{T-1} k\Big(\frac{j}{S_T}\Big) \hat{\Gamma}_T(j), \qquad \text{where} \tag{A.13}$$

$$\hat{\Gamma}_T(l) = \begin{cases} \frac{1}{T} \sum_{t=l+1}^{T} \mathrm{E}[y_t y_{t-l}'] & \text{for} \quad j \geq 0 \\ \frac{1}{T} \sum_{t=-l+1}^{T} \mathrm{E}[y_{t+l} y'] & \text{for} \quad j < 0 \end{cases}, \qquad \text{with} \tag{A.14}$$

$$\hat{y}_t' = (r_{tj} - \hat{\mu}_1, r_{tn} - \hat{\mu}_n, r_{tn}^2 - \hat{\gamma}_n). \tag{A.15}$$

The bandwidth is denoted by S_T, r_{tj} indicates the return of portfolio j at time t, $T/(T-4)$ is a small-sample degrees of freedom adjustment introduced to offset the effect of the estimation of the 4-vector v in the computation of the $\hat{\Gamma}_T(j)$ (Ledoit & Wolf, 2008). For the kernel $k(\cdot)$, one often opts for one with $q = 2$, where $1 \leq q \leq \infty$ determines the smoothness of the kernel at the origin (Ledoit & Wolf, 2008). We use the Parzen kernel. For the bandwidth S_T we use the automatic bandwidth selection of Andrews (1991).

B Appendix: Data

Table B1 describes the complete dataset, consisting of 42 FMPs for the time period of 4 February 1997 until 27 April 2021. The country-specific FMPs are the Indo-Asia Pacific Region (AP), the European Union (EU), Japan (JP), North America (NA) and the United Kingdom (UK). The sector-specific FMPs are the remaining FMPs that are not part of the 13-FMP dataset. Especially the country FMPs JP and AP hold a higher standard deviation, indicating a more volatile market than other countries. Moreover, especially the energy and telecommunication sectors seem more volatile. All the FMPs are non-normal distributed due to a kurtosis deviating largely from 3 and skewness unequal to zero.

Table B1: Descriptive statistics of the FMP return for period 4 February 1997 until 27 April 2021.

	Mean	SD	SR	Kurtosis	Skewness
Market	0.183	2.365	0.077	5.914	-0.371
Within Value	0.021	0.272	0.077	5.849	0.418
Within Quality	0.020	0.180	0.112	7.504	0.312
Within Momentum	0.021	0.360	0.058	8.389	-1.108
Within Size	0.020	0.326	0.063	7.366	-0.268
Within LowVol	0.016	0.520	0.031	5.253	-0.082
Within Axioma Market Sensitivity	0.015	0.649	0.024	14.161	-0.038
Within Axioma Liquidity	0.019	0.276	0.070	5.273	-0.020
AP	0.017	1.896	0.009	8.616	-0.289
EU	-0.006	1.309	-0.004	7.106	0.280
JP	-0.065	2.411	-0.027	5.500	0.170
NA	0.025	0.759	0.033	8.155	0.123
UK	-0.024	1.380	-0.017	7.923	0.074
Automobiles & Components	0.009	1.195	0.008	11.917	0.177
Banks	-0.033	0.987	-0.034	9.143	-0.027
Capital Goods	0.013	0.702	0.019	5.264	-0.124
Commercial & Professional Services	-0.034	0.832	-0.041	4.930	-0.104
Consumer Durables & Apparel	-0.027	0.808	-0.034	3.867	0.016
Consumer Services	-0.016	0.978	-0.016	18.759	-0.803
Diversified Financials	0.007	0.954	0.008	6.038	-0.037
Energy	-0.049	1.453	-0.034	5.997	-0.207
Food & Staples Retailing	0.000	0.955	0.000	11.054	0.706
Food, Beverage & Tobacco	-0.007	0.752	-0.009	5.132	0.347
Health Care Equipment & Services	0.039	1.125	0.035	6.220	-0.017
Household & Personal Products	0.001	1.003	0.001	6.906	-0.066
Insurance	-0.011	0.875	-0.012	7.297	0.019
Materials	0.008	0.945	0.009	5.714	0.147
Media	0.003	0.965	0.004	5.578	-0.017
Pharmaceuticals, Biotechnology & Life Sciences	0.039	0.987	0.039	5.448	0.191
Real Estate	0.015	0.931	0.016	7.656	-0.437
Retailing	0.027	1.006	0.027	5.491	0.264
Semiconductors & Semiconductor Equipment	0.024	1.435	0.017	5.403	0.106
Software & Services	0.045	1.022	0.044	5.156	-0.130
Technology Hardware & Equipment	0.020	0.984	0.020	4.614	0.017
Telecommunication Services	0.004	1.138	0.004	5.904	0.022
Transportation	-0.013	0.859	-0.016	5.722	-0.072
Utilities	0.019	1.088	0.018	5.962	-0.163
Across Value	-0.009	0.868	-0.010	6.884	0.071
Across LowVol	0.010	1.697	0.006	6.460	0.283
Across Quality	0.053	0.675	0.078	6.765	-0.220
Across Momentum	0.065	1.280	0.051	6.918	-0.338
Across Size	0.007	1.294	0.005	5.689	-0.144

Note. This table gives the statistical characteristics of full dataset of 42 FMPs in percentages (Mean and SD).

Table B2: Correlation characteristics of the FMP return for period 4 February 1997 until 27 April 2021.

	Mean	Median	Min	Max
K=13	-0.004	-0.009	-0.625	0.643
K=42	-0.012	-0.010	-0.674	0.633

Note. K indicates the FMP dimension.

C Appendix: Simulation results

C.1 Parameters

Tables C1 and C2 contain the parameter estimates of the different dynamic models, except the GOGARCH that uses matrices as parameters and are therefore not possible to show in this paper. To quantify time-variation, we also present the average min-max ranges of the parameter estimates. We find that the average parameter estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{g}$ are in line with other research; the values for $\hat{\alpha}$ and $\hat{g}$ are small and for $\hat{\beta}$ are large where they sum up close to one. We find that the estimates of the shrunk models attain similar parameter values and that the ADCC model has a lower average parameter value for $\hat{\beta}$ than the other DCC models. This is in line with the findings of Cappiello et al. (2006). The min-max range indicates that the parameter estimates depend on time-variation. Note that especially $\hat{\beta}$ has a wide range, indicating that the parameters change a lot over time. Moreover, the average $\hat{g}$ estimate is fairly small, indicating that the asymmetric information that can be captured is limited for the considered simulations.

Table C1: Average parameter estimates for the DCC models for different simulations with 13 FMPs.

Normal EWM	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC
$\hat{\alpha}$	0.004	0.007	0.005	0.004	0.007	0.005	0.000
	[0.000, 0.031]	[0.000, 0.056]	[0.000, 0.038]	[0.000, 0.031]	[0.000, 0.056]	[0.000, 0.038]	[0.000, 0.000]
$\hat{\beta}$	0.943	0.956	0.949	0.943	0.956	0.949	0.929
	[0.796, 0.997]	[0.827, 0.997]	[0.737, 0.997]	[0.796, 0.997]	[0.827, 0.997]	[0.737, 0.997]	[0.000, 1.000]
$\hat{g}$							0.002
							[0.000, 0.040]
***t*-dist EWM**	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC
$\hat{\alpha}$	0.006	0.006	0.006	0.006	0.006	0.006	0.002
	[0.000, 0.072]	[0.000, 0.066]	[0.000, 0.071]	[0.000, 0.072]	[0.000, 0.066]	[0.000, 0.071]	[0.000, 0.024]
$\hat{\beta}$	0.926	0.940	0.933	0.926	0.940	0.933	0.708
	[0.040, 0.999]	[0.633, 0.996]	[0.526, 0.9960.526,]	[0.040, 0.999]	[0.633, 0.996]	[0.526, 0.996]	[0.000, 1.000]
$\hat{g}$							0.007
							[0.000, 0.039]
Normal DCC	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC
$\hat{\alpha}$	0.026	0.021	0.025	0.026	0.021	0.025	0.005
	[0.000, 0.164]	[0.000, 0.156]	[0.000, 0.153]	[0.000, 0.164]	[0.000, 0.156]	[0.000, 0.153]	[0.000, 0.024]
$\hat{\beta}$	0.928	0.948	0.937	0.928	0.948	0.937	0.906
	[0.716, 0.997]	[0.786, 0.997]	[0.773, 0.998]	[0.716, 0.997]	[0.786, 0.997]	[0.773, 0.998]	[0.289, 1.000]
$\hat{g}$							0.001
							[0.000, 0.012]
***t*-dist DCC**	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC
$\hat{\alpha}$	0.001	0.002	0.001	0.001	0.002	0.001	0.002
	[0.000, 0.067]	[0.000, 0.019]	[0.000, 0.070]	[0.000, 0.067]	[0.000, 0.019]	[0.000, 0.070]	[0.000, 0.018]
$\hat{\beta}$	0.949	0.944	0.949	0.949	0.944	0.949	0.689
	[0.315, 0.989]	[0.710, 0.989]	[0.315, 0.989]	[0.315, 0.989]	[0.710, 0.989]	[0.315, 0.989]	[0.000, 1.000]
$\hat{g}$							0.010
							[0.000, 0.030]

Note. The minimum and maximum parameter estimates are displayed in brackets below the average parameter estimates. The average estimates are close to zero, but never exactly zero. The average parameters are based on 250 simulated covariances for 100 iterations. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or *t*-distribution for the returns. The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007.

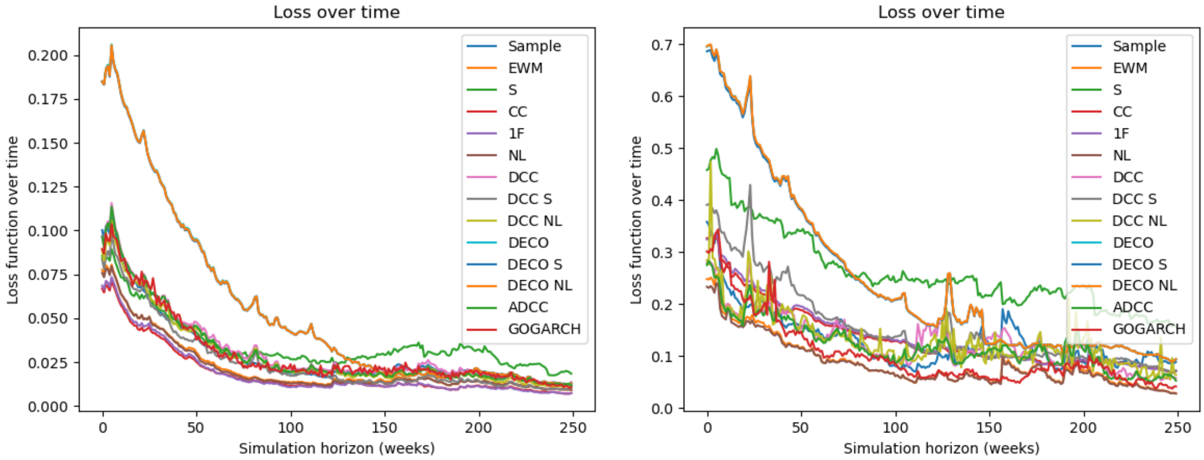
Table C2: Average parameter estimates for the DCC models for different simulations with 42 FMPs.

Normal EWM	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC
$\hat{\alpha}$	0.000 [0.000, 0.001]	0.001 [0.000, 0.013]	0.001 [0.000, 0.015]	0.000 [0.000, 0.001]	0.001 [0.000, 0.013]	0.001 [0.000, 0.015]	0.001 [0.000, 0.006]
$\hat{\beta}$	0.953 [0.911, 0.986]	0.950 [0.910, 1.000]	0.953 [0.911, 0.986]	0.953 [0.911, 0.986]	0.950 [0.910, 1.000]	0.953 [0.911, 0.986]	0.518 [0.963, 1.000]
$\hat{g}$							0.000 [0.000, 0.001]
***t*-dist EWM**	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC
$\hat{\alpha}$	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$\hat{\beta}$	0.966 [0.912, 0.987]	0.964 [0.918, 0.988]	0.963 [0.916, 0.987]	0.966 [0.912, 0.987]	0.964 [0.918, 0.988]	0.963 [0.916, 0.987]	0.888 [0.102, 0.999]
$\hat{g}$							0.000 [0.000, 0.001]
Normal DCC	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC
$\hat{\alpha}$	0.006 [0.000, 0.094]	0.012 [0.000, 0.097]	0.012 [0.000, 0.100]	0.006 [0.000, 0.094]	0.012 [0.000, 0.097]	0.012 [0.000, 0.100]	0.000 [0.000, 0.004]
$\hat{\beta}$	0.892 [0.000, 0.985]	0.859 [0.000, 1.000]	0.854 [0.000, 0.999]	0.892 [0.000, 0.985]	0.859 [0.000, 1.000]	0.854 [0.000, 0.999]	0.582 [0.000, 1.000]
$\hat{g}$							0.000 [0.000, 0.001]
***t*-dist DCC**	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC
$\hat{\alpha}$	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]
$\hat{\beta}$	0.964 [0.914, 0.989]	0.968 [0.912, 0.995]	0.964 [0.913, 0.988]	0.964 [0.914, 0.989]	0.968 [0.912, 0.995]	0.964 [0.913, 0.988]	0.652 [0.000, 0.999]
$\hat{g}$							0.000 [0.000, 0.003]

Note. The minimum and maximum parameter estimates are displayed in brackets below the average parameter estimates. The average estimates are close to zero, but never exactly zero. The average parameters are based on the 250 simulated covariances for 100 iterations. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns. The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007.

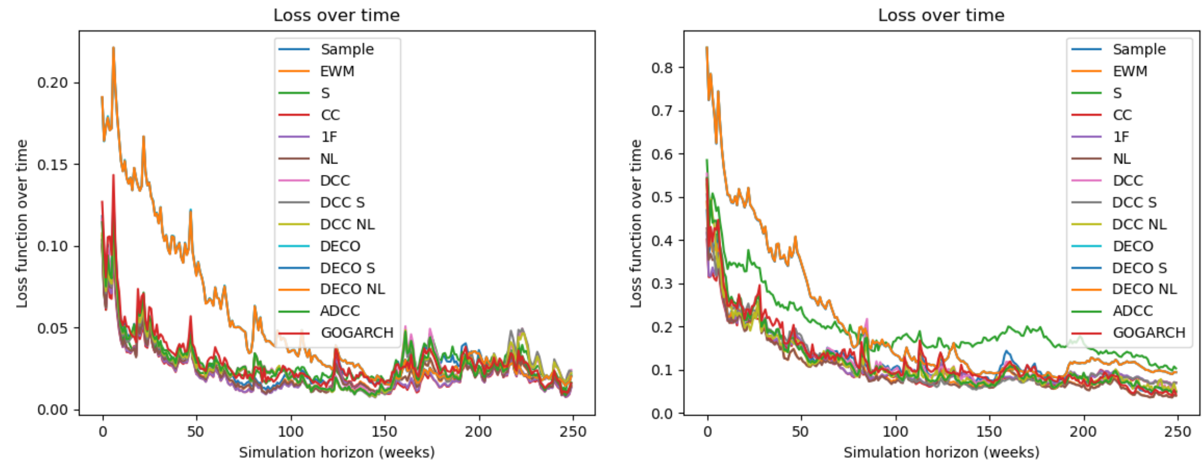
C.2 Direct evaluation

The following figures give the performance of the Engle et al. (2019) loss function over time for the dataset consisting of 13 FMPs and rolling window of 150 weeks. The simulation is based on the first 550 weekly datapoints, from 4 February 1997 until 14 August 2007. The simulation is reiterated for 100 times to provide reliable results.



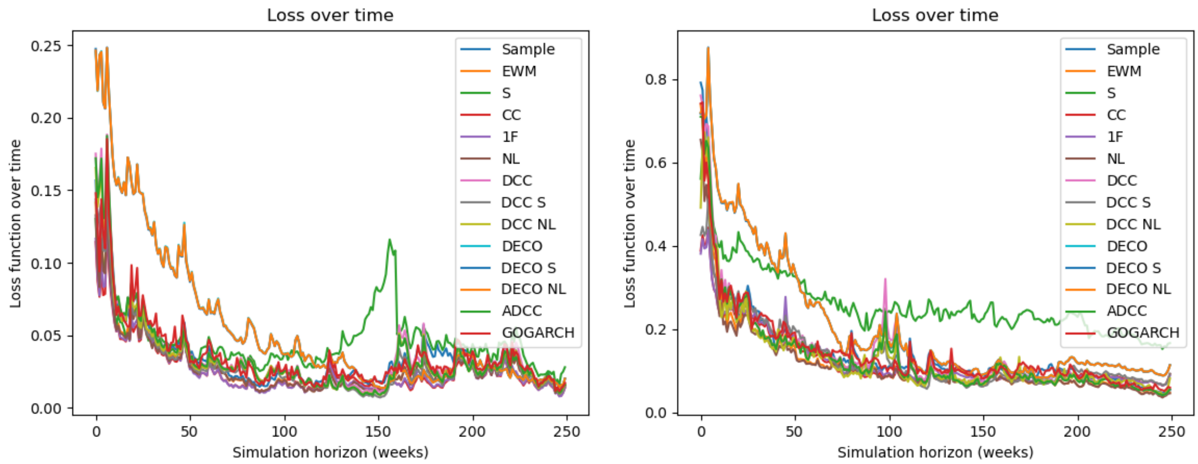
(a) 13 FMPs (b) 42 FMPs

Figure C1: Average mean-variance loss function of Engle et al. (2019) of the 250 simulated covariances for the estimation models for a rolling window of 150 weekly observations with simulation data by the t-distribution of the EWM covariance for 100 iterations.



(a) 13 FMPs (b) 42 FMPs

Figure C2: Average mean-variance loss function of Engle et al. (2019) of the 250 simulated covariances for the estimation models for rolling window of 150 weekly observations with simulation data by the normal distribution of the DCC covariance for 100 iterations.



(a) 13 FMPs

(b) 42 FMPs

Figure C3: Average mean-variance loss function of Engle et al. (2019) of the 250 simulated covariances for the estimation models for a rolling window of 150 weekly observations with simulation data by the t-distribution of the DCC covariance for 100 iterations.

Table C3: PRIAL of the mean-variance loss for estimation window of 150 weeks and 13 FMPs with the S model as benchmark. Values are given in percentages.

	Benchmarks		EWM Shrinkage			DCC models							
	Sample	EWM	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
EWM normal	-14.146	21.150	**27.479**	26.743	21.224	-25.118	-13.484	-18.919	-149.171	-148.843	-147.538	-19.867	-7.169
EWM t-dist	-6.040	21.113	**28.193**	27.561	23.817	-4.851	9.588	2.107	-129.416	-128.030	-128.169	0.937	-8.271
DCC normal	-5.836	13.343	**20.888**	18.864	15.397	-4.846	1.675	-1.490	-83.171	-83.240	-82.908	2.173	-11.072
DCC t-dist	-6.106	11.661	**21.477**	19.401	13.708	-5.022	10.797	2.512	-80.704	-80.128	-80.665	-1.984	-8.832

Note. Highest value is indicated in bold. The loss used to calculate the PRIAL is averaged over the 250 losses from the covariance estimates and 100 simulations for each estimator. PRIAL gives the percentage relative improvement in average loss compared to the S model. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns. The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007.

Table C4: PRIAL of the LW loss for estimation window of 150 weeks and 42 FMPs with the LW as benchmark. Values are given in percentages.

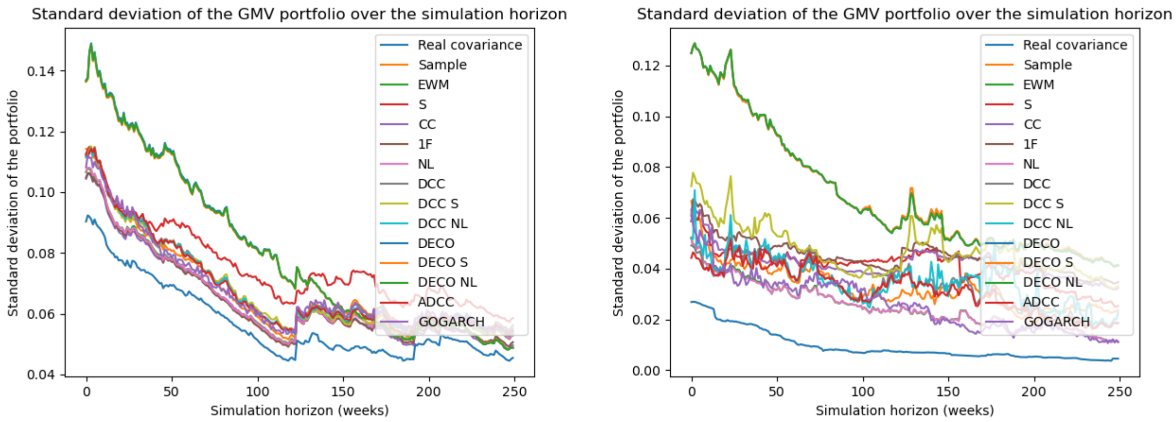
	Benchmarks		EWM Shrinkage			DCC models							
	Sample	EWM	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
EWM normal	33.469	64.973	36.647	36.222	**65.750**	36.795	28.396	31.226	-25.545	-25.456	-25.725	42.721	63.019
EWM t-dist	50.103	66.686	48.175	47.092	**67.788**	51.073	40.819	49.745	5.927	6.634	5.844	54.073	60.284
DCC normal	37.742	46.388	43.846	42.444	**48.524**	37.032	40.765	40.910	-2.855	-2.669	-2.933	39.348	35.993
DCC t-dist	45.534	54.653	54.250	53.246	**57.471**	46.929	50.049	51.801	20.700	21.000	20.626	48.507	46.732

Note. Highest value is indicated in bold. The loss used to calculate the PRIAL is averaged over the 250 losses from the covariance estimates and 100 simulations for each estimator. PRIAL gives the percentage relative improvement in average loss compared to the S model. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns. The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007.

C.3 Indirect evaluation

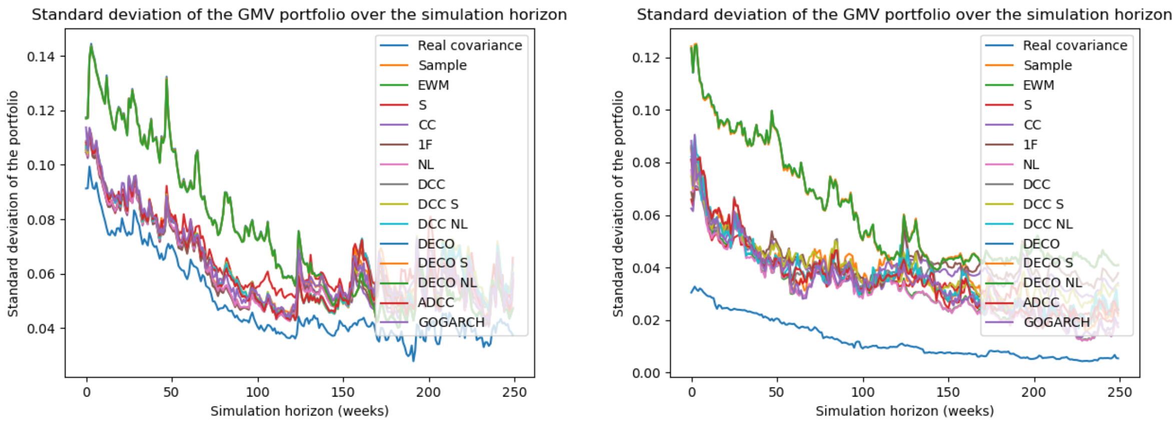
C.3.1 Global minimum variance portfolio

The following figures give the standard deviation of the GMV portfolios over time, for a rolling window of 150 weeks. The simulation is based on the first 550 weekly datapoints, from 4 February 1997 until 14 August 2007. The simulation is reiterated for 100 times to provide reliable results.



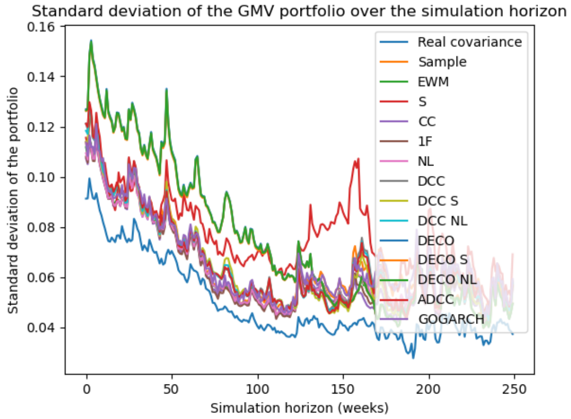
(a) 13 FMPs (b) 42 FMPs

Figure C4: Average volatility of the GMV portfolios based on simulation data of 250 weekly returns and 100 iterations by the t-distribution simulation of the EWM covariance for a rolling window of 150 weekly observations.

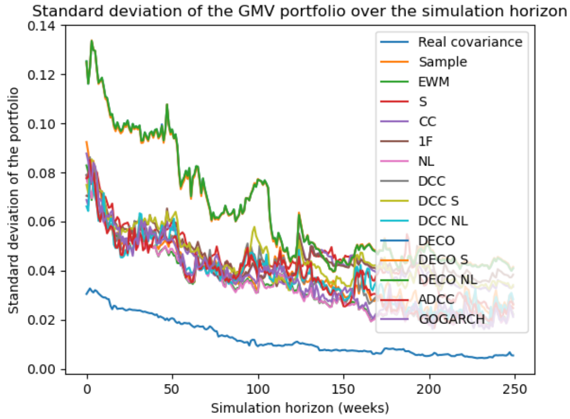


(a) 13 FMPs (b) 42 FMPs

Figure C5: Average volatility of the GMV portfolios based on simulation data of 250 weekly returns and 100 iterations by the normal distribution simulation of the DCC covariance for a rolling window of 150 weekly observations.



(a) 13 FMPs

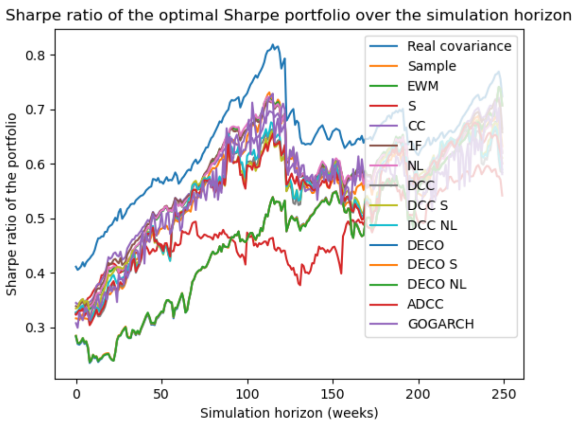


(b) 42 FMPs

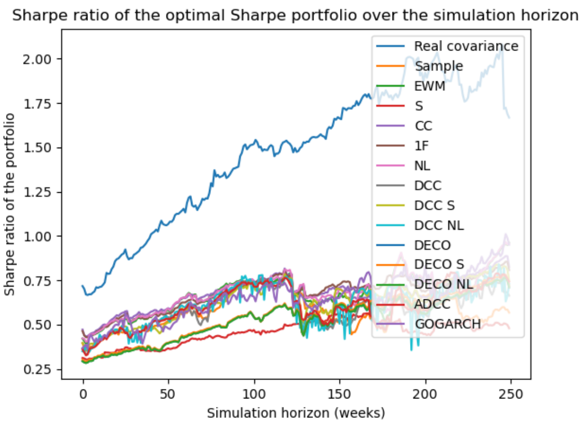
Figure C6: Average volatility of the GMV portfolios based on simulation data of 250 weekly returns and 100 iterations by the t-distribution simulation of the DCC covariance for a rolling window of 150 weekly observations.

C.3.2 Sharpe ratio portfolio

The simulation is based on the first 550 weekly datapoints, from 4 February 1997 until 14 August 2007. The simulation is reiterated for 100 times to provide reliable results.

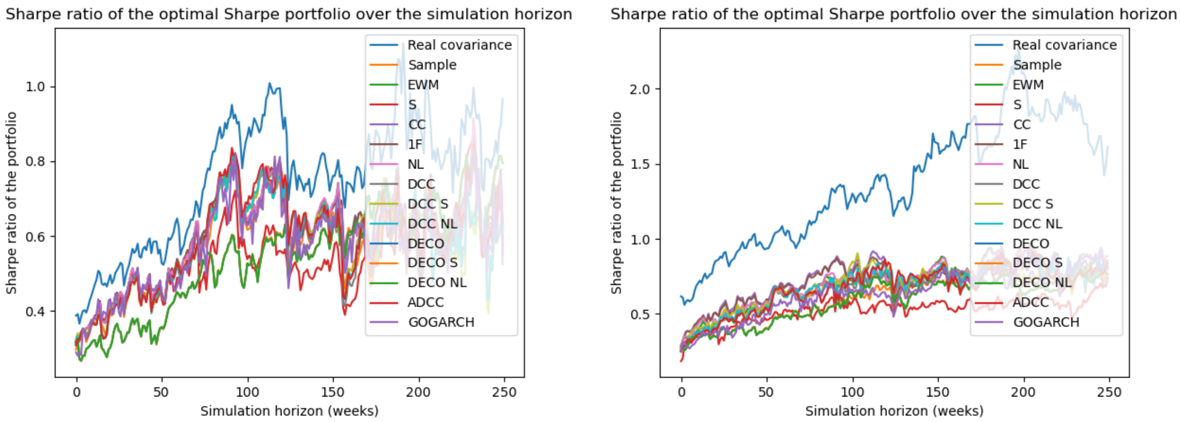


(a) 13 FMPs



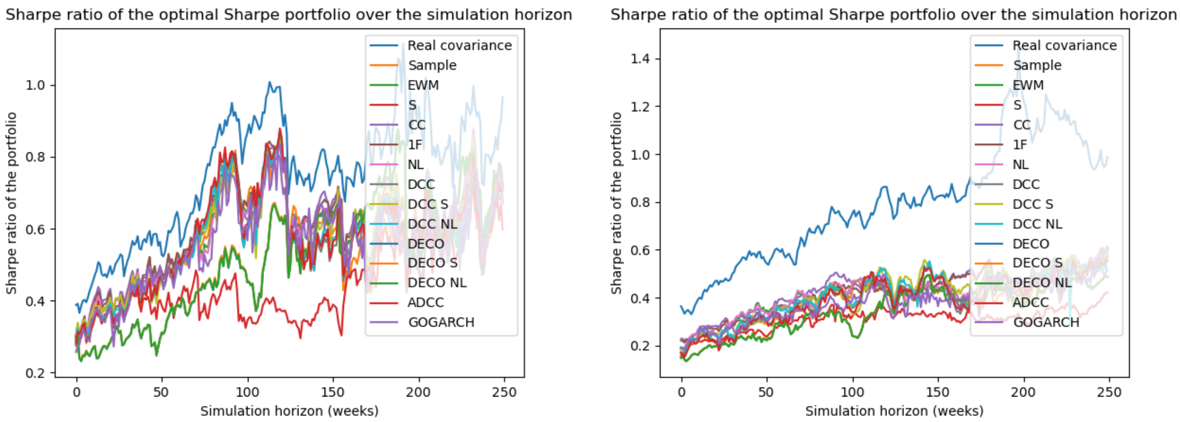
(b) 42 FMPs

Figure C7: Average Sharpe ratio of the SR portfolios based on simulation data of 250 weekly returns and 100 iterations by the t-distribution simulation and the EWM covariance for a rolling window of 150 weekly observations.



(a) 13 FMPs (b) 42 FMPs

Figure C8: Average Sharpe ratio of the SR portfolios based on simulation data of 250 weekly returns and 100 iterations by the normal distribution simulation and the DCC covariance for a rolling window of 150 weekly observations.



(a) 13 FMPs (b) 42 FMPs

Figure C9: Average Sharpe ratio of the SR portfolios based on simulation data of 250 weekly returns and 100 iterations by the t-distribution simulation and the DCC covariance for a rolling window of 150 weekly observations.

D Appendix: Empirical results

Table D1 contains the parameter estimates of the different dynamic models for the empirical study for both $K = 13$ and $K = 42$. To quantify time-variation, we also present the min-max ranges of the parameter estimates. We find similar outcomes compared to the simulation results. We observe again that average parameter estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{g}$ attain values in line with the outcomes of Engle (2002), Cappiello et al. (2006) and Engle & Kelly (2012). Moreover, the average $\hat{g}$ estimate is fairly small, indicating that the asymmetric information that can be captured is limited for the considered simulations. The min-max range indicates that the parameter estimates depend on time-variation. Note that the range for $\hat{\beta}$ widens when K becomes larger, indicating that the dispersion between observations over time is large.

Table D1: Average parameter estimates for the different DCC models for 4 February 1997 until 27 April 2021.

$K = 13$	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC
$\hat{\alpha}$	0.070	0.070	0.071	0.070	0.070	0.071	0.020
	[0.000, 0.348]	[0.000, 0.331]	[0.000, 0.328]	[0.000, 0.348]	[0.000, 0.331]	[0.000, 0.328]	[0.000, 0.054]
$\hat{\beta}$	0.789	0.807	0.795	0.789	0.807	0.795	0.770
	[0.003, 0.973]	[0.041, 0.974]	[0.040, 0.969]	[0.003, 0.973]	[0.041, 0.974]	[0.040, 0.969]	[0.381, 1.000]
$\hat{g}$							0.003
							[0.000, 0.028]
$K = 42$	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC
$\hat{\alpha}$	0.019	0.024	0.024	0.019	0.024	0.024	0.008
	[0.000, 0.068]	[0.000, 0.055]	[0.000, 0.083]	[0.000, 0.068]	[0.000, 0.055]	[0.000, 0.083]	[0.000, 0.025]
$\hat{\beta}$	0.805	0.836	0.804	0.805	0.836	0.804	0.400
	[0.000, 0.985]	[0.000, 0.998]	[0.000, 0.988]	[0.000, 0.985]	[0.000, 0.998]	[0.000, 0.988]	[0.000, 1.000]
$\hat{g}$							0.000
							[0.000, 0.002]

Note. The minimum and maximum parameter estimates are displayed in brackets below the average parameter estimates. The average estimates are close to zero but never exactly zero. The parameters are based on full dataset of 13 and 42 FMPs for 4 February 1997 until 27 April 2021.

Table D2: Average weight statistics for the different estimators for 4 February 1997 until 27 April 2021.

	Benchmarks			EWM shrinkage				DCC models							
$K = 13$	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Max	0.077	0.274	0.239	0.188	0.246	0.247	0.217	0.260	0.254	0.261	0.272	0.272	0.274	0.261	0.244
Min	0.077	-0.044	-0.036	-0.010	-0.016	-0.019	-0.021	-0.032	-0.011	-0.023	0.004	0.003	0.003	-0.035	-0.045
Median	0.077	0.039	0.038	0.045	0.033	0.033	0.041	0.037	0.034	0.035	0.060	0.060	0.060	0.037	0.039
SD	0.000	0.093	0.089	0.071	0.086	0.088	0.083	0.090	0.085	0.089	0.083	0.083	0.083	0.090	0.092
$K = 42$	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Max	0.024	0.101	0.108	0.045	0.111	0.119	0.107	0.097	0.135	0.095	0.169	0.179	0.169	0.093	0.102
Min	0.024	-0.012	-0.011	0.001	-0.004	-0.008	-0.010	-0.009	-0.002	-0.011	0.004	0.004	0.004	-0.008	-0.012
Median	0.024	0.023	0.022	0.024	0.017	0.016	0.022	0.022	0.016	0.022	0.011	0.010	0.011	0.022	0.022
SD	0.000	0.022	0.023	0.012	0.024	0.027	0.023	0.021	0.027	0.022	0.035	0.037	0.035	0.021	0.023
$K = 13$ **Cons**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Max	0.077	0.270	0.231	0.187	0.243	0.246	0.214	0.250	0.252	0.253	0.272	0.272	0.274	0.254	0.235
Min	0.077	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.003	0.003	0.000	0.000
Median	0.077	0.034	0.035	0.044	0.032	0.031	0.039	0.036	0.034	0.035	0.060	0.060	0.060	0.035	0.032
SD	0.000	0.085	0.082	0.069	0.083	0.084	0.078	0.084	0.084	0.085	0.083	0.083	0.083	0.084	0.082
$K = 42$ **Cons**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
Max	0.024	0.105	0.113	0.045	0.109	0.114	0.110	0.099	0.135	0.093	0.169	0.179	0.169	0.092	0.102
Min	0.024	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.004	0.004	0.000	0.000
Median	0.024	0.021	0.021	0.024	0.017	0.017	0.021	0.022	0.016	0.022	0.011	0.010	0.011	0.022	0.022
SD	0.000	0.022	0.023	0.012	0.024	0.026	0.022	0.021	0.027	0.021	0.035	0.037	0.035	0.020	0.022

Note. The 0.000 values are exactly zero due to the imposed short-selling constraint.

E Appendix: Robustness analysis

E.1 Moving window

We perform the same approach for the simulations as before to check for the robustness of the simulation results in Section 5 and its dependence on the estimation window. However, with estimation windows $RW = 300$ and $RW = 2000$. We discuss only the direct and indirect evaluation results of $K = 13$ for simplicity. Further details on how the simulations are conducted are discussed in Section 3. For all estimation windows, we end up with 250 covariance estimates.

Table E1 gives the direct evaluation results quantified by the loss function discussed in Engle et al. (2019). Table E2 shows the indirect GMV portfolio results. Overall, the outcomes are consistent with the outcomes of $RW = 150$ and signal robustness of the CC estimator.

The direct results in Table E1 shows that the best-performing model is dependent on the type of simulation. For the EWM-based simulations, we find that the static CC model is the most accurate in estimating the covariance matrix, similar to $RW = 150$. On the other hand, we observe that the dynamic models, the ADCC and DCC S, perform the best for the DCC-based simulations. Note that the static models are the second- or third-best model after these dynamic model. Overall, we find that the average losses for $RW = 150$ are smaller compared to $RW = 300$, indicating that implementing a window of 150 might give more beneficial sample properties.

Table E1: Average loss for different simulations for a rolling window of 300 weeks and 13 FMPs.

	Benchmarks		EWM shrinkage				DCC models							
	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
EWM normal	0.031	0.028	0.032	**0.027**	0.027	0.028	0.034	0.033	0.033	0.079	0.079	0.079	0.039	0.040
EWM t-dist	0.031	0.028	0.032	**0.026**	0.027	0.027	0.034	0.035	0.034	0.076	0.076	0.076	0.038	0.038
DCC normal	0.058	0.055	0.058	0.050	0.052	0.054	0.053	0.053	0.053	0.073	0.073	0.073	**0.048**	0.066
DCC t-dist	0.063	0.056	0.082	0.052	0.053	0.056	0.051	**0.048**	0.050	0.070	0.070	0.070	0.049	0.069

Note. Lowest average loss defined by the loss function of Engle et al. (2019) is indicated in bold. The loss is averaged over the 250 losses from the covariance estimates and 100 simulations for each estimator. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns. The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007.

Table E2: Average annualised performance of the GMV portfolios for a rolling window of 300 weeks and 13 FMPs.

		Benchmarks			EWM shrinkage				DCC models							
Normal EWM	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.544	2.066	0.653	0.635	0.675	**0.634**	0.635	0.638	0.662	0.662	0.662	0.727	0.727	0.727	0.674	0.681
SR	2.477	0.654	1.854	1.910	1.769	1.935	1.939	1.901	1.882	1.907	1.900	1.871	1.875	1.872	1.796	1.901
IR		0.015	-2.999	-3.171	-3.045	-2.900	-2.804	-3.189	-2.110	-1.765	-1.855	0.201	0.242	0.221	-2.727	-1.035
Turnover	1.459	0.364	1.414	1.494	1.163	1.372	1.338	1.513	3.587	3.679	3.592	3.412	3.421	3.421	3.565	10.687
***t*-dist EWM**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.559	2.039	0.653	0.647	0.688	**0.645**	0.647	0.648	0.681	0.681	0.679	0.740	0.740	0.740	0.695	0.677
SR	2.669	0.619	2.199	2.142	1.884	2.222	2.203	2.108	1.965	1.977	1.987	1.977	1.981	1.977	1.871	1.893
IR		-0.841	-1.204	-2.387	-3.871	-1.350	-1.552	-2.879	-3.054	-2.916	-2.874	-0.461	-0.421	-0.473	-3.849	-4.008
Turnover	1.484	0.364	1.473	1.536	1.187	1.385	1.356	1.532	3.938	4.105	3.981	3.593	3.599	3.606	3.781	10.178
Normal DCC	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.544	2.203	0.670	0.654	0.670	**0.647**	0.650	0.655	0.664	0.653	0.665	0.719	0.721	0.719	0.661	0.668
SR	2.663	1.314	2.157	2.152	2.554	2.242	2.226	2.189	1.983	2.069	2.016	2.424	2.413	2.422	2.007	2.236
IR		4.988	-0.028	-0.746	4.622	0.039	0.003	-0.277	-2.303	-1.757	-1.887	4.475	4.415	4.453	-2.294	0.746
Turnover	5.710	0.468	1.589	1.484	1.130	1.384	1.334	1.525	5.598	5.516	5.616	5.244	5.261	5.245	5.507	11.452
***t*-dist DCC**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.804	2.696	1.018	1.007	1.070	**0.997**	0.999	1.003	1.103	1.099	1.094	1.144	1.144	1.143	1.111	1.028
SR	1.962	1.307	1.726	1.775	1.751	1.696	1.684	1.778	1.724	1.714	1.750	1.551	1.550	1.554	1.741	1.775
IR		5.383	2.116	2.529	2.844	1.448	1.313	2.509	3.922	3.703	4.108	2.101	2.100	2.130	4.551	2.856
Turnover	5.710	0.468	1.917	1.523	1.122	1.453	1.397	1.620	4.427	4.176	4.388	3.721	3.724	3.722	4.320	13.765

Note. The SDs of the GMV portfolios over time are calculated as $SD = \sqrt{W_t' \Sigma_t W_t}$, and the SR is defined as $SR = \frac{W_t' \mu}{\sqrt{W_t' \Sigma_t W_t}}$. The SD is given in percentages.
The GMV portfolio performance is based on 250 weekly simulated observations and results are annualised by multiplying by 52.
The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns.

Tables E3 and E4 give the direct and indirect evaluation results for $RW = 2000$. Overall, we observe a shift in the best-performing models. We now find that the dynamic models outperform the static estimators. It indicates that the static models are better at capturing small sample properties while the dynamic ones would benefit from a larger sample.

For $RW = 2000$, average mean-variance losses per model are depicted in Table E3. The results show that the ADCC estimator is the clear winner. In contrast to the two smaller estimation windows, most of the dynamic models show higher accuracy than the static ones. We find that the average losses of the dynamic models for $RW = 2000$ are smaller than for $RW = 150$ and $RW = 300$, but the static estimators have no clear trend. This indicates that the large estimation window might only be useful for dynamic models while adding limited value for the static ones.

Table E3: Average loss for different simulations for a rolling window of 2000 weeks and 13 FMPs.

	Benchmarks		EWM shrinkage				DCC models							
	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
EWM normal	0.017	0.013	0.022	0.013	0.012	0.013	0.009	0.009	0.009	0.087	0.087	0.087	**0.008**	0.009
EWM t-dist	0.037	0.028	0.083	0.037	0.035	0.027	0.020	0.020	0.020	0.156	0.156	0.156	**0.018**	0.021
DCC normal	0.017	0.010	0.020	0.009	0.009	0.010	0.003	0.003	0.003	0.027	0.027	0.027	**0.003**	0.176
DCC t-dist	0.004	0.002	0.003	0.002	0.002	0.002	0.012	0.012	0.012	0.010	0.010	0.010	**0.001**	0.002

Note. Lowest average loss defined by the loss function of Engle et al. (2019) is indicated in bold. The loss averaged over the 250 losses from the covariance estimates and 100 simulations for each estimator. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns. The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007.

Table E4 shows the performance of the GMV portfolio for each covariance estimator for $RW = 2000$. The results differ from the ones in Table E3. The leading model is the GOGARCH; it is the best in three out of four simulations. Note that the ADCC is never the winner, while it attains the most accurate covariance estimates according to the average loss outcomes in Table E3. We observe that the dynamic models perform better than the static ones, in line with the direct evaluation. As such, applying a static model might be the best when working with a dataset with a rather short history and a long rolling window is not possible. Based on the simulation results, the performance of a dynamic model could be improved by enlarging the estimation window, using the large sample properties. Note that an estimation window of 2000 is difficult to realise in practice because many datasets do not have that much history in weeks.

Table E4: Average annualised performance of the GMV portfolios for a rolling window of 2000 weeks and 13 FMPs.

		Benchmarks			EWM shrinkage				DCC models							
Normal EWM	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.639	1.949	0.744	0.735	0.763	0.740	0.739	0.735	0.693	0.695	0.694	0.858	0.858	0.858	0.698	**0.692**
SR	1.122	0.258	0.986	1.043	0.786	1.242	1.216	1.041	1.029	1.046	1.040	1.129	1.129	1.127	1.044	0.980
IR		-0.797	0.459	1.476	-2.412	6.233	5.669	1.430	-0.162	0.341	0.152	3.636	3.642	3.612	0.438	-1.413
Turnover	1.673	0.104	1.646	1.749	1.338	1.542	1.490	1.759	0.787	0.783	0.797	0.560	0.559	0.558	0.837	2.615
***t*-dist EWM**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.812	2.426	0.907	0.866	1.083	0.861	0.858	0.865	0.805	**0.802**	0.805	0.912	0.912	0.912	0.809	0.811
SR	0.581	-0.939	-0.280	-0.174	-0.218	-0.260	-0.266	-0.173	-0.047	-0.042	-0.041	-0.240	-0.239	-0.239	-0.030	-0.164
IR		-6.369	-8.127	-9.111	-4.972	-8.873	-9.205	-9.101	-10.729	-10.257	-10.563	-5.388	-5.382	-5.381	-10.568	-11.816
Turnover	1.735	0.104	2.301	1.821	1.334	1.503	1.387	1.826	0.995	0.979	0.987	0.605	0.608	0.603	1.116	4.767
Normal DCC	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.584	2.529	0.835	0.832	0.835	0.827	0.824	0.832	0.811	0.811	0.810	0.862	0.862	0.862	0.817	**0.806**
SR	-0.038	-0.391	-0.056	-0.150	-0.109	-0.167	-0.162	-0.151	-0.130	-0.132	-0.131	-0.132	-0.133	-0.133	-0.128	-0.103
IR		-2.463	-0.546	-2.471	-1.409	-2.857	-2.723	-2.491	-2.217	-2.248	-2.243	-1.915	-1.916	-1.915	-2.162	-1.272
Turnover	4.127	0.156	1.527	1.622	1.130	1.537	1.492	1.621	3.626	3.619	3.619	3.372	3.372	3.372	3.747	9.561
***t*-dist DCC**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.336	1.199	0.357	0.356	0.370	0.353	0.354	0.356	0.422	0.422	0.422	0.372	0.372	0.373	0.344	**0.342**
SR	-0.125	-0.308	-0.033	-0.126	-0.075	-0.168	-0.163	-0.127	-0.196	-0.235	-0.228	-0.257	-0.235	-0.270	-0.154	-0.213
IR		-2.706	2.229	-0.236	0.991	-1.487	-1.318	-0.261	-1.650	-2.445	-2.292	-3.210	-2.703	-3.466	-1.165	-2.895
Turnover	2.360	0.156	1.440	1.650	1.206	1.527	1.489	1.662	7.585	8.568	9.016	2.167	2.287	2.316	2.313	4.820

Note. The SDs of the GMV portfolios over time are calculated as $SD = \sqrt{W_t' \Sigma_t W_t}$, and the SR is defined as $SR = \frac{W_t' \mu}{\sqrt{W_t' \Sigma_t W_t}}$. The SD is given in percentages.
The GMV portfolio performance is based on 250 weekly simulated observations and results are annualised by multiplying by 52.
The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns.

E.2 Short-selling constraints

Table E5 shows the constrained GMV portfolio performance in terms of various statistics for simulations with $K = 13$. The lowest SD is indicated in bold. We find that the results are consistent with the unconstrained GMV portfolios.

When comparing the results with the ones in Table 4, we find that the SD of the optimal GMV portfolio, based on the real covariance, is higher due to the addition of the constraints. The best-performing model is still the CC, being the top performer in three out of four simulations in terms of SD, consistent with the unconstrained results.

The standard deviations of the sample, EWM and unconstrained dynamic models are closer to the standard deviation based on the real GMV portfolio, suggesting better performance compared to the unconstrained case. However, their performance is still inferior to two of the shrinkage methods in all simulations. Note that in some cases, the improvement is even absolute, giving a lower SD for the constrained compared to the unconstrained portfolios. Especially the ADCC model shows improvement as it now often outperforms the DCC S model, which is the best estimator for the unrestricted GMV portfolios. Moreover, the turnover is lower for the constrained case as weights are no longer allowed to take on extreme values. This is beneficial for portfolio managers that have to adhere to a certain turnover maximum.

The results for the constrained GMV portfolio for $K = 42$ are given in Table E6. The leading estimator is still the NL model, equal to the outcomes of the unrestricted GMV portfolios in Table 5. Similarly to the outcomes for $K = 13$, we find that implementing the constraint has a shrinkage effect on most models and that the turnover decreases compared to the unconstrained case. Both are considered beneficial.

Also, we again observe that the SD of some models has decreased for the constrained optimisation and these estimators now attain lower SDs than for the GMV portfolios without short-selling constraints. For example, the sample, EWM, NL and GOGARCH estimation models where the constrained portfolios show a relatively large improvement compared to their unconstrained counterparts. This again signifies that the short-selling constraint works as a shrinkage mechanism and confirms the findings of Jagannathan & Ma (2003).

We find that the short-selling constraints are more effective for $K = 42$ than $K = 13$, leading to lower standard deviations for the estimated GMV portfolios. This signals that estimating a large dimension is more prone to error, such that shrinkage techniques can have a larger positive impact.

Table E5: Average annualised performance of the constrained GMV portfolios for a rolling window of 150 weeks and 13 FMPs.

		Benchmarks				EWM shrinkage						DCC models				
Normal EWM	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.426	1.642	0.494	0.468	0.497	**0.464**	0.464	0.470	0.495	0.493	0.493	0.571	0.570	0.568	0.489	0.480
SR	3.888	1.410	3.405	3.605	3.296	3.624	3.617	3.561	3.310	3.273	3.281	2.842	2.841	2.854	3.349	3.504
IR	-	5.421	-4.353	-5.768	-3.421	-7.768	-7.102	-5.809	-11.576	-13.087	-13.222	-10.433	-10.539	-10.595	-11.968	-2.871
Turnover	1.203	0.312	1.306	1.163	0.893	1.055	1.032	1.210	2.462	2.604	2.539	2.194	2.219	2.206	2.462	8.311
***t*-dist EWM**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.426	1.642	0.482	0.463	0.502	0.460	**0.459**	0.464	0.477	0.477	0.475	0.590	0.588	0.589	0.472	0.479
SR	3.888	1.410	3.661	3.738	3.344	3.727	3.706	3.700	3.674	3.613	3.651	2.928	2.938	2.936	3.683	3.592
IR	-	0.092	1.661	-0.447	-1.085	-2.256	-3.193	-1.014	2.204	0.453	1.275	-3.757	-3.651	-3.660	2.750	-1.982
Turnover	1.203	0.312	1.293	1.046	0.895	0.973	0.897	1.114	2.153	2.119	2.214	2.142	2.136	2.140	2.384	6.837
Normal DCC	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.373	1.549	0.449	0.437	0.454	**0.431**	0.433	0.436	0.450	0.452	0.450	0.516	0.515	0.515	0.440	0.451
SR	4.386	1.494	3.945	4.046	3.710	4.057	4.041	4.027	3.716	3.583	3.650	3.081	3.085	3.088	3.845	3.941
IR	-	3.238	2.381	0.561	-2.420	0.809	0.464	-0.045	-4.711	-5.748	-5.963	-4.638	-4.597	-4.452	-4.196	1.054
Turnover	5.440	0.416	1.274	1.078	0.879	0.994	0.971	1.158	4.291	4.111	4.304	2.779	2.750	2.776	2.844	7.383
***t*-dist DCC**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.373	1.549	0.462	0.450	0.487	**0.449**	0.449	0.453	0.465	0.466	0.467	0.556	0.555	0.556	0.463	0.462
SR	4.386	1.494	3.664	3.751	3.306	3.748	3.731	3.716	3.612	3.559	3.586	3.000	3.002	2.997	3.624	3.641
IR	-	8.170	-0.186	-0.366	1.673	-1.102	-1.955	-0.055	1.644	1.251	1.678	0.530	0.452	0.510	1.002	-0.573
Turnover	5.440	0.416	1.364	1.108	0.864	1.023	0.973	1.170	2.799	2.852	2.857	2.801	2.800	2.787	3.016	7.122

Note. The SDs of the constrained GMV portfolios over time are calculated as $SD = \sqrt{W_t' \Sigma_t W_t}$, and the SR is defined as $SR = \frac{W_t' \mu}{\sqrt{W_t' \Sigma_t W_t}}$. The SD is given in percentages.

The constrained GMV portfolio performance is based on 250 weekly simulated observations and results are annualised by multiplying by 52.

The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns.

Table E6: Average annualised performance of the constrained GMV portfolios for a rolling window of 150 weeks and 42 FMPs.

		Benchmarks			EWM shrinkage				DCC models							
Normal EWM	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.069	0.738	0.258	0.182	0.232	0.288	0.305	**0.181**	0.251	0.299	0.259	0.472	0.472	0.472	0.241	0.184
SR	6.830	1.098	2.008	2.814	2.918	4.001	3.933	2.844	2.071	3.509	2.380	2.586	2.611	2.556	2.457	2.786
IR	-	6.912	3.824	1.577	8.507	21.748	20.895	1.825	5.174	20.241	8.633	12.866	13.061	12.640	7.019	1.574
Turnover	1.367	0.364	1.679	1.395	0.648	1.162	1.068	1.407	2.344	2.257	13.066	2.015	2.050	1.978	3.386	3.860
***t*-dist EWM**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.069	0.738	0.231	0.177	0.295	0.307	0.326	**0.177**	0.235	0.342	0.251	0.495	0.496	0.495	0.227	0.184
SR	6.830	1.098	2.625	3.323	2.479	3.870	3.879	3.332	2.480	3.413	2.822	2.432	2.489	2.427	2.731	3.119
IR	-	2.137	1.737	2.260	2.413	10.814	11.299	2.316	-0.148	7.496	0.578	5.157	5.499	5.102	0.340	2.979
Turnover	1.367	0.364	1.993	1.651	0.645	1.160	1.058	1.653	2.646	2.422	11.741	1.985	2.071	2.007	3.234	4.305
Normal DCC	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.091	0.727	0.259	0.224	0.256	0.290	0.310	**0.221**	0.257	0.279	0.254	0.437	0.436	0.437	0.253	0.246
SR	5.543	1.116	2.836	3.149	2.714	4.129	3.989	3.151	2.713	3.842	3.114	2.819	2.849	2.794	2.846	2.796
IR	-	5.739	9.607	9.800	8.404	19.474	20.262	9.687	9.901	18.747	12.449	13.686	13.983	13.500	11.326	9.327
Turnover	3.467	0.468	1.934	1.702	0.719	1.251	1.124	1.740	3.078	3.539	8.074	2.638	2.663	2.604	3.692	5.112
***t*-dist DCC**	Real	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.091	0.727	0.271	0.238	0.319	0.316	0.333	**0.238**	0.266	0.325	0.267	0.459	0.459	0.459	0.258	0.257
SR	5.543	1.116	2.337	2.595	2.249	3.694	3.658	2.589	2.314	3.351	2.764	2.522	2.572	2.510	2.567	2.428
IR	-	1.603	-2.563	-3.831	2.292	8.422	9.692	-3.800	-5.845	4.829	-3.561	3.390	3.609	3.294	-5.042	-1.717
Turnover	3.467	0.468	1.961	1.653	0.701	1.254	1.128	1.701	3.246	3.307	11.152	2.813	2.904	2.824	3.926	4.774

Note. The SDs of the constrained GMV portfolios over time are calculated as $SD = \sqrt{W_t' \Sigma_t W_t}$, and the SR is defined as $SR = \frac{W_t' \mu}{\sqrt{W_t' \Sigma_t W_t}}$. The SD is given in percentages.
The constrained GMV portfolio performance is based on 250 weekly simulated observations and results are annualised by multiplying by 52.
The simulations are based on the first 550 weeks from 4 February 1997 until 14 August 2007. The different simulation setups are combinations of the EWM and DCC covariance estimate with a normal or t-distribution for the returns.

E.3 Subperiod analysis

In the subperiod portfolio analysis, we evaluate the performance of the covariance estimators in a low- and high-volatility period. The low-volatility period represents a period of expansion, while the high-volatility period corresponds to a recession period. We use the empirical dataset and base the split on the weekly returns in Figure 1a, where the economic event of the burst of the internet bubble in 2000-2002 causes a lot of uncertainty and a large deviation in the returns. This leads in high volatility. Therefore, we take the period from 6 January 1998 until 30 December 2003 as the high-volatility period. For the low-volatility period, we again base the split on Figure 1a. The period where the returns appear to deviate the least indicates low volatility, corresponding to data from 6 January 2010 until 30 December 2015. Both periods contain 313 observations and we apply a moving window of 150 weekly observations. We consider both FMP dimensions. We test whether the performance of the models deviates compared to the full dataset by constructing the GMV and constrained GMV portfolios. We consider the same performance measures as for the full empirical analysis.

The descriptive statistics for both periods are given in Tables E7 and E8 for $K = 13$ and $K = 42$, respectively. One can observe that the standard deviation in the high-volatility period is higher for all FMPs. Additionally, we find that the average returns are not following a certain trend for the two volatility period, but depend on the FMP.

Table E7: Descriptive statistics of the 13 FMP returns in percentages for a high-volatility and low-volatility period.

	High-volatility period					Low-volatility period				
	Mean	SD	SR	Kurtosis	Skewness	Mean	SD	SR	Kurtosis	Skewness
Market	0.115	2.977	0.039	3.799	0.044	0.265	2.087	0.127	7.517	-1.019
Within Value	0.062	0.332	0.186	3.843	0.316	0.010	0.178	0.059	3.280	0.170
Within Quality	0.030	0.251	0.121	3.783	0.182	0.019	0.120	0.161	3.725	0.186
Within Momentum	0.029	0.442	0.065	4.935	-0.732	0.041	0.249	0.163	3.387	-0.499
Within Size	0.051	0.492	0.104	4.850	-0.330	0.023	0.206	0.112	3.091	-0.257
Within LowVol	0.009	0.665	0.014	3.598	0.055	0.032	0.406	0.079	3.584	0.212
Within Axioma Market Sensitivity	0.022	0.646	0.034	3.428	0.073	-0.005	0.641	-0.007	5.253	0.098
Within Axioma Liquidity	0.015	0.322	0.046	4.183	0.165	0.029	0.224	0.130	3.178	-0.188
Across Value	0.030	1.107	0.027	5.098	0.053	-0.021	0.530	-0.040	3.263	-0.030
Across LowVol	0.019	2.426	0.008	4.252	0.339	0.049	1.133	0.043	3.863	0.301
Across Quality	0.110	0.917	0.119	4.212	-0.425	0.069	0.436	0.159	3.130	0.127
Across Momentum	0.108	1.516	0.071	5.953	0.040	0.109	0.907	0.120	5.073	0.109
Across Size	0.039	1.444	0.027	4.234	0.050	0.012	1.005	0.012	3.418	-0.256

Note. This table gives the statistical characteristics of the high- and low-volatility periods for 13 FMPs in percentages (Mean and SD). The high-volatility period is from 6 January 1998 until 30 December 2003 and the low-volatility period is from 6 January 2010 until 30 December 2015.

Table E8: Descriptive statistics of the 42 FMP returns in percentages for a high-volatility and low-volatility period.

	High-volatility period					Low-volatility period				
	Mean	SD	SR	Kurtosis	Skewness	Mean	SD	SR	Kurtosis	Skewness
Market	0.115	2.977	0.039	3.799	0.044	0.265	2.087	0.127	7.517	-1.019
Within Value	0.062	0.332	0.186	3.843	0.316	0.010	0.178	0.059	3.280	0.170
Within Quality	0.030	0.251	0.121	3.783	0.182	0.019	0.120	0.161	3.725	0.186
Within Momentum	0.029	0.442	0.065	4.935	-0.732	0.041	0.249	0.163	3.387	-0.499
Within Size	0.051	0.492	0.104	4.850	-0.330	0.023	0.206	0.112	3.091	-0.257
Within LowVol	0.009	0.665	0.014	3.598	0.055	0.032	0.406	0.079	3.584	0.212
Within Axioma Market Sensitivity	0.022	0.646	0.034	3.428	0.073	-0.005	0.641	-0.007	5.253	0.098
Within Axioma Liquidity	0.015	0.322	0.046	4.183	0.165	0.029	0.224	0.130	3.178	-0.188
AP	0.098	2.347	0.042	5.372	-0.126	-0.094	1.396	-0.068	4.428	-0.139
EU	-0.008	1.449	-0.005	3.759	-0.043	-0.080	1.373	-0.058	4.120	0.236
JP	0.037	2.991	0.012	4.366	0.233	-0.011	2.188	-0.005	6.545	-0.225
NA	0.011	0.870	0.013	4.848	0.456	0.054	0.634	0.085	4.472	-0.201
UK	-0.053	1.519	-0.035	3.706	-0.121	-0.017	1.106	-0.015	5.662	0.422
Automobiles & Components	-0.072	1.321	-0.054	10.481	-0.577	0.148	0.895	0.165	3.923	0.449
Banks	0.002	1.131	0.002	3.444	0.096	-0.054	0.674	-0.080	4.730	0.263
Capital Goods	0.019	0.728	0.026	3.638	0.102	0.009	0.571	0.015	3.854	-0.082
Commercial & Professional Services	-0.085	1.132	-0.075	3.904	-0.042	-0.053	0.596	-0.088	2.792	-0.188
Consumer Durables & Apparel	-0.059	0.945	-0.063	3.578	0.182	0.013	0.689	0.019	3.319	0.080
Consumer Services	-0.036	1.112	-0.033	4.748	-0.055	0.038	0.748	0.051	3.343	-0.026
Diversified Financials	0.034	1.201	0.028	3.609	-0.124	0.013	0.685	0.019	3.229	0.357
Energy	-0.002	1.967	-0.001	4.129	-0.006	-0.084	0.969	-0.086	5.783	-0.429
Food & Staples Retailing	-0.003	1.155	-0.003	3.334	-0.086	0.000	0.678	0.001	2.974	-0.045
Food, Beverage & Tobacco	-0.051	0.974	-0.052	3.568	0.308	0.001	0.547	0.001	2.828	0.101
Health Care Equipment & Services	0.116	1.529	0.076	4.515	0.120	0.006	0.775	0.007	4.804	-0.703
Household & Personal Products	0.027	1.404	0.019	5.269	-0.003	-0.042	0.675	-0.062	5.194	-0.074
Insurance	-0.010	1.058	-0.010	4.382	0.570	0.011	0.548	0.021	3.387	-0.170
Materials	-0.011	1.056	-0.010	5.459	-0.035	-0.065	0.740	-0.087	3.771	-0.199
Media	0.043	1.316	0.033	4.083	0.035	0.045	0.695	0.065	5.863	-0.641
Pharmaceuticals, Biotechnology & Life Sciences	0.064	1.298	0.049	4.396	0.107	0.113	0.612	0.185	3.551	-0.231
Real Estate	0.013	1.134	0.011	3.554	0.368	0.076	0.608	0.124	3.376	0.011
Retailing	-0.006	1.338	-0.005	3.480	0.148	0.058	0.685	0.085	2.889	0.075
Semiconductors & Semiconductor Equipment	0.134	2.055	0.065	3.660	0.119	-0.022	0.985	-0.023	4.075	0.173
Software & Services	-0.021	1.479	-0.014	3.469	-0.151	0.046	0.693	0.067	4.416	0.513
Technology Hardware & Equipment	0.076	1.289	0.059	3.608	0.018	-0.057	0.716	-0.080	3.526	-0.270
Telecommunication Services	-0.040	1.578	-0.025	3.752	0.089	-0.007	0.709	-0.010	3.662	-0.245
Transportation	-0.065	0.932	-0.070	5.832	-0.283	-0.013	0.661	-0.020	3.183	-0.064
Utilities	0.007	1.538	0.005	4.295	-0.246	0.011	0.716	0.015	5.116	0.253
Across Value	0.030	1.107	0.027	5.098	0.053	-0.021	0.530	-0.040	3.263	-0.030
Across LowVol	0.019	2.426	0.008	4.252	0.339	0.049	1.133	0.043	3.863	0.301
Across Quality	0.110	0.917	0.119	4.212	-0.425	0.069	0.436	0.159	3.130	0.127
Across Momentum	0.108	1.516	0.071	5.953	0.040	0.109	0.907	0.120	5.073	0.109
Across Size	0.039	1.444	0.027	4.234	0.050	0.012	1.005	0.012	3.418	-0.256

Note. This table gives the statistical characteristics of the high- and low-volatility periods for 42 FMPs in percentages (Mean and SD). The high-volatility period is from 6 January 1998 until 30 December 2003 and the low-volatility period is from 6 January 2010 until 30 December 2015.

The results of the GMV and constrained GMV portfolios for both periods and dimensions are given in Tables E9 and E10. We find that the optimal model for the GMV portfolio is the same for the high- and low-volatility period and consistent with the outcomes of the full dataset in Table 8. The CC estimator is the best for $K = 13$ and the NL model is the optimal one for the $K = 42$ dimension.

In terms of dynamic models for 13 FMPs, the results show that the DCC S is the most accurate in a high-volatility period and the ADCC estimator is the best in a low-volatility period. These outcomes are inconsistent with the full dataset results and indicate that the dynamic models are not robust for changing economic circumstances in this dimension. For $K = 42$, the best estimator is the GOGARCH in both periods. This is consistent with the results of the full dataset in Table 8.

Logically, we observe that SDs are higher for the high-volatility period compared to the low-volatility period in both dimensions. Moreover, the SRs attain higher values for the high-volatility period for most of the models. This holds for both dimensions. This could be because high volatility is often paired with high return. This comes from the risk-reward theory, which states that there is no return without risk.

For the constrained GMV portfolio, we find a deviation in the top estimator for the low-volatility period compared to the full dataset for $K = 13$. It gives the 1F estimator as the best instead of the CC. We also observe a deviating best model for $K = 42$, but for the high-volatility period. The optimal dynamic estimators are consistent with the results for the subperiod GMV portfolios for both $K = 13$ and $K = 42$.

A potential explanation for the differences for the two volatility periods could be that models react differently to constraints in low-volatility periods compared to high-volatility periods. This could result in another optimal model when splitting the two.

Lastly, when comparing the constrained and unconstrained portfolios, we find that the constrained portfolios outperform the unconstrained ones in the low-volatility period for all estimators. In such a period of expansion, most factors are doing well, such that non-negative weights are favorable. However, in high-volatility periods, we observe the reverse. The unconstrained GMV portfolios show lower SDs compared to the constrained portfolios for all estimators. This because in high-volatility periods, indicating periods of crisis, constraints limit the performance of the portfolios. Due to the bad performance of many factors, it would be more beneficial to go short. This is however not possible due to the constraints.

Table E9: Average annualised performance of the GMV and constrained GMV portfolios in high- and low-volatility periods for 13 FMPs.

	Benchmarks			EWM shrinkage			DCC models								
$K = 13$ **high vol**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	2.036	0.659	0.640	0.668	**0.637*****	0.638	0.639	0.653	0.648	0.657	0.838	0.837	0.837	0.672	0.665
SR	0.676	2.316	2.505	2.370	2.527	2.526	2.486	2.388	2.415	2.348	2.146	2.159	2.163	2.373	2.138
Turnover	0.364	1.399	1.370	1.062	1.239	1.227	1.578	4.358	4.149	4.230	3.355	3.376	3.363	4.039	9.773
$K = 13$ **low vol**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	1.629	0.514	0.505	0.525	**0.496*****	0.497	0.505	0.515	0.520	0.521	0.539	0.540	0.539	0.495	0.513
SR	1.660	1.426	1.643	1.947	1.797	1.791	1.743	1.311	1.469	1.399	1.942	1.913	1.919	1.362	1.506
Turnover	0.312	1.660	2.039	1.538	1.758	1.741	2.190	6.485	6.787	6.997	2.656	2.639	2.651	3.649	9.583
$K = 13$ **cons high vol**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	2.036	0.686	0.651	0.680	**0.649*****	0.651	0.652	0.657	0.656	0.662	0.838	0.837	0.837	0.675	0.676
SR	0.676	2.331	2.553	2.363	2.548	2.555	2.531	2.343	2.349	2.290	2.149	2.159	2.163	2.369	2.231
Turnover	0.364	1.274	1.241	1.012	1.145	1.132	1.444	3.924	3.840	3.905	3.349	3.373	3.358	3.627	8.398
$K = 13$ **cons low vol**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	1.629	0.499	0.493	0.523	0.492	**0.491*****	0.494	0.506	0.507	0.507	0.535	0.535	0.535	0.492	0.489
SR	1.660	1.771	1.838	2.040	1.908	1.922	1.908	1.604	1.615	1.596	2.001	1.983	1.989	1.670	1.848
Turnover	0.312	1.394	1.588	1.402	1.514	1.445	1.814	5.474	5.438	5.629	2.627	2.606	2.617	2.848	7.096

Note. A significant decrease of the SD of best-performing GMV portfolio compared to the SD of the GMV of the real covariance is indicated with a *,** and *** for a p-value below 0.1, 0.05 and 0.01, respectively. Using a two-sided test by Ledoit & Wolf (2011) and HAC standard errors.
K indicates the dimension of the FMP returns. The weekly performance is annualised by multiplying by 52 and a rolling window of 150 observations is used. The low-volatility period is from 6 January 2010 until 30 December 2015 and the high-volatility period is from 6 January 1998 until 30 December 2003. The portfolios are rebalanced for each new observations, thus for each week. The SD is given in percentages.

Table E10: Average annualised performance of the GMV and constrained GMV portfolios in high- and low-volatility periods for 42 FMPs.

	Benchmarks			EWM shrinkage				DCC models							
$K = 42$ **high vol**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	1.082	0.318	0.267	0.356	0.386	0.416	**0.266*****	0.318	0.415	0.347	0.727	0.720	0.727	0.345	0.301
SR	0.568	0.300	0.270	0.604	1.759	1.949	0.298	-0.145	1.403	0.146	1.441	1.494	1.440	0.003	0.100
Turnover	0.468	2.144	2.005	0.761	1.277	1.316	2.062	4.057	3.356	12.492	3.020	3.138	2.964	4.915	6.003
$K = 42$ **low vol**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.667	0.124	0.097	0.182	0.245	0.246	**0.097*****	0.197	0.217	0.185	0.419	0.425	0.419	0.206	0.100
SR	1.451	-0.381	-0.190	1.528	2.209	2.317	-0.172	0.309	1.993	0.681	2.151	2.115	2.117	0.548	-0.087
Turnover	0.416	1.484	1.449	0.719	1.508	1.460	1.479	3.773	4.466	12.239	2.056	2.180	2.062	5.151	3.923
$K = 42$ **cons high vol**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	1.082	0.303	**0.252*****	0.355	0.384	0.413	0.253	0.304	0.414	0.335	0.727	0.720	0.727	0.331	0.278
SR	0.568	0.400	0.360	0.603	1.774	1.936	0.413	-0.031	1.436	0.197	1.443	1.493	1.440	0.168	0.295
Turnover	0.468	1.876	1.716	0.764	1.260	1.278	1.797	3.586	3.319	10.387	3.021	3.141	2.967	4.223	4.915
$K = 42$ **cons low vol**	1/N	Sample	EWM	S	CC	1F	NL	DCC	DCC S	DCC NL	DECO	DECO S	DECO NL	ADCC	GOGARCH
SD	0.667	0.120	0.094	0.182	0.245	0.243	**0.094*****	0.192	0.215	0.177	0.419	0.425	0.419	0.199	0.095
SR	1.451	-0.267	-0.062	1.539	2.238	2.403	-0.061	0.146	2.080	0.506	2.149	2.121	2.120	0.495	0.062
Turnover	0.416	1.264	1.318	0.723	1.505	1.428	1.289	3.343	4.391	10.058	2.068	2.193	2.073	4.279	3.314

Note. A significant decrease of the SD of best-performing GMV portfolio compared to the SD of the GMV of the real covariance is indicated with a *,** and *** for a p-value below 0.1, 0.05 and 0.01, respectively. Using a two-sided test by Ledoit & Wolf (2011) and HAC standard errors.
K indicates the dimension of the FMP returns. The weekly performance is annualised by multiplying by 52 and a rolling window of 150 observations is used. The low-volatility period is from 6 January 2010 until 30 December 2015 and the high-volatility period is from 6 January 1998 until 30 December 2003. The portfolios are rebalanced for each new observations, thus for each week. The SD is given in percentages.