Dynamic portfolio selection with sector-specific regularization

Christian M. Hafner[a,b,*], Linqi Wang[a,c]

[a] *Louvain Institute of Data Analysis and Modelling in Economics and Statistics (LIDAM), UCLouvain, Belgium*

[b] *Institute of Statistics, Biostatistics, and Actuarial Sciences (ISBA), UCLouvain, Belgium*

[c] *Louvain Finance (LFIN), UCLouvain, Belgium*

Abstract

A new algorithm is proposed for dynamic portfolio selection that takes a sector-specific structure into account. Regularizations with respect to within- and between-sector variations of portfolio weights, as well as sparsity and transaction cost controls, are considered. The model includes two special cases as benchmarks: a dynamic conditional correlation model with shrinkage estimation of the unconditional covariance matrix, and the equally weighted portfolio. An algorithm is proposed for the estimation of the model parameters and the calibration of the penalty terms based on cross-validation. In an empirical study, it is shown that the within-sector regularization contributes significantly to the reduction of out-of-sample volatility of portfolio returns. The model improves the out-of-sample performance of both the DCC with nonlinear shrinkage and the equally-weighted portfolio.

Keywords: Dynamic conditional correlation, Cross-validation, Shrinkage, Industry sectors

JEL Classification: C14, C43, Z11

*Corresponding author.

Email addresses: `christian.hafner@uclouvain.be` (Christian M. Hafner), `linqi.wang@uclouvain.be` (Linqi Wang)

1. Introduction

Portfolio selection remains one of the central topics of empirical finance since its inception by Markowitz (1952). It is especially challenging in the context of portfolio selection with a large number of assets, where it is well known that naive implementation of the Markowitz (1952) results using sample means and variance-covariance matrices leads to poor out-of-sample performance due to unstable and ill-conditioned estimates. Remedies have been proposed such as shrinkage methods as in Ledoit and Wolf (2004) and Ledoit and Wolf (2017) and weight regularization as in DeMiguel et al. (2009a), Brodie et al. (2009) and Fastrich et al. (2015), which stabilize the ill-conditioned optimization problem. Statistical justifications for regularization techniques in a portfolio selection context are given by e.g. Puelz et al. (2015) and Fisher et al. (2020). Candelon et al. (2012) reformulate traditional shrinkage estimators in terms of a linear regression framework and impose a "double" shrinkage by adding additional weight penalties such as lasso or ridge, which further stabilizes portfolio weights and decreases portfolio turnover. Similarly, Ao et al. (2019) show an equivalent representation of the mean-variance optimization problem as an unconstrained regression, which they augment by lasso-type constraints to obtain sparsity and obtain consistency of mean and risk as both the number of assets and sample size increase. Furthermore, as shown by Jagannathan and Ma (2003), a no-short-sale constraint also serves to stabilize weights and regularize the optimization problem, as it is equivalent to shrinking large elements of the covariance matrix. They show empirically that short-sale constrained portfolios perform as well as those constructed using covariance matrices estimated from factor models and shrinkage estimators. DeMiguel et al. (2009a) and Fan et al. (2012) generalize this idea by introducing a gross-exposure constraint that bridges the gap between an unconstrained and no-short-sale constrained portfolio. Fan et al. (2012) give a theoretical justification for the empirical results in Jagannathan and Ma (2003).

Although the main objective of portfolio selection is typically formulated in terms of maximization of the Sharpe ratio, which corresponds to the tangency portfolio of Markowitz (1952), the estimation and prediction of asset mean returns have proven to be particularly difficult and noisy, see e.g. Best and Grauer (1991). Many studies, starting at least with Chan et al. (1999), therefore neglect the mean and concentrate on the minimization of the portfolio variance to obtain estimates of the global minimum variance portfolio (GMV). As shown by Jagannathan and Ma (2003), the GMV outperforms many competing approaches involving estimation of the mean in terms of Sharpe ratio. This has even been shown for the naive equally weighted portfolio by DeMiguel et al. (2009b). We therefore refrain from estimating the mean in this paper and consider the problem of estimating the GMV portfolio.

In a dynamic framework, it is important to take time-varying volatilities and correlations into account. Many alternative modelling strategies exist, including Bayesian dynamic linear models, see e.g. Puelz et al. (2020), but we follow a large part of the financial econometrics literature by implementing a GARCH-type model combined with dynamic conditional correlations (DCC) as in Engle (2002). These models can be used to predict volatilities and correlations up to a certain investment horizon, see Baillie and Bollerslev (1992) and Engle and Sheppard (2001). In large dimensions, issues arise for the estimation of DCC-type models that can be addressed by the composite likelihood method as in Pakel et al. (2020), combined with shrinkage of the sample covariance matrix that is used for correlation targeting as in Hafner and Reznikova (2012). See also Morana (2019) for an alternative approach to DCC estimation in high dimensions based on regularised semiparametric methods. For portfolio selection with many assets, Engle et al. (2019) implement a DCC-GARCH model with composite likelihood estimation and nonlinear shrinkage and show its superior out-of-sample performance with respect to benchmark portfolios, in particular the ones not taking the dynamics of volatilities and correlations into account.

Despite the vast literature on portfolio selection, not many studies have taken information about the industry sectors explicitly into account, although it is well known that this information helps to improve predictability of mean returns, volatilities and correlations. Companies within a given industry sector often compete in the same product market and co-move regarding product and technology innovations. They react similarly to permanent shifts in supply and demand conditions, as well as the regulatory environment. As the industry goes through expansions and contractions, companies' growth opportunities and investing and financing decisions are correlated. For example, Moskowitz and Grinblatt (1999) document the strong and persistent effects of industry components in stock returns, where industry momentum strategies are significant and more profitable than individual stock momentum strategies. Hou (2007) argue that the lead-lag effect is driven by an intra-industry phenomenon, where returns on big firms lead returns on small firms within the same industry, which is primarily caused by stock prices' slow response to negative information. Hence, there is information clustering at the industry level. Hong et al. (2007) find that stock markets react with a delay to information contained in industry returns about their fundamentals and that information diffuses only gradually across markets. Brito et al. (2018) forecast very large realized covariance matrices of returns using standard firm-level factors (e.g., size, value and profitability) and imposing additional sectoral restrictions in the residual covariance matrix. Kurose and Omori (2020) propose a multi-block equicorrelation structure to estimate multivariate stochastic volatility models in large dimensions and apply their model to an asset allocation exercise with sector as blocks. For portfolio selection, Chen et al. (2020) take a sector-specific structure into account by allowing to incorporate investor preferences with respect to industry sectors. They propose a sparse-group selection with the objective of being sparse across sectors but diversified within favored sectors. Fan et al. (2016) include sector specific information into a factor model where industry ETFs are included as additional factors to the CAPM or the three-factor model of Fama and French (1993).

Sparsity and diversification, although conflicting in nature, both are desirable properties of portfolios, since diversification allows to reduce portfolio weight variability, while sparsity keeps the number of invested assets and hence transaction costs under control. Moreover, both have a stabilizing effect on portfolio weights, if diversification is understood as shrinking towards a fixed weighting scheme. There are however many ways to introduce these properties into the selection procedure. The choice of Chen et al. (2020) is to impose sparsity between industry sectors and diversification within sectors via the sparse group lasso of Friedman et al. (2010) and Simon et al. (2013), an extension of the group lasso of Yuan and Lin (2006). It is not *a priori* clear whether this is the best strategy, as one might want to allow for sector-wide diversification as well, rather than investing in only a few sectors. Furthermore, the study of Chen et al. (2020) is using a rolling window updating mechanism. From an efficiency point of view, it would be preferable to have a genuine dynamic model that can be estimated using all available historical data, and that produces analytical forecasts of volatilities and correlations, as mentioned above.

In this paper, we adopt the perspective of a portfolio manager who optimally allocates funds across assets subject to a set of criteria motivated by investment preferences or constraints. Relevant examples of such criteria include restrictions on sectoral exposures, guided for instance by preliminary screening of relevant investable sectors (Chen et al. (2020)), or constraints on the distribution of portfolio weights across assets, e.g. by favouring sparse asset allocations (Puelz et al. (2020)), and over time by promoting stable asset allocation minimizing transaction costs (Hautsch and Voigt (2019)).

To this end, we propose a new portfolio selection procedure in a dynamic framework that explicitly accounts for these economically motivated criteria by regularising portfolio weights accordingly. More specifically, building on the DCC-GARCH model of Engle et al. (2019), which allows for the modelling of time-varying volatilities and correlations in large dimensions, we add restrictions on portfolio weights to reflect economically motivated criteria such as between- and within-sector diversification. We also consider penalty terms promoting sparsity in the resulting asset allocation and controlling for transaction costs by stabilizing variations in portfolio weights over time. Our proposed framework is modular as it allows to combine different penalty terms to obtain allocations satisfying multiple criteria simultaneously, e.g. combining between-/within-sector diversification with a cost penalty to obtain diversified portfolios with weights that are more stable over time. In addition, our framework can flexibly accommodate any dynamic input model to generate the portfolio weights, and is easy to implement as the penalty parameter controls the degree to which the resulting allocation will promote the imposition of the economically motivated criteria.

Our contribution is twofold: First, we introduce an algorithm for optimally choosing the penalization parameters in a data-driven way based on cross-validation in a dynamic context. The algorithm uses a partition of the training period into blocks, for which the combination of DCC-GARCH and regularization delivers portfolio variances that can be minimized with respect to the regularization parameters. These parameters are then used to construct portfolio weights in the out-of-sample evaluation period. The proposed algorithm generalizes classical cross-validation procedures, as e.g. in DeMiguel et al. (2009a), to take the time series structure of the data into account.

Our second contribution is a large scale empirical study to investigate how a regularization with respect to the sector structure helps to improve out-of-sample performance of portfolio return variances. We select stocks of the S&P Total index up to dimension 500, with daily frequency for the returns and a monthly frequency for portfolio re-balancing. We obtain several important findings. First, diversification is clearly more important within sectors than between sectors. Second, promoting sparsity of portfolio weights strongly contributes to reducing portfolio variances. This control for sparsity reduces negative weights and effectively moves the optimal portfolio towards a no short-sale constrained portfolio. The best performing model in large cross-sections combines within-sector diversification with a control for sparsity. The optimal portfolios significantly outperform benchmark portfolios such as the equally weighted or the DCC-GARCH portfolio without sector-wise regularization. Third, when a no-short-sale restriction is added, we observe that the best performing portfolios all feature the within-sector regularization. However, the difference with respect to the DCC-GARCH benchmark is smaller and not statistically significant for larger portfolios. Thus, in all considered scenarios, the penalty term for within-sector variability turns out to be important. This is in line with the results of Chen et al. (2020), extends them to a dynamic framework, and can be explained and motivated by the economic similarities of companies belonging to the same industry.

The remainder of this paper is organized as follows. The next section presents the model and the algorithm to estimate portfolio weights. Section 3 applies our methodology to stock returns of the S&P Total index, and Section 4 concludes. Some complementary results related to the empirical study are collected in an appendix.

2. The portfolio selection methodology

In this section we develop our portfolio selection methodology. First, some useful notation is introduced. Then, the DCC-GARCH model used for modelling volatilities and correlations is presented, including the way how monthly forecasts are generated from averaging daily forecasts. We then expose competing weight regularizations for the construction of global minimum variance portfolios, and finally present an algorithm for efficiently calibrating the penalization parameters using dynamic cross-validation.

2.1. Notation

We adopt the following notation throughout the paper. The subscript $i \in \{1, \ldots, N\}$ indexes the assets in the portfolio, where N denotes the total number of assets available; the subscript $k \in \{1, \ldots, K\}$ indexes the Global Industry Classification Standard (GICS) sectors, where K is the total number of sectors; and the subscript $t \in \{1, \ldots, T\}$ is the time index, where T is the sample size. Furthermore, we use the following notation:

- $y_{i,t}$: return for asset i at date t, stacked into $y_t := \big(y_{1,t}, \ldots, y_{N,t}\big)'$.
- $\mathcal{F}_{t-1}$: information set generated by $\{y_{t-1}, y_{t-2}, \ldots\}$.
- $\sigma^2_{i,t} := \text{Var}\big(y_{i,t}|\mathcal{F}_{t-1}\big)$: conditional variance of the i-th asset at date t.
- $\varepsilon_{i,t} := y_{i,t}/\sigma_{i,t}$: devolatized series at date t, stacked into $\varepsilon_t := \big(\varepsilon_{1,t}, \ldots, \varepsilon_{N,t}\big)'$.
- D_t : N-dimensional diagonal matrix with the i-th diagonal element being $\sigma_{i,t}$.
- $R_t := \text{Corr}\big(y_t|.\mathcal{F}_{t-1}\big) = \text{Cov}\big(\varepsilon_t|\mathcal{F}_{t-1}\big)$: conditional correlation matrix at date t.
- $H_t := \text{Cov}\big(y_t|\mathcal{F}_{t-1}\big)$: conditional covariance matrix at date t and $\text{Diag}\,(H_t) = D_t^2$.
- $C := \mathbb{E}\,(R_t) = \text{Corr}\,(y_t) = \text{Cov}\,(\varepsilon_t)$: unconditional correlation matrix.

2.2. Dynamic conditional covariance matrix estimation

In our empirical application, we use daily return data to forecast covariance matrices but we adopt the common practice of monthly re-balancing for the portfolio construction. This creates a mismatch between the frequency used for estimation and for forecasting, which we address with the average-forecasting approach used in De Nard et al. (2021) for DCC-GARCH model forecasting. Based on daily covariance estimates, daily forecasts are generated, and then they are iterated to deliver predictions for the horizon of interest. Specifically, at any portfolio construction date h, we retrieve the forecasts of the covariance matrices for all days in the upcoming month, namely $t = h, h+1, \cdots, h+L-1$, then average those L daily forecasts and use this averaged forecast to construct the portfolio composition at date h.

For the dynamics of the univariate volatilities, we choose a GARCH(1,1) process:

$$\sigma^2_{i,t} = \omega_i + \delta_{1,i} y^2_{i,t-1} + \delta_{2,i}\sigma^2_{i,t-1}, \tag{1}$$

where $(\omega_i, \delta_{1,i}, \delta_{2,i})$ are the parameters for asset i. For the evolution of the conditional correlation matrix over time, we assume that it is governed by a DCC(1,1) model:

$$Q_t = (1 - \alpha - \beta)\, C + \alpha \varepsilon_{t-1} \varepsilon'_{t-1} + \beta Q_{t-1}, \tag{2}$$

where (α, β) are the DCC model parameters, and C is the unconditional correlation matrix of ε_t which can be estimated in high dimensions using shrinkage techniques, see our discussion below. The matrix Q_t here can be interpreted as a conditional pseudo-correlation matrix or a conditional covariance matrix of devolatized residuals. For the reason that its diagonal elements, although close to one, are not exactly equal to one, we obtain the conditional correlation matrix and the conditional covariance matrix as

$$R_t := \mathrm{Diag}\,(Q_t)^{-1/2}\, Q_t\, \mathrm{Diag}\,(Q_t)^{-1/2}\,, \tag{3}$$

$$H_t := D_t R_t D_t, \tag{4}$$

and for estimation by quasi maximum likelihood, we assume a conditional normal distribution, i.e. $y_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, H_t)$.

We use approximate correlation targeting by first estimating C using shrinkage of the sample correlation matrix of ε_t, and then maximize the likelihood with respect to α and β. Due to the inconsistency argument of Aielli (2013) this is not an exact correlation targeting, which would require a modification of the DCC model. However, results are typically very similar, so for simplicity we stick to the original version of the DCC model.

With respect to the shrinkage method for C, we adopt the nonlinear shrinkage estimator of Ledoit and Wolf (2017) and Ledoit and Wolf (2020). This estimator is particularly appropriate to handle the problem of ill-conditioned sample correlation matrices arising in high dimensions, see also Zhao et al. (2020). Furthermore, the composite likelihood method developed by Pakel et al. (2020) provides a way to overcome the computational hurdle associated with the estimation of the DCC model in high dimensions. Engle et al. (2019) provide empirical evidence that a DCC model combining nonlinear shrinkage technique with an estimation carried out via composite likelihood method - called DCC-NL - has superior out-of-sample performance, compared to a DCC model with linear shrinkage, in a Markowitz portfolio setting with a large number of assets ($N \geq 100$). Note that the DCC-NL model, by regularizing the sample covariance matrix, also implicitly stabilizes portfolio weights. We will investigate in the empirical section whether additional explicit sector-wise penalties are further stabilizing the weights and improving the performance of the DCC-NL Markowitz portfolio.

To determine the average of the L forecasts of the conditional covariance matrix $H_{h+l} = D_{h+l}R_{h+l}D_{h+l}$, for $l = 0, 1, \ldots, L-1$, we propose a three-step approach where D_{h+l} and R_{h+l} are predicted separately, as described in the following.

For the multi-step ahead forecast of conditional univariate volatilities, we follow the approach of Baillie and Bollerslev (1992). The forecasts that minimize the mean square prediction error, for given parameters, are determined by the conditional expectations of conditional variances. Let us denote the l-step ahead forecast by $\hat{\sigma}^2_{i,h}(l) := \mathbb{E}\left[\sigma^2_{i,h+l}|\mathcal{F}_{h-1}\right]$, which for the GARCH(1,1) case can be written as

$$\hat{\sigma}^2_{i,h}(l) = \sum_{j=0}^{l-1} \omega_i \left(\delta_{1,i} + \delta_{2,i}\right)^j + \left(\delta_{1,i} + \delta_{2,i}\right)^l \sigma^2_{i,h}. \tag{5}$$

The forecasts of the diagonal matrix D_{h+l} can be constructed as $\hat{D}_h(l) = \mathrm{Diag}\left(\hat{\sigma}_{1,h}, \ldots, \hat{\sigma}_{N,h}\right)$.

Due to the nonlinearity of the DCC model, there is no exact analytical expression for the conditional expectations $\mathbb{E}\left[R_{h+l}|\mathcal{F}_{h-1}\right]$. However, Monte Carlo simulation evidence of Engle and Sheppard (2001) suggests that the approximation $\mathbb{E}\left[R_{h+l}|\mathcal{F}_{h-1}\right] \approx \hat{R}_h(l)$ has a negligible bias, where

$$\hat{R}_h(l) = \sum_{j=0}^{l-1} (1-\alpha-\beta)\, \hat{C}\, (\alpha+\beta)^j + (\alpha+\beta)^l R_h, \tag{6}$$

where $\hat{C}$ is a nonlinear shrinkage estimator of the unconditional correlation matrix of ε_t proposed by Ledoit and Wolf (2020).

By using the forecasts $\hat{D}_h(l)$ and $\hat{R}_h(l)$, the forecasts of the conditional covariance matrix using DCC-GARCH can be finally computed as $\hat{H}_h(l) := \hat{D}_h(l)\hat{R}_h(l)\hat{D}_h(l)$, for $l = 0, 1, \ldots, L-1$. Thus, to obtain the estimated conditional covariance matrix on portfolio construction day h, we average over the L forecasts:

$$\hat{H}_h := \frac{1}{L} \sum_{l=0}^{L-1} \hat{H}_h(l). \tag{7}$$

This matrix $\hat{H}_h$ will then be used for portfolio selection.

2.3. Global minimum-variance portfolio

We follow a large part of the empirical literature and consider the problem of estimating the global minimum variance (GMV) portfolio. Several reasons motivate our choice. First, it is more difficult to estimate means accurately than the covariance matrix of asset returns and the errors in estimating means have a larger impact on portfolio weights than the errors in the estimates of covariance matrices. Furthermore, as demonstrated by extensive empirical evidence, e.g. Jagannathan and Ma (2003), GMV portfolios have shown very good out-of-sample performance even in terms of criteria that take the mean into account, such as the Sharpe ratio.

The standard GMV problem for a given covariance matrix H_t is formulated as

$$\begin{aligned} &\min_{w} w' H_t w \\ &\text{s.t. } w'\mathbf{1} = 1, \end{aligned} \tag{8}$$

where $\mathbf{1}$ denotes a vector of ones with dimension $N \times 1$. It has the analytical solution

$$w_t = \frac{H_t^{-1}\mathbf{1}}{\mathbf{1}' H_t^{-1}\mathbf{1}}. \tag{9}$$

The natural strategy in practice is to replace the unknown H_t in Equation (9) by an estimator $\hat{H}_t$, yielding a feasible portfolio allocation strategy

$$\hat{w}_t = \frac{\hat{H}_t^{-1}\mathbf{1}}{\mathbf{1}' \hat{H}_t^{-1}\mathbf{1}}. \tag{10}$$

In order to take a sector structure into account, we modify the optimization problem above to the following

$$\begin{aligned} &\min_{w \in \mathbf{R}^N} w' H_t w + P_{\lambda t}(w) \\ &\text{s.t. } w'\mathbf{1} = 1, \end{aligned} \tag{11}$$

where $P_{\lambda t}(w)$ is a penalty function, to be specified in the following. The penalty function is parameterized by a parameter $\lambda \geq 0$, possibly a vector, that determines the strength of the imposed penalty. It will also depend on time, so that t is added as a subscript of the penalty function. The optimization problem reduces to the standard GMV problem if $\lambda = 0$, with analytical solution given in Equation (9). For $\lambda > 0$, Equation (11) does not have an analytical solution in general, but can typically be written as a convex programming problem for which efficient numerical algorithms are available. In our numerical implementations, we use the R package CVXR for disciplined convex programming, see Fu et al. (2020).

We denote the K industry sectors by $G_1, G_2, \ldots, G_K$, and by p_k the number of stocks in the k-th sector G_k, with average weight $m_{kt} = \sum_{i \in G_k} w_{it}/p_k$. In this paper, we conduct a comparison of the following candidates for the penalty function $P(w)$.

1. WITHIN (W): penalize weight differences of assets within the same sector

$$P_{\lambda t}(w) = \lambda \sum_{k=1}^{K} \sum_{i \in G_k} |w_{it} - m_{kt}|. \tag{12}$$

2. BETWEEN (B): penalize weight differences of assets between sectors

$$P_{\lambda t}(w) = \lambda \sum_{k=1}^{K} |m_{kt} - \frac{1}{N}|. \tag{13}$$

3. GROUP LASSO (Gl):

$$P_{\lambda t}(w) = \lambda \sum_{k=1}^{K} |m_{kt}|. \tag{14}$$

4. SPARSE (S):

$$P_{\lambda t}(w) = \lambda \sum_{i=1}^{N} |w_{it}|. \tag{15}$$

5. COST (C):

$$P_{\lambda h}(w) = \lambda \sum_{i=1}^{N} |w_{ih} - w_{i,h-1}^{+}|, \quad h \geq 2, \tag{16}$$

   where $w_{i,h}^{+} := w_{i,h} y_{i,h}/(1 + w_h' y_h)$ with $y_{i,h}$ being the return of asset i between two consecutive re-balancing dates $h-1$ and h.

The first two penalty functions encourage weight similarities of assets belonging to the same sector ("within"), and across different sectors ("between"), respectively. If one believes that the dynamic properties of stock returns are more homogeneous within a sector than between different sectors, then it would intuitively make sense to obtain better GMV portfolio results with the penalty Equation (12) than with the penalty Equation (13). This is however an empirical question that we are going to address in the next section. Note that an alternative specification of the within and between penalties could use an L_2 norm rather than L_1, or a combination of both as in the elastic net of Zou and Hastie (2005).

The third penalty function displayed in Equation (14) promotes sparsity of portfolio weights between sectors via a lasso-type penalty as introduced by Yuan and Lin (2006). The fourth penalty function in Equation (15) promotes global sparsity of portfolio weights via a lasso-type penalty as introduced by Tibshirani (1996) in a regression context. It is well known that as the strength of this penalty increases, the solution converges to the short-sale constrained portfolio, see DeMiguel et al. (2009a). This term thus controls the overall degree of allowed short sales.

Finally, the penalty function in Equation (16) promotes a reduction of variability of portfolio weights from one period to the next, and hence intends to reduce transaction costs, similar to Chen et al. (2020) and Hautsch and Voigt (2019). The vector w^+_{h-1} is the allocation of portfolio weights right before re-balancing at time $h-1$. As the strength of this penalty increases, the optimal portfolio allocation, i.e. the number of shares per asset, would be decided at the beginning of the evaluation period. The portfolio weights would vary during the evaluation period only because of changing asset prices. We note that the transaction costs are not explicitly taken into account in the portfolio construction and performance evaluation, we instead control for the weight variation over time and use it as an indicator on the relative performance in terms of transaction costs for different portfolios.

We can combine the above penalties in various forms. For example, if we want to penalize both the weights differences between and within sectors, we construct

$$P_{\lambda t}(w) = \lambda_1 \sum_{k=1}^{K} |m_{kt} - \frac{1}{N}| + \lambda_2 \sum_{k=1}^{K} \sum_{i \in G_k} |w_{it} - m_{kt}|. \tag{17}$$

where now the penalty parameter is a vector of two components, $\lambda = (\lambda_1, \lambda_2)'$. We will call this penalty "BW" for between- and within-sector regularization. The BW portfolio shrinks the asset allocation towards the equally weighted portfolio, which has often been chosen as a benchmark for investment strategies and was shown in DeMiguel et al. (2009b) to perform well out of sample. Analogously, we call SW and CW penalties that combine the within term with sparsity and transaction cost control, respectively. Furthermore, combining the sparse and group lasso penalties, we obtain the sparse group lasso regularization of Friedman et al. (2010). See also Babii et al. (2020) and Babii et al. (2021) who provide theoretical results for the sparse group lasso in a time series framework.

We first investigate individually how the above regularization terms perform compared to the DCC-GARCH estimation without penalty. We will then study whether all three sector-wise penalization terms are equally important and whether combining sector-wise penalties with global controls for sparsity and cost efficiency further improves the performance.

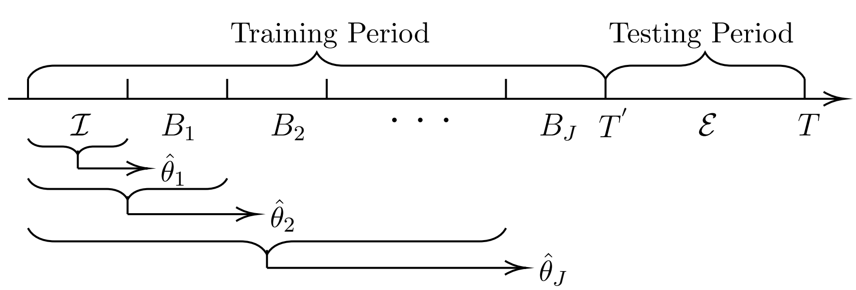
2.4. The dynamic portfolio selection algorithm

The previous section presented the objective function for portfolio optimization including various penalty terms depending on the sector structure. The problem is how to choose the parameters λ that determine the strength of the corresponding penalties in a dynamic framework. In the following we propose an algorithm that allows to combine estimation of the dynamic conditional covariance matrix with a cross-validation method to select the penalty parameter. We call θ the vector of all unknown parameters contained in the DCC-GARCH model.

1. Divide the data into a training set $\mathcal{T} = [1, T']$ and a evaluation set, $\mathcal{E} = (T', T]$.

2. Divide $\mathcal{T}$ into an initialization period $\mathcal{I}$ and J blocks B_j: $\mathcal{T} = \mathcal{I} \bigcup_{j=1}^{J} B_j$. The length of each block is m.

3. Estimate a DCC-GARCH model on $\mathcal{I}$ to get $\hat{\theta}_1$ and the conditional covariance matrix forecasts of $H_t(\hat{\theta}_1)$, on $\mathcal{I} \bigcup B_1$ to get $\hat{\theta}_2$ and $H_t(\hat{\theta}_2)$, etc. until $\mathcal{I} \bigcup_{j=1}^{J-1} B_j$ to get $\hat{\theta}_J$ and $H_t(\hat{\theta}_J)$.



4. Fix a $\lambda > 0$ and calculate the optimal portfolio allocation for each re-balancing date t in B_1 as

$$w_t(\lambda) = \arg\min_{w} (w' H_t(\hat{\theta}_1) w + P_{\lambda t}(w)), \quad t \in B_1, \tag{18}$$

with $P_\lambda(w)$ defined for different cases from Equation (12) to Equation (17), and the portfolio variance over B_1 is

$$Q_1(\lambda) = \frac{1}{m} \sum_{t \in B_1} (w_t(\lambda)' y_t)^2. \tag{19}$$

5. Repeat step 5 for $j = 2, \ldots, J$ to get $Q_2(\lambda), \ldots, Q_J(\lambda)$, and calculate $Q(\lambda) = \frac{1}{J} \sum_{j=1}^{J} Q_j(\lambda)$.

6. Repeat steps 5 and 6 for different values of λ on a grid, and calculate

$$\lambda^* = \arg\min_{\lambda} Q(\lambda). \tag{20}$$

7. Estimate a DCC-GARCH model on the training set to get $\hat{\theta}$ and the conditional covariance matrix forecasts of $H_t(\hat{\theta})$.

8. For $t \in \mathcal{E}$, using the optimal penalty parameter λ^* of step 6 and the forecasts $H_t(\hat{\theta})$ of step 7, calculate the optimal portfolio allocation as

$$w_t = \arg\min_{w} (w' H_t(\hat{\theta}) w + P_{\lambda^* t}(w)), \quad t \in \mathcal{E}. \tag{21}$$

9. Calculate the out-of-sample portfolio variance

$$V = \frac{1}{T - T'} \sum_{t \in \mathcal{E}} (w_t' y_t)^2. \tag{22}$$

10. Compare V with benchmarks such as the classical GMV without sector regularization ($\lambda = 0$), and one that imposes weight equality across assets ($1/N$).

The proposed algorithm nests a block-version of time series cross-validation (steps 4 to 6) into the portfolio selection problem. We use blocks rather than single observations as test sets, because the forecasting targets are variances, not observations themselves. The choice of the block-size m should balance the estimation uncertainty of out-of-sample volatilities $Q_j(\lambda)$ in Equation (19) with that of their expectation, $Q(\lambda)$. Note that in this form, the proposed portfolio selection algorithm is novel. In the following empirical analysis we investigate its performance when applied to a large asset data set.

3. Empirical application

In our empirical analysis, we investigate the question of how adding a sector-specific regularization structure will impact portfolio performance with respect to benchmark portfolios. We will examine the out-of-sample performance of Markowitz portfolios based on combining a dynamic conditional correlation model with intra- and inter-sector penalization using historical stock data.

3.1. Data description

We use historical data on daily returns for the S&P Total Index component stocks with data available since at least January 03, 2000. The selected stocks belong to 10 different Global Industry Classification Standard (GICS) sectors which have at least 50 companies in the S&P Total Index. Therefore, Communication Services is the only sector in the S&P Total Index which is not covered in the investment universes considered in our analysis. The sample period spans from January 03, 2000 to March 31, 2021, with a total of 5304 observations per stock. For simplicity, we adopt the common convention that 21 consecutive trading days constitute one month. In this manner, the cross-validation period for the penalization terms ranges from January 17, 2007 to July 13, 2016, resulting in a total of 112 months. The out-of-sample period is composed of the remaining 56 months, starting on July 14, 2016 and ending on March 16, 2021.

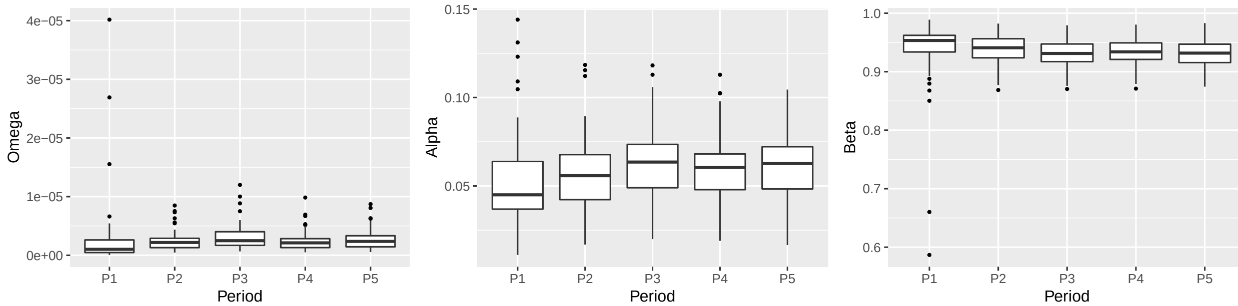
To be consistent with common practice, all the portfolios are re-balanced monthly to achieve a lower turnover and avoid an unreasonable amount of transaction costs. For simplicity, we assume the portfolio weights are fixed from one day to the next within a month. This will reduce the transaction cost but not eliminate it as the number of shares over time does not remain constant. At any investment date, the conditional covariance matrix is estimated using the 3-step approach for the DCC-GARCH model described in Section 2.2. The availability of daily returns ensures a sufficient estimation precision of the DCC-GARCH model at each step of the proposed algorithm. We consider the following portfolio sizes: $N \in \{50, 100, 250, 500\}$, composed respectively of the first 5, 10, 25, and 50 largest market capitalization stocks – evaluated on March 31, 2021 – from the S&P Total Index component companies belonging to the 10 GICS sectors. In our application, each investment universe is nested by the larger ones, e.g. our investment universe with $N = 50$ is a subset of the one with $N = 100$.

3.2. Dynamic covariance matrix estimation

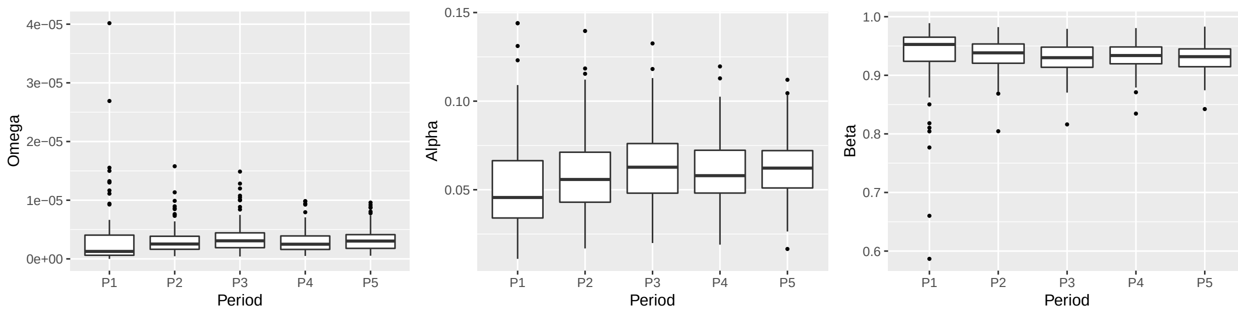
We follow the procedure described in Section 2.2, where a univariate GARCH(1,1) model is adopted for the conditional variances of asset returns, and a DCC(1,1) model is fitted to obtain time-varying conditional correlation matrices using historical daily returns. To address the challenges arising in large dimensions, we use nonlinear shrinkage for the unconditional correlation matrix targeting and the estimation is carried out via the composite likelihood method. In order to gain insights on the robustness of the parameter estimates over time and on the effect of sample size, we consider five samples of increasing size. Each period starts on 2000-01-04, and ends respectively on 2007-01-12, 2009-07-09, 2011-11-04, 2014-03-12, and 2016-07-12. In this manner, the sample size for estimation starts from roughly 1800 observations and then increases by around 600 observations each period. These periods will also serve as blocks for our cross-validation algorithm introduced in Section 2.4, so that after the initialization period (Period 1) with 1764 observations, there are $J = 4$ blocks, each consisting of $m = 588$ observations.

Figure 1 summarizes the GARCH estimates via box-plots for the different sub-periods and dimensions. It can be observed that the estimates for all parameters (ω, α, β) are quite stable over time, especially after Period 2 where the effect of adding more data becomes marginal. Although there are some outliers, the estimates of ω and α are close to zero for most of the assets, while the estimates for the persistence coefficient β are close to one, which are typical findings for financial asset return data. Parameter estimates for the DCC model are reported in Table 1, where again estimates of α are close to zero and estimates of β are close to 1, indicating high persistence in the dynamic correlations. Furthermore, similar to the volatility models, parameter estimates stabilize as the sample size increases. In addition, all parameter estimates are significantly different from zero using estimated asymptotic standard errors.

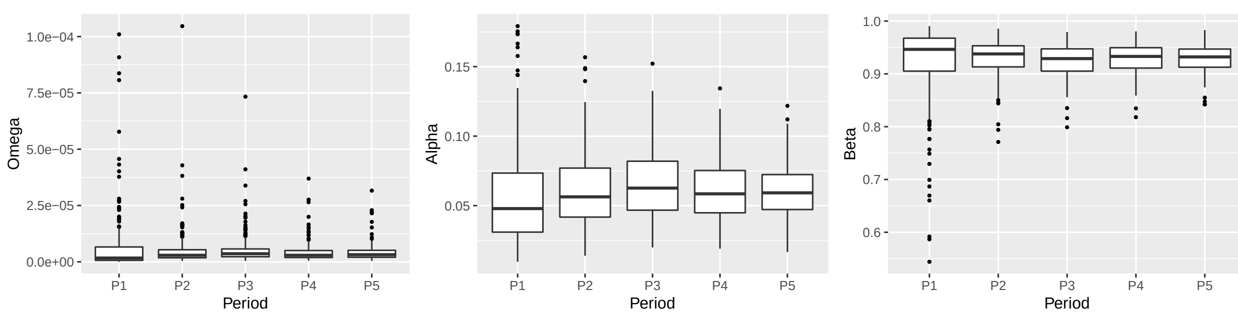
After estimation, we have used various diagnostics such as Portmanteau tests for autocorrelation in standardized residuals, as well as their squares and cross-products, which do not yield any evidence against the chosen model specifications. To save space, we do not



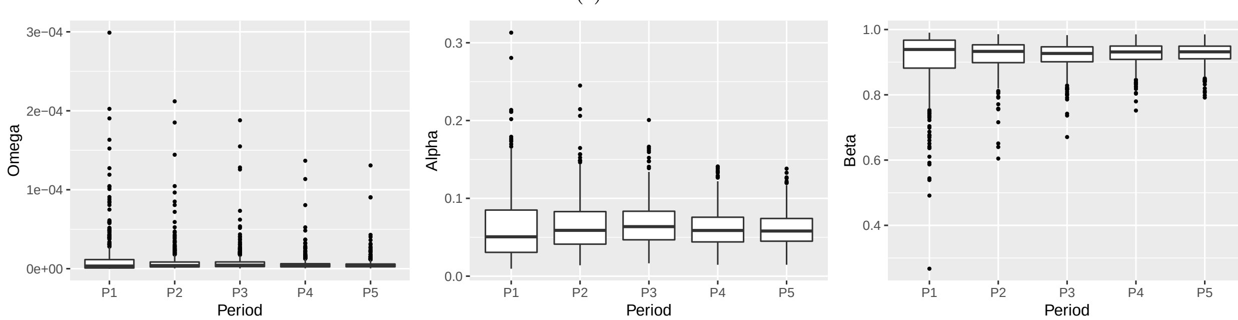
(a) N=50



(b) N=100



(c) N=250



(d) N=500

Figure 1: Boxplots of GARCH(1,1) parameter estimates for the five periods of augmenting sample sizes (P1 to P5), and for four investment universes $N \in \{50, 100, 250, 500\}$. The sub-periods considered in the estimation are: P1 from 2000-01-04 to 2007-01-12; P2 from 2000-01-04 to 2009-07-09; P3 from 2000-01-04 to 2011-11-04; P4 from 2000-01-04 to 2014-03-12; P5 from 2000-01-04 to 2016-07-12.

Table 1: DCC parameter estimates for the five periods of augmenting sample sizes (Period 1 to Period 5), and for three investment universes $N \in \{50, 100, 250, 500\}$.

	$N = 50$		$N = 100$		$N = 250$		$N = 500$	
	α	β	α	β	α	β	α	β
Period 1	0.0126	0.9797	0.0116	0.9815	0.0109	0.9825	0.0095	0.9856
	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0001)	(0.0003)	(0.0004)	(0.0015)
Period 2	0.0145	0.9782	0.0137	0.9803	0.013	0.9821	0.0116	0.9844
	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0001)
Period 3	0.0196	0.9711	0.0192	0.9725	0.0178	0.9754	0.0156	0.9794
	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0001)
Period 4	0.018	0.9737	0.0171	0.9762	0.0167	0.9773	0.0149	0.9803
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Period 5	0.0183	0.9735	0.0168	0.9766	0.0167	0.9773	0.0146	0.9807
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)

Note: The numbers in parentheses are asymptotic standard errors of the corresponding parameter estimates. The sub-periods considered in the estimation are: Period 1 from 2000-01-04 to 2007-01-12; Period 2 from 2000-01-04 to 2009-07-09; Period 3 from 2000-01-04 to 2011-11-04; Period 4 from 2000-01-04 to 2014-03-12; Period 5 from 2000-01-04 to 2016-07-12.

report these diagnostics here, but keep them available upon request. We therefore continue with the construction of GMV portfolios under alternative types of regularization.

3.3. GMV portfolio performance without short-sale constraints

We study the performance of investment strategies including sector-specific regularization with respect to benchmark portfolios in the context of the GMV portfolio where the most important performance measure is the out-of-sample volatility. Since the GMV portfolio is designed to minimize the variance rather than to maximize the Sharpe Ratio, all portfolios are primarily evaluated by the magnitude of the volatility reduction. We also report the Sharpe ratio to demonstrate that GMV portfolios typically perform well even in terms of criteria that include average out-of-sample returns (see e.g. Jagannathan and Ma (2003)).

We include the following eleven portfolios in our empirical analysis.

1. EQ: the equally-weighted portfolio with weights given by $1/N$, which is a standard benchmark advocated by DeMiguel et al. (2009b).
2. NP (no penalty): the DCC-GARCH with nonlinear shrinkage and maximum composite likelihood.
3. W: as NP but with additional penalty for within-sector variation.
4. B: as NP but with additional penalty for between-sector variation.
5. S: as NP but with additional L_1-norm penalty to promote sparsity.
6. Gl: as NP but with additional L_1-norm penalty to promote sparsity in the selected sectors.
7. C: as NP but with additional penalty to control transaction costs.
8. BW: as W but with additional penalty for between-sector variation.
9. SW: as W but with additional L_1-norm penalty to promote sparsity.
10. SGl: as Gl but with additional L_1-norm penalty to promote sparsity within the selected sectors.
11. CW: as W but with additional penalty to control transaction costs.

In order to obtain the optimal value of the penalization parameter(s) λ, we solve the optimization program specified in Equation (20) via grid search. The resulting optimal values of $\log(\lambda)$ are reported in Table 6 of the appendix. We then construct the asset allocation for the eleven portfolios considered on each monthly re-balancing date, and the portfolio weights are assumed to remain constant within the month. The out-of-sample portfolio performance measurements are computed with daily return data and then annualized by following the convention of multiplying the average return by 252 and the standard deviation by $\sqrt{252}$. The (annualized) Sharpe ratio is obtained as the ratio of average return and standard deviation. The statistics, with volatility in percentage points, are presented in

Table 2: Out-of-sample annualized portfolio standard deviation (in percentage points) and Sharpe ratio

	W	B	S	Gl	C	BW	SW	SGl	CW	NP	$1/N$
						N=50					
SD	17.410**	18.013	16.936***	18.049	16.909***	16.662***	16.933***	**16.449******	16.884***	18.191	19.868
SR	0.819	0.908	0.955	0.984	0.954	1.031	0.898	0.973	0.899	0.860	0.729
						N=100					
SD	17.231***	18.854	**16.800******	18.871	16.833***	17.073***	16.886***	16.806***	16.860***	18.755	20.269
SR	0.703	0.917	1.044	0.884	1.035	0.722	0.795	1.048	0.784	0.872	0.648
						N=250					
SD	18.650***	19.579	**17.568****	19.634	17.573**	19.060	17.592**	17.572**	17.581**	19.269	20.539
SR	0.644	0.644	0.919	0.680	0.919	0.599	0.898	0.920	0.894	0.675	0.573
						N=500					
SD	18.825**	19.657	17.636***	19.604	17.632***	19.464	**17.577******	17.635***	17.588***	19.464	21.511
SR	0.777	0.682	0.942	0.682	0.914	0.723	0.900	0.942	0.865	0.699	0.518

Note: Significant outperformance of the portfolios over the NP portfolio in terms of SD is denoted by asterisks: $*$, $**$, and $***$ indicate significance at the 10%, 5%, and 1% level respectively. All portfolios outperform the equally-weighted benchmark at a significance level of 5% or lower.

Table 2. The best performing portfolio with respect to the volatility criterion is marked in bold.

Focusing first on the three portfolios with sector-wise regularizations, namely the W, B, and Gl portfolios, we conclude that the W portfolio has the highest contribution to reducing portfolio volatility. While the combination of within- and between-sector (BW) regularization is further improving the performance of the W portfolio in terms of volatility for the moderate investment universes ($N \in \{50, 100\}$), adding additional between-sector penalization is deteriorating the volatility performance of the W portfolio in larger dimensions (i.e. $N \in \{250, 500\}$). Additionally, the observation that shrinking different sector weights towards equivalence lacks in efficacy is consistent with the intuition that assets belonging to different sectors may have quite different characteristics and tend to be less correlated. Promoting within-sector diversification is more important than between-sector diversification or concentrating the allocation on a few sectors. Furthermore, promoting sparsity concentrates the weight allocation on fewer assets, and adding a cost penalization term stabilizes the weight evolution of each asset over time. We can observe from Table 2 that both approaches contribute to better portfolio performance in terms of volatility.

In addition, for each of the portfolios with regularizations, we conduct the test of equality in standard deviations with respect to the two benchmark portfolios, with a two-sided p-value obtained by the prewhitened HAC_{PW} method described in Ledoit and Wolf (2011). It can be concluded that the nine regularized portfolios outperform the equally-weighted benchmark and the outperformance is statistically significant at all conventional levels. Furthermore, controlling for "within" ,"sparse" and "cost" significantly improves the portfolio volatility (at the 5% level) when considering the NP portfolio as benchmark. We also test whether the regularized portfolios are significantly outperforming each other. The p-values are reported in Table 8 and Table 9 of the appendix. We observe that the "within" portfolio is not significantly outperformed by other penalized portfolios at the 5% level for $N \in \{100, 250\}$ and at the 1% level for $N = 500$. However, the "between" and "group sparse" portfolios are almost always significantly outperformed for $N \in \{100, 250, 500\}$ at the 1% level. This confirms that promoting within-sector diversification is more beneficial than between-sector diversification or concentrating the allocation on a few sectors. In addition, portfolios controlling for sparsity and transaction cost are rarely significantly outperformed, especially in moderate and large investment universes.

In terms of the Sharpe ratio, we conclude that adding a sparsity regularization and a cost penalty term contributes to an improved portfolio performance. Additionally, the "between" regularized portfolio (B) and the portfolio focusing on a few selected sectors (Gl) display an improvement in performance relative to both benchmarks in the small investment universes ($N \in \{50, 100\}$). The outperformance in terms of Sharpe ratio tends to diminish for portfolios with a larger number of assets ($N \in \{250, 500\}$). We note that if we changed our performance target from GMV portfolio to Sharpe ratio maximization, including additional features such as a momentum signal for expected returns could further improve the performance, see e.g. De Nard et al. (2021).

We also provide some descriptive statistics for the portfolio weights w_t over time. In each holding period, namely one month, we compute the following four characteristics:

- Min: Minimum weight of all assets in the portfolio.
- Max: Maximum weight of all assets in the portfolio.
- SD: Standard deviation of weights of all assets in the portfolio.
- MAD-EW: Mean absolute deviation from equal weights, $1/N$.
- MDiv-Sec-W: Mean diversification within sectors, computed as

$$\text{MDiv-Sec-W} = \frac{1}{T - T'} \sum_{t=1}^{T-T'} \sum_{k=1}^{K} \sum_{i \in G_k} |w_{it} - m_{kt}|.$$

- MDiv-Sec-B: Mean diversification between sectors, computed as

$$\text{MDiv-Sec-B} = \frac{1}{T - T'} \sum_{t=1}^{T-T'} \sum_{k=1}^{K} |m_{kt} - \frac{1}{N}|.$$

- No. Active: Number of stocks with absolute weight larger than $1\% * 1/N$.

For each characteristic, we then report the average statistics over the out-of-sample period, as summarized in Table 3. We observe that the portfolios with the least dispersed weights are those including within-sector regularization, either alone or in combination with one of the other regularization terms. For each dimension, portfolios with "within", "sparse" and "cost" penalizations contribute the most to reducing the distance to the equally weighted portfolio as measured by the MAD-EW criterion. Additionally, we observe that portfolios with significantly improved volatilities (i.e. W, S, C) have overall smaller values for the MDiv-Sec-W statistic indicating higher within-sector diversification, but larger values for the MDiv-Sec-B statistic suggesting lower between-sector diversification. Conversely, portfolios with poor out-of-sample performance (i.e. B, Gl) exhibit higher between-sector diversification and lower within-sector diversification. This observation is consistent with the conclusion that promoting within-sector diversification is more important than between-sector diversification or concentrating the allocation on only a few sectors. Furthermore, the number of active positions indicates that the portfolios with sector-wise regularizations (i.e. W, B, Gl) are diversified across all available assets while portfolios controlling for sparsity and cost are concentrating their weights on 40% of the investment universe for the moderate portfolio sizes ($N \in \{50, 100\}$) and on 20% of the investment universe for larger portfolio sizes ($N \in \{250, 500\}$).

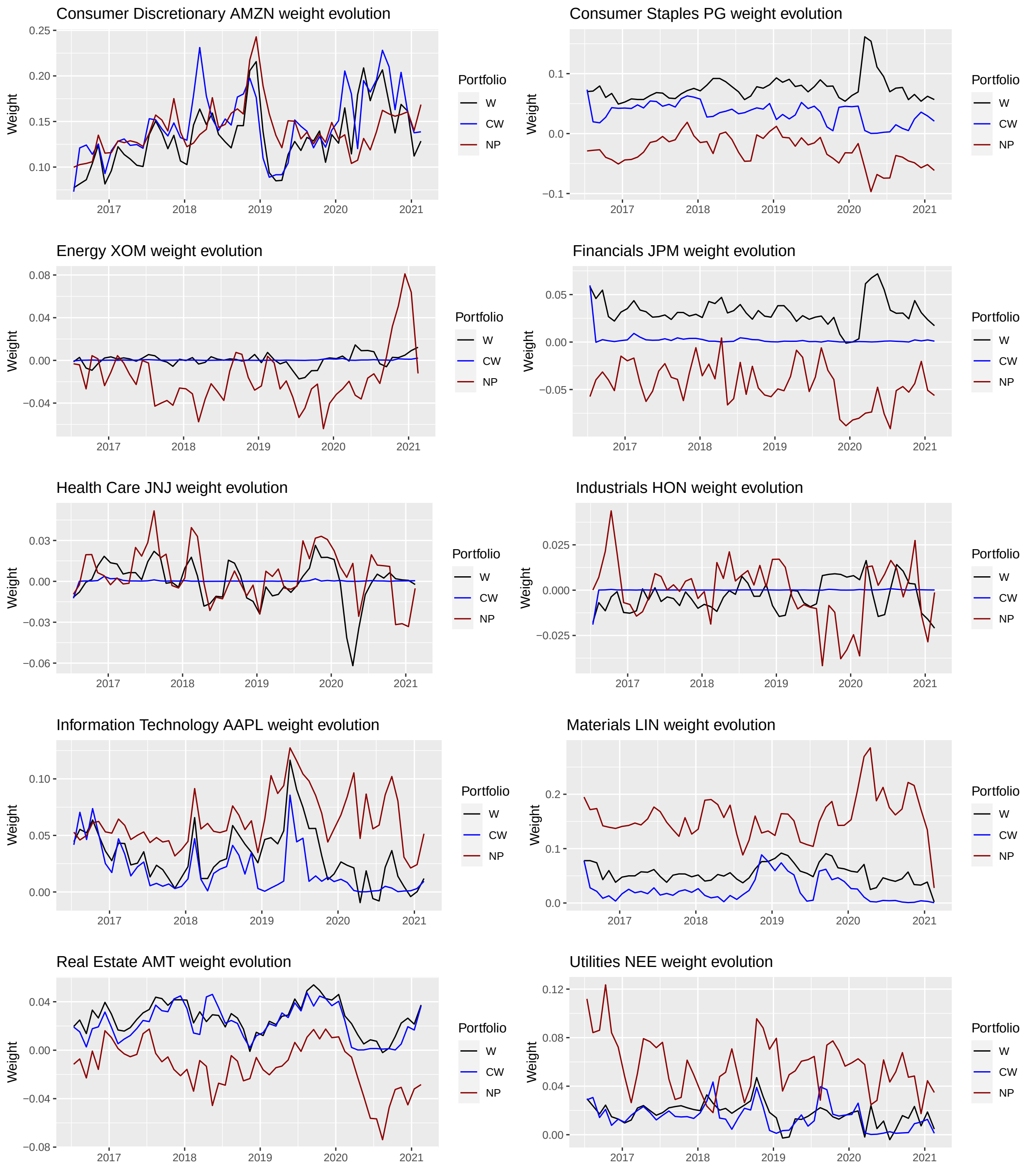


Figure 2: Out-of-sample period weight evolution

We also present the weight evolution during the out-of-sample period as weight variations provide us with further indications on the relative performance of the different portfolios in terms of transaction costs. Figure 2 illustrates the weight variation over time for component stocks in the case where the investment universe is composed of 50 assets. The asset with the highest sector-wise market capitalization is chosen as a representative of each sector, and in such a way we obtain ten stocks representing the ten sectors in the portfolio. From Figure 2, it can be concluded that the DCC-GARCH portfolio (NP) is subject to large changes from one re-balancing date to another, which further deteriorates the portfolio return when transaction costs are taken into account. The black curve indicates that adding the within-sector penalization structure is stabilizing the weights across time, resulting in a further improved relative portfolio performance – compared to the benchmark NP portfolio – when accounting for transaction costs. Additionally, it can be observed from the blue curve that including another regularization term based on transaction costs is further stabilizing the weight evolution over time.

Table 3: Statistics of portfolio weights

	W	B	S	Gl	C	BW	SW	SGl	CW	NP	$1/N$
						N=50					
Min	-0.070	-0.084	0.000	-0.078	-0.001	-0.073	0.000	0.002	-0.001	-0.074	0.020
Max	0.212	0.212	0.332	0.232	0.329	0.191	0.318	0.277	0.310	0.235	0.020
SD	0.049	0.065	0.057	0.066	0.057	0.046	0.053	0.042	0.052	0.066	0.000
MAD-EW	0.034	0.050	0.032	0.050	0.032	0.031	0.028	0.020	0.027	0.050	0.000
MDiv-Sec-W	1.060	2.475	1.352	2.407	1.356	1.237	1.085	0.937	1.075	2.317	0.000
MDiv-Sec-B	0.265	0.124	0.193	0.158	0.193	0.203	0.200	0.122	0.201	0.191	0.000
No. Active	50	50	20	50	27	50	27	50	32	50	50
						N=100					
Min	-0.046	-0.062	0.000	-0.062	-0.001	-0.045	0.000	0.000	-0.001	-0.061	0.010
Max	0.200	0.163	0.306	0.170	0.305	0.188	0.333	0.307	0.329	0.174	0.010
SD	0.033	0.041	0.038	0.041	0.039	0.033	0.039	0.038	0.039	0.041	0.000
MAD-EW	0.020	0.030	0.017	0.030	0.017	0.020	0.016	0.017	0.016	0.030	0.000
MDiv-Sec-W	1.752	2.943	1.566	2.915	1.610	1.808	1.347	1.584	1.356	2.887	0.000
MDiv-Sec-B	0.124	0.063	0.087	0.072	0.087	0.113	0.103	0.087	0.104	0.082	0.000
No. Active	100	100	40	100	25	99	55	29	46	100	100
						N=250					
Min	-0.030	-0.028	0.000	-0.028	-0.001	-0.029	0.000	0.000	-0.001	-0.028	0.004
Max	0.126	0.127	0.305	0.132	0.302	0.103	0.291	0.305	0.290	0.140	0.004
SD	0.018	0.019	0.024	0.019	0.024	0.017	0.024	0.024	0.024	0.019	0.000
MAD-EW	0.011	0.013	0.007	0.013	0.007	0.012	0.007	0.007	0.007	0.013	0.000
MDiv-Sec-W	2.831	3.229	1.779	3.230	1.808	2.914	1.779	1.779	1.795	3.208	0.000
MDiv-Sec-B	0.023	0.015	0.038	0.015	0.038	0.013	0.038	0.038	0.038	0.019	0.000
No. Active	249	249	57	249	60	249	57	70	59	249	250
						N=500					
Min	-0.014	-0.016	0.000	-0.016	-0.000	-0.014	0.000	0.000	-0.000	-0.016	0.002
Max	0.067	0.072	0.238	0.072	0.235	0.058	0.232	0.238	0.229	0.078	0.002
SD	0.009	0.009	0.015	0.009	0.015	0.009	0.015	0.015	0.015	0.010	0.000
MAD-EW	0.006	0.006	0.004	0.006	0.004	0.006	0.004	0.004	0.004	0.006	0.000
MDiv-Sec-W	2.866	3.137	1.845	3.157	1.870	2.936	1.847	1.851	1.862	3.129	0.000
MDiv-Sec-B	0.009	0.007	0.017	0.007	0.017	0.006	0.017	0.017	0.017	0.009	0.000
No. Active	499	499	111	498	111	498	73	99	104	498	500

3.4. GMV portfolio performance with short-sale constraints

Given that some markets and fund managers are explicitly subject to short-sale restrictions in their activities, we also consider the added benefits of our approach when a non-negativity constraint is applied to all portfolio weights. We report the values of the optimal penalization parameters λ for all cases in Table 7 of the appendix.

Comparing the GMV portfolio results in Table 2 with the no-short-sale portfolio results in Table 4, we observe that the optimal no-short-sale portfolios have smaller out-of-sample volatility than the GMV portfolios for all considered investment universes. This finding suggests that the imposed no-short-sale constraint helps decreasing the actual portfolio risk, but further improvements can still be obtained. In addition, our previous conclusion that including within-sector regularization is strongly beneficial still holds, especially in small and moderate investment universes. Whereas adding between-sector shrinkage effectively improves performance for small to moderate dimensions, it tends to deteriorate overall portfolio performance in large dimensions ($N = 500$), although the differences are small.

Furthermore, although within-sector penalization reduces portfolio volatility, this improvement in performance is only statistically significant in the case of $N = 50$. The outperformance in terms of volatility is not statistically significant for larger portfolios ($N \in \{100, 250, 500\}$).

Finally, we take a closer look at the portfolio construction by reporting the weights statistics in Table 5. Clearly, there is now a stronger effect of the penalization term, which is shrinking the weights towards equality, especially in larger portfolios. Moreover, the W portfolio has higher within-sector diversification while the between-sector diversification increases for larger portfolios ($N \in \{250, 500\}$) and decreases for moderate portfolio sizes ($N \in \{50, 100\}$). Additionally, the number of active positions decreases significantly compared to the corresponding portfolios without short-sale constraint, and this effect gets stronger for larger investment universes.

Table 4: Out-of-sample annualized portfolio standard deviation (in percentage points) and Sharpe ratio (with short-sale constraint)

	W	B	C	BW	CW	NP	$1/N$
				N=50			
SD	16.370**	16.401**	16.910**	16.472**	**16.356****	16.931	19.868
SR	0.866	1.109	0.959	1.014	0.873	0.953	0.729
				N=100			
SD	**16.589**	16.681*	16.793	16.669	16.602	16.798	20.269
SR	0.781	1.093	1.050	0.809	0.781	1.046	0.648
				N=250			
SD	16.381	17.565	17.582	**16.378**	16.403	17.554	20.539
SR	0.598	0.915	0.905	0.596	0.601	0.914	0.573
				N=500			
SD	17.515	17.648	17.610*	17.605	**17.501**	17.625	21.511
SR	0.828	0.949	0.940	0.838	0.811	0.946	0.518

Note: Significant outperformance of the portfolios over the NP portfolio in terms of SD is denoted by asterisks: $*$, $**$, and $***$ indicate significance at the 10%, 5%, and 1% level respectively. The outperformances of all portfolios against the equally-weighted benchmark are significant at a level of 5% or lower.

Table 5: Statistics of portfolio weights (with short-sale constraint)

	W	B	C	BW	CW	NP	1/N
N=50							
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.020
Max	0.206	0.299	0.330	0.206	0.207	0.331	0.020
SD	0.040	0.054	0.057	0.038	0.040	0.057	0.000
MAD-EW	0.025	0.030	0.031	0.024	0.025	0.031	0.000
MDiv-Sec-W	0.496	1.343	1.340	0.586	0.494	1.345	0.000
MDiv-Sec-B	0.214	0.184	0.192	0.204	0.213	0.193	0.000
No. Active	33	28	25	39	36	23	50
N=100							
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.010
Max	0.290	0.285	0.306	0.278	0.285	0.306	0.010
SD	0.035	0.037	0.038	0.034	0.034	0.038	0.000
MAD-EW	0.015	0.017	0.017	0.014	0.014	0.017	0.000
MDiv-Sec-W	1.136	1.572	1.569	1.162	1.110	1.565	0.000
MDiv-Sec-B	0.106	0.086	0.087	0.104	0.106	0.087	0.000
No. Active	68	51	43	75	72	51	100
N=250							
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.004
Max	0.050	0.303	0.304	0.049	0.053	0.304	0.004
SD	0.009	0.024	0.024	0.009	0.010	0.024	0.000
MAD-EW	0.005	0.007	0.007	0.005	0.005	0.007	0.000
MDiv-Sec-W	0.799	1.762	1.759	0.767	0.872	1.762	0.000
MDiv-Sec-B	0.042	0.038	0.038	0.042	0.042	0.038	0.000
No. Active	141	111	113	147	144	117	250
N=500							
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.002
Max	0.223	0.190	0.237	0.183	0.222	0.238	0.002
SD	0.015	0.013	0.015	0.013	0.015	0.015	0.000
MAD-EW	0.004	0.004	0.004	0.004	0.004	0.004	0.000
MDiv-Sec-W	1.829	1.858	1.829	1.852	1.819	1.834	0.000
MDiv-Sec-B	0.018	0.011	0.017	0.012	0.018	0.017	0.000
No. Active	125	168	188	117	153	187	500

4. Conclusions and outlook

In a dynamic framework, considering various regularizations for portfolio weights, we have shown that both sparsity and sector-wise regularizations are important for reducing out-of-sample portfolio volatilities. Among the sector-wise regularizations, we observe that controlling for within-sector weight variations has the largest contribution to reducing out-of-sample portfolio volatility compared to promoting between-sector diversification or concentrating the portfolio weights on a few sectors. In a scenario without short-sale constraints, the best-performing portfolio always includes a control for sparsity that implicitly reduces negative weights and moves the portfolio closer to one with short sales constraints. In large dimensions, the optimal portfolio combines the sparsity regularization with a penalty for within-sector variation of weights. The optimal portfolios significantly outperform two benchmark portfolios. Adding short-sale constraints, the best-performing portfolio always includes a "within" penalty, but the statistical significance disappears when portfolio size is larger.

Our results can be extended in various directions. First, other portfolios than the GMV can be considered such as the tangency portfolio, or factor-risk-parity portfolios as in Lassance et al. (2021). Second, our methodology could be applied to international diversification, where the structure is not only based on industry sectors but also countries. In a similar vein as our paper, it would be interesting to investigate whether, for example, within-country diversification is more relevant than between-country diversification. Moreover, additional classes of assets can be considered such as bonds, foreign exchange, commodities and alternative assets, for which other types of regularizations might be useful, depending on the objectives. We leave these topics for future research.

Appendix

Table 6: Logarithm of penalty tuning parameters, log(λ), via cross validation (without short-sale constraint)

	W	B	S	Gl	C	BW		SW		SGl		CW	
N=50	-15.8	-16.6	-9.3	4.9	-8.1	-16.6	-15.8	-9.4	-15.8	-9.3	4.9	-8.2	-15.8
N=100	-15.5	-18.1	-11.4	-11.5	-13.8	-18.2	-15.6	-11.6	-15.6	-11.4	-11.4	-14.0	-15.6
N=250	-17.6	-17.6	-10.9	-5.8	-10.8	-17.6	-17.5	-11.0	-17.5	-11.1	-5.7	-10.9	-17.6
N=500	-17.8	-18.4	-11.8	-4.6	-12.7	-18.5	-17.8	-11.8	-17.9	-11.8	-4.8	-2.0	-17.8

Table 7: Logarithm of penalty tuning parameters, log(λ), via cross validation (with short-sale constraint)

	W	B	C	BW		CW	
N=50	-15.0	-16.6	-8.3	-16.7	-15.1	-8.4	-15.0
N=100	-14.9	-18.0	-12.4	-17.9	-15.0	-12.2	-14.9
N=250	-13.6	-20.9	-11.2	-21.0	-13.6	-11.0	-13.7
N=500	-17.0	-17.3	-11.6	-17.4	-17.0	-11.4	-16.9

Table 8: **P-values for statistical test of difference in volatility (without short-sale constraint)**

	W	B	S	Gl	C	BW	SW	SGl	CW	NP	1/N
						N=50					
W		0.212		0.261						0.038	0.001
B				0.842						0.314	0.010
S	0.130	0.062		0.088						0.005	0.000
Gl										0.556	0.029
C	0.100	0.049	0.462	0.076			0.788			0.003	0.000
BW	0.014	0.000	0.546	0.002	0.570		0.446		0.508	0.000	0.000
SW	0.082	0.052	0.970	0.078						0.002	0.000
SGl	0.013	0.009	0.056	0.021	0.044	0.638	0.072		0.095	0.001	0.000
CW	0.054	0.042	0.576	0.078	0.765		0.011			0.002	0.000
NP											0.016
						N=100					
W		0.000		0.000						0.000	0.000
B				0.806							0.002
S	0.306	0.000		0.000	0.205	0.492	0.466	0.226	0.642	0.000	0.000
Gl											0.005
C	0.334	0.000		0.000		0.539	0.658		0.834	0.000	0.000
BW	0.087	0.000		0.000						0.000	0.000
SW	0.387	0.000		0.000		0.619				0.000	0.000
SGl	0.312	0.000		0.000	0.205	0.501	0.496		0.672	0.000	0.000
CW	0.352	0.000		0.000		0.564	0.602			0.000	0.000
NP		0.333		0.059							0.004

Note: The p-values are from the tests of outperformance where the row portfolios are benchmarked against the column portfolios, i.e. we test whether portfolio i significantly outperforms portfolio j in terms of volatility for row i and column j of the table. An empty cell indicates portfolio i is not outperforming portfolio j.

Table 9: **P-values for statistical test of difference in volatility (without short-sale constraint)**

	W	B	S	Gl	C	BW	SW	SGl	CW	NP	1/N
						N=250					
W		0.000		0.000		0.000				0.001	0.000
B				0.390							0.017
S	0.174	0.006		0.004	0.762	0.058	0.889	0.454	0.929	0.021	0.001
Gl											0.035
C	0.181	0.007		0.004		0.061	0.922		0.956	0.023	0.001
BW		0.002		0.004						0.158	0.000
SW	0.139	0.004		0.002		0.040				0.014	0.000
SGl	0.176	0.006		0.004	0.960	0.059	0.908		0.952	0.022	0.001
CW	0.143	0.004		0.002		0.042	0.911			0.014	0.000
NP		0.000		0.004							0.001
						N=500					
W		0.002		0.001		0.000				0.011	0.000
B											0.000
S	0.044	0.000		0.001		0.003				0.001	0.000
Gl		0.189									0.000
C	0.043	0.000	0.799	0.000		0.003		0.825		0.001	0.000
BW		0.502		0.560						0.999	0.000
SW	0.025	0.000	0.266	0.000	0.278	0.001		0.270	0.262	0.000	0.000
SGl	0.044	0.000	0.679	0.001		0.003				0.001	0.000
CW	0.026	0.000	0.416	0.000	0.434	0.001		0.422		0.000	0.000
NP		0.000		0.000							0.000

Note: The p-values are from the tests of outperformance where the row portfolios are benchmarked against the column portfolios, i.e. we test whether portfolio i significantly outperforms portfolio j in terms of volatility for row i and column j of the table. An empty cell indicates portfolio i is not outperforming portfolio j.

Table 10 reports the total running time for the out-of-sample period of 56 months using the optimal penalization parameter found in the cross-validation step. The running time increases moderately with respect to portfolio size. Note that the optimal weights of the portfolios with W, B, S, Gl penalty terms at time t are independent of the optimal weights of other re-balancing dates and thus the computation can be implemented with parallel programming. However, the optimal allocation at time t of the portfolios controlling for the cost penalty is dependent on the optimal weights of the previous re-balancing date $t-1$, and thus can only be computed sequentially.

Table 10: Running time for the out-of-sample period (in minutes)

	W	B	S	Gl	C	BW	SW	SGl	CW
N=50	0.33	0.05	0.03	0.03	0.13	0.37	0.33	0.04	2.05
N=100	0.32	0.06	0.03	0.03	0.13	0.36	0.33	0.05	2.09
N=250	0.32	0.07	0.05	0.05	0.22	0.38	0.37	0.07	2.17
N=500	0.53	0.19	0.16	0.13	0.65	0.56	0.53	0.22	2.67

Note: Computations were carried out on a 2020 iMac with the following specifications: Intel Core i9 10-Core processor (3.6GHz) and 72 GB RAM (2133 MHz DDR4).

Acknowledgements

We would like to thank Eric Ghysels, Nathan Lassance, Francesco Roccazzella and Jonas Striaukas for helpful comments about a first draft of this paper. Both authors gratefully acknowledge financial support of the research grant ARC 18-23/089 of the Belgian Federal Science Policy Office.

References

Aielli, G. P. (2013): "Dynamic conditional correlation: on properties and estimation". *Journal of Business & Economic Statistics* 31.3, pp. 282–299.

Ao, M., L. Yingying, and X. Zheng (2019): "Approaching mean-variance efficiency for large portfolios". *The Review of Financial Studies* 32.7, pp. 2890–2919.

Babii, A., E. Ghysels, and J. Striaukas (2020): "High-dimensional Granger causality tests with an application to VIX and news". *Available at SSRN 3615718*.

— (2021): "Machine learning time series regressions with an application to nowcasting". *Journal of Business & Economic Statistics*, pp. 1–23.

Baillie, R. T. and T. Bollerslev (1992): "Prediction in dynamic models with time-dependent conditional variances". *Journal of Econometrics* 52.1-2, pp. 91–113.

Best, M. J. and R. R. Grauer (1991): "On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results". *The Review of Financial Studies* 4.2, pp. 315–342.

Brito, D., M. C. Medeiros, and R. Ribeiro (2018): "Forecasting large realized covariance matrices: The benefits of factor models and shrinkage". *Available at SSRN 3163668*.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris (2009): "Sparse and stable Markowitz portfolios". *Proceedings of the National Academy of Sciences* 106.30, pp. 12267–12272.

Candelon, B., C. Hurlin, and S. Tokpavi (2012): "Sampling error and double shrinkage estimation of minimum variance portfolios". *Journal of Empirical Finance* 19.4, pp. 511–527.

Chan, L. K., J. Karceski, and J. Lakonishok (1999): "On portfolio optimization: Forecasting covariances and choosing the risk model". *The Review of Financial Studies* 12.5, pp. 937–974.

Chen, J., G. Dai, and N. Zhang (2020): "An application of sparse-group lasso regularization to equity portfolio optimization and sector selection". *Annals of Operations Research*, pp. 1–20.

De Nard, G., O. Ledoit, and M. Wolf (2021): "Factor models for portfolio selection in large dimensions: The good, the better and the ugly". *Journal of Financial Econometrics* 19.2, pp. 236–257.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009a): "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms". *Management Science* 55.5, pp. 798–812.

DeMiguel, V., L. Garlappi, and R. Uppal (2009b): "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" *The Review of Financial Studies* 22.5, pp. 1915–1953.

Engle, R. (2002): "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models". *Journal of Business & Economic Statistics* 20.3, pp. 339–350.

Engle, R. F., O. Ledoit, and M. Wolf (2019): "Large dynamic covariance matrices". *Journal of Business & Economic Statistics* 37.2, pp. 363–375.

Engle, R. F. and K. Sheppard (2001): "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH". *National Bureau of Economic Research, Working Paper* 8554.

Fama, E. F. and K. R. French (1993): "Common risk factors in the returns on stocks and bonds". *Journal of Financial Economics* 33, pp. 3–56.

Fan, J., A. Furger, and D. Xiu (2016): "Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data". *Journal of Business & Economic Statistics* 34.4, pp. 489–503.

Fan, J., J. Zhang, and K. Yu (2012): "Vast portfolio selection with gross-exposure constraints". *Journal of the American Statistical Association* 107.498, pp. 592–606.

Fastrich, B., S. Paterlini, and P. Winker (2015): "Constructing optimal sparse portfolios using regularization methods". *Computational Management Science* 12.3, pp. 417–434.

Fisher, J. D., D. W. Puelz, and C. M. Carvalho (2020): "Monotonic effects of characteristics on returns". *Annals of Applied Statistics* 14.4, pp. 1622–1650.

Friedman, J., T. Hastie, and R. Tibshirani (2010): "A note on the group lasso and a sparse group lasso". *arXiv preprint arXiv:1001.0736.*

Fu, A., B. Narasimhan, and S. Boyd (2020): "CVXR: An R Package for Disciplined Convex Optimization". *Journal of Statistical Software* 94.1, pp. 1–34.

Hafner, C. M. and O. Reznikova (2012): "On the estimation of dynamic conditional correlation models". *Computational Statistics & Data Analysis* 56.11, pp. 3533–3545.

Hautsch, N. and S. Voigt (2019): "Large-scale portfolio allocation under transaction costs and model uncertainty". *Journal of Econometrics* 212.1, pp. 221–240.

Hong, H., W. Torous, and R. Valkanov (2007): "Do industries lead stock markets?" *Journal of Financial Economics* 83.2, pp. 367–396.

Hou, K. (2007): "Industry information diffusion and the lead-lag effect in stock returns". *The Review of Financial Studies* 20.4, pp. 1113–1138.

Jagannathan, R. and T. Ma (2003): "Risk reduction in large portfolios: Why imposing the wrong constraints helps". *The Journal of Finance* 58.4, pp. 1651–1683.

Kurose, Y. and Y. Omori (2020): "Multiple-block dynamic equicorrelations with realized measures, leverage and endogeneity". *Econometrics and Statistics* 13, pp. 46–68.

Lassance, N., V. DeMiguel, and F. Vrins (2021): "Optimal portfolio diversification via independent component analysis". *Operations Research*.

Ledoit, O. and M. Wolf (2004): "A well-conditioned estimator for large-dimensional covariance matrices". *Journal of Multivariate Analysis* 88.2, pp. 365–411.

— (2011): "Robust performances hypothesis testing with the variance". *Wilmott* 2011.55, pp. 86–89.

— (2017): "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks". *The Review of Financial Studies* 30.12, pp. 4349–4388.

— (2020): "Analytical nonlinear shrinkage of large-dimensional covariance matrices". *The Annals of Statistics* 48.5, pp. 3043–3065.

Markowitz, H. (1952): "Portfolio Selection". *The Journal of Finance* 7.1, pp. 77–91.

Morana, C. (2019): "Regularized semiparametric estimation of high dimensional dynamic conditional covariance matrices". *Econometrics and Statistics* 12, pp. 42–65.

Moskowitz, T. J. and M. Grinblatt (1999): "Do industries explain momentum?" *The Journal of Finance* 54.4, pp. 1249–1290.

Pakel, C., N. Shephard, K. Sheppard, and R. F. Engle (2020): "Fitting vast dimensional time-varying covariance models". *Journal of Business & Economic Statistics*, pp. 1–17.

Puelz, D., C. M. Carvalho, and P. R. Hahn (2015): "Optimal ETF selection for passive investing". *arXiv preprint arXiv:1510.03385*.

Puelz, D., P. R. Hahn, and C. M. Carvalho (2020): "Portfolio selection for individual passive investing". *Applied Stochastic Models in Business and Industry* 36.1, pp. 124–142.

Simon, N., J. Friedman, T. Hastie, and R. Tibshirani (2013): "A sparse-group lasso". *Journal of Computational and Graphical Statistics* 22.2, pp. 231–245.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso". *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1, pp. 267–288.

Yuan, M. and Y. Lin (2006): "Model selection and estimation in regression with grouped variables". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.1, pp. 49–67.

Zhao, Z., O. Ledoit, and H. Jiang (2020): "Risk reduction and efficiency increase in large portfolios: leverage and shrinkage". *University of Zurich, Department of Economics, Working Paper* 328.

Zou, H. and T. Hastie (2005): "Regularization and variable selection via the elastic net". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2, pp. 301–320.