Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis Quantitative Finance

Master's Thesis

Jorrit Wijckmans (643993)

Supervisor: dr. A Quaini

Abstract

This research evaluates portfolio performance using weights derived from the direct estimation of the precision matrix via Reduced-Rank regression. Our empirical study compares this method's effectiveness against a Nodewise regression method and the equally weighted (1/N) portfolio. The findings reveal that for historical monthly S&P500 data, Reduced-Rank regression is outperformed by both alternative methods, with Nodewise regression achieving the highest Sharpe Ratio, even when transaction costs are considered. In the analysis of historical daily S&P500 data, Reduced-Rank Regression shows commendable performance until transaction costs are included, at which point the 1/N portfolio demonstrates superior Sharpe Ratio performance in out-of-sample testing. Additionally, the potential for improving Reduced-Rank Regression's performance by easing some of the study's constraints is noted, suggesting a possible avenue for future research.

Contents

1 Introduction

Markowitz (1952) seminal work on optimal portfolio selection revolutionized the field of investment management and financial theory. This work laid the foundation for Modern Portfolio Theory (MPT), which introduces the concept of optimizing the balance between risk and return in investment portfolios. The theory posits that investors can achieve this balance by diversifying their investment portfolios.

The methodology for portfolio selection as introduced by Markowitz (1952) necessitates the estimation of both the expected return vector and the covariance matrix of returns. Green et al. (2013) note the extensive literature on the estimation of expected returns, citing more than 300 papers focused on this issue. In contrast, the subject of estimating the covariance matrix has not been as extensively explored. What is widely acknowledged in the field, however, is that the standard approach of using the sample covariance matrix is not suitable for this purpose.

Accurately estimating the covariance matrix is a fundamental aspect of portfolio construction in financial asset management. This estimation is particularly vital because the inverse covariance matrix, in this thesis referred to as the precision matrix, plays a critical role in determining the optimal weights for both the mean-variance portfolio and the Global Minimum Variance portfolio (GMV). However, estimating the covariance matrix becomes increasingly complex in high-dimensional settings where the number of assets, denoted as N, is large compared to the length T of the time series. In such high-dimensional scenarios, traditional estimation methods encounter significant challenges. The primary issue is that the sample covariance matrix can become singular or nearly singular, meaning it cannot be inverted, or its inversion becomes highly unstable. This singularity leads to unreliable estimates of the precision matrix, which in turn, can severely distort the portfolio optimization process, resulting in inaccurate asset allocations.

To address the challenges of high-dimensionality in covariance matrix estimation, several approaches have been developed. Factor models, whose performances are extensively studied by Fan et al. (2011, 2013), are based on the assumption that excess returns on assets are driven by a limited set of either latent or observable factors. Within this framework, models focus on estimating the factor-based covariance matrix of these excess returns. The advantage of using a factor-based structure is that it simplifies the covariance matrix, making it lower-dimensional and, consequently, more manageable to invert. In their comprehensive retrospective, Ledoit and Wolf (2022b) revisit over 15 years of research on diverse shrinkage methods, which can be categorized into two types: linear and nonlinear. Linear shrinkage stands out for its simplicity in terms of comprehension, derivation, and implementation. However, nonlinear shrinkage offers a further enhancement in performance. This is particularly true when it incorporates specific stylized facts, like time-varying co-volatility or the integration of factor models, which can significantly refine the effectiveness of these techniques.

The majority of approaches in the estimation of covariance matrices initially concentrate on deriving the covariance matrix and then inverting it to acquire the precision matrix. Pioneering this field, Stevens (1998) directly estimated the precision matrix through a novel methodology later referred to as Nodewise regression. This work set the stage for further advancements. Meinshausen and Bühlmann (2006) notably advanced this area of study by employing p Lasso regressions. This technique allowed for the crafting of a sparse precision matrix of dimension $p \times p$, an innovative step in simplifying and enhancing the precision matrix's usability. Building on these works, Callot et al. (2021) made significant contributions by establishing consistency results for the variance and the corresponding weights of the mean-variance portfolio and of the GMV portfolio. Both results are obtained in high dimensional settings.

In this thesis, we adopt Stevens (1998)'s approach to directly estimate the precision matrix. We combine this with a methodology introduced by Negahban and Wainwright (2011), Chen et al. (2013), and Yuan et al. (2007) by integrating Reduced-Rank regression. Reduced-rank regression imposes a rank constraint on the coefficient matrix, effectively reducing the number of parameters to be estimated.

The unweighted nuclear norm, in contrast to the adaptive nuclear norm described by Chen et al. (2013), treats all singular values of a matrix equally. In this formulation, singular values are summed without assigning differential weights, resulting in a convex norm. This promises a sophisticated and accurate method of directly estimating the precision matrix.

The thesis is structured as follows. We begin with an in-depth exploration of the existing literature, focusing specifically on various methodologies employed to address the challenges of high-dimensionality in covariance matrix estimation. Following this, we delve into the core of our methodology. Here, we elaborate on our approach to directly estimate the precision matrix. This is achieved through the implementation of Reduced-Rank regression. Furthermore, we delve into the methodology behind Nodewise regression and the 1/N portfolio. This is followed by an overview of the dataset selected for our studies. In the subsequent section, a simulation study is conducted to evaluate the efficacy of the three aforementioned methods across various data-generating scenarios. Finally, we present and analyze the findings from an empirical study, drawing conclusions on the practical applicability of our research. We conclude with a discussion of these findings and provide final remarks.

2 Literature

Covariance matrix estimation is crucial for portfolio optimization. While the sample covariance matrix is unbiased, it can incur significant estimation errors that may be amplified during portfolio optimization. The reliance on the precision matrix for calculating optimal weights in both mean-variance and minimum-variance portfolios underscores the importance of precise estimation. Consequently, substantial research efforts have been put into refining covariance matrix estimation techniques, with several of these explored in the current analysis.

2.1 Factor Models

Factor models represent one methodology investigated in this context, exemplified in the work by Chan et al. (1999), who employ a handful of factors to encapsulate the overall covariance structure. This approach enhances the robustness of the covariance matrix estimation and improves the Sharpe ratio. Factors might be preselected, such as the three Fama-French factors, or extracted from the dataset using techniques like Principal Component Analysis (PCA). An empirical evaluation by Chan et al. (1999) demonstrates that PCA's first factor explains a significant portion of assets' variability. They also found that adding more factors does not consistently enhance covariance matrix estimation, particularly when the market factor is a significant explanatory component.

The performance of factor models in financial analysis has been extensively studied, with notable contributions from Fan et al. (2011, 2013). These models specialize in estimating factor-based covariance matrices for excess returns, leveraging the advantage of their low-dimensional factor structure, which allows for straightforward matrix inversion. Further research by Fan et al. (2016) and Ait-Sahalia and Xiu (2017) delved into the accuracy of precision matrices estimated using factor models. However, their studies did not extend to the examination of portfolio variances and weights in scenarios where the number of assets exceeds the number of observations.

In a related development, Fan et al. (2008) introduced a covariance matrix estimator that operates under the assumption of observable factors and a diagonal residual covariance matrix, implying conditional sparsity. This work also established the convergence rates of portfolio risk in larger portfolios using this estimator. Building on this, Fan et al. (2015) developed the Principal Orthogonal Complement Thresholding (POET) estimator, tailored for scenarios with unobserved factors. In high-dimensional contexts, Ding (2017) presented insights on optimizing global variance ratios through factor models. Additionally, Ao and Zheng (2017) contributed to the discourse on portfolio selection, applying the Lasso technique in cases of heteroskedastic asset returns.

2.2 Shrinkage Techniques

Shrinkage techniques, as proposed by Ledoit and Wolf (2003), are another strategy for improving covariance estimates. These techniques involve taking a convex combination of the sample covariance matrix and a target matrix, creating a balance that reduces the impact of estimation errors on portfolio allocation. The benefits of such shrunk covariance matrices, as presented by Ledoit and Wolf (2003, 2004a), surpass those of conventional sample covariance matrices and factor models. However, selecting the optimal shrinkage intensity is a challenge, especially when dealing with a large number of assets.

Ledoit and Wolf (2003) introduced the concept of shrinkage to enhance covariance estimation. This method involves creating a convex combination of the sample covariance matrix with a specified shrinkage target matrix. In their subsequent work, Ledoit and Wolf (2004a) elaborated on various types of target matrices such as identity, single-factor, and second-moment matrices. The process of linear shrinkage in covariance matrix estimation uses a predetermined shrinkage intensity to decide the extent of dependence on the target covariance matrix. As detailed by Ledoit and Wolf (2003), the linear combination of sample and target covariance matrices mitigates the influence of estimation errors in portfolio allocation. This is achieved by shrinking the overestimated covariance estimators towards their target matrix counterparts. The findings presented by Ledoit and Wolf (2003, 2004a) demonstrate that covariance matrices modified through shrinkage tend to surpass traditional sample covariance matrices, as well as factor models and PCA-based covariance estimators in performance.

Determining the appropriate shrinkage intensity poses a challenge when implementing linear shrinkage of the covariance matrix. In their studies, Ledoit and Wolf (2003, 2004b,a) propose unique shrinkage intensities tailored to various target matrices. Parameters like the time period, variances, and correlations of the sample and factor are critical in determining the shrinkage factor. However, it is important to note that the shrinkage intensity remains consistent across all assets. This uniformity can lead to a diminished effectiveness of linearly shrunk covariance matrices in high-dimensional scenarios where the number of assets N is significantly large.

Ledoit and Wolf (2012) presented a non-linear shrinkage approach for estimating the covariance matrix. This method utilizes the spectral decomposition of the sample covariance matrix, enabling the eigenvalues to be shrunk at varied intensities. The key advantage of this non-linear shrinkage estimator is its ability to apply distinct shrinkage intensities to individual assets. According to Ledoit and Wolf (2012), this non-linear approach is more effective than linear shrinkage in high-dimensional environments. The non-linear shrinkage method by Ledoit and Wolf (2012), while offering significant improvements in the accuracy of covariance matrix estimation, especially in high-dimensional settings, necessitates complex calculations including eigenvalue decomposition, non-linear adjustments, and possibly iterative optimization, all of which contribute to its computational intensity. To address this challenge, Ledoit and Wolf (2020) introduced an analytical technique for non-linear covariance shrinkage, offering a solution that matches the accuracy of numerical methods but is significantly more computationally efficient.

Ledoit and Wolf (2022a) developed various non-linear methods to shrink the covariance matrix. One such method is the linear-inverse shrinkage estimator, which applies a refined version of Stein shrinkage to the eigenvectors of the sample covariance matrix. Similarly, the quadratic-inverse shrinkage estimator shrinks the eigenvectors using both the inverse Stein method and a minimum-variance loss function. Additionally, Ledoit and Wolf (2022b) introduced a technique that combines linear- and quadratic-inverse shrinkage estimators through a geometric average, enhancing covariance estimation.

Within financial return analysis, the presumption that returns are independent and identically distributed (i.i.d.) is frequently considered unsuitable, given the data's inherent dependencies and heterogeneity. To tackle this issue, Engle (1982) suggested employing auto-regressive moving average (ARMA) models for estimating the variance, paving the way for the creation of multivariate GARCH models. To enhance

this approach, Bauwens et al. (2006) developed the Dynamic Conditional Correlation (DCC) models. These models provide increased flexibility in modelling correlations and consistency in variances. The DCC model, initially presented by Engle (2002), extends the GARCH model to a multivariate format, incorporating either a large static or time-varying covariance matrix, as further elaborated by Ledoit and Wolf (2022b). Its estimation process involves a two-step method: initially applying a GARCH model to univariate data, followed by estimating the covariance matrix based on the selected multivariate distribution.

Estimating the dynamic conditional correlation matrix in high-dimensional environments, akin to challenges faced with sample covariance estimators, can be computationally demanding. To address this, Engle et al. (2019) introduced both linear and non-linear shrinkage estimators within the DCC model framework. These estimators, named DCC-LS (Linear Shrinkage) and DCC-NLS (Non-Linear Shrinkage), commence by fitting a univariate GARCH model, followed by applying either linear or non-linear shrinkage to the unconditional correlation matrix. Engle et al. (2019) demonstrate that the DCC-LS and DCC-NLS estimators outperform other estimators in high-dimensional settings, particularly when the number of assets N is large. The implementation of DCC-shrinkage estimation notably enhances portfolio robustness in cases where the number of assets surpasses the time periods, $N > T$. Consequently, Engle et al. (2019) conclude that dynamic covariance estimation is particularly effective for data with daily or weekly returns.

De Nard et al. (2021) introduced a novel covariance estimator that combines factor structures with the dynamic, time-varying conditional heteroskedasticity of residuals in high-dimensional contexts. This advanced estimator considers time variability across both the factors and the residual matrix, which is derived from regressing the asset returns on these factors. An alternative approach within this model allows for the use of a time-invariant residual matrix over the entire observed period. Their findings indicate that this model surpasses several existing models in performance, including Ledoit and Wolf (2015) non-linear shrinkage estimators and the DCC-shrinkage estimator developed by Engle et al. (2019)

2.3 Direct Estimation of the Precision Matrix

The determination of optimal weights for both mean-variance and minimum-variance optimization is inherently linked to the precision matrix. In high-dimensional scenarios, particularly when the number of assets N surpasses the time periods T, the stability of the sample covariance matrix is compromised, as noted by Stevens (1998). Addressing this, Stevens (1998) proposes characterizing the precision matrix directly. His methodology related the precision matrix to the regression coefficients and residual variances of the regressions of each asset's excess return onto the excess returns of the other assets. The residual variances represent the unhedgeable risk associated with the assets. The outcome is that the diagonal entries of the inverse matrix are equal to the reciprocal of $\sigma_{ii}\left(1 - R^2\right)$. Here, σ_{ii} is a measure of the variability or spread of the ith stock's returns, and the term $(1 - R^2)$ adjusts this variance to account for the proportion of the ith stock's variance that is not explained by other stocks. As for the elements not on the main diagonal of the inverse matrix, it is found that they can be alternatively described as the regression coefficient $\frac{-\beta_{ij}}{\sigma_{ii}(1-R^2)}$ or $\frac{-\beta_{ji}}{\sigma_{ii}(1-R^2)}$, which are equal expressions.

Meinshausen and Bühlmann (2006) extended this research by approaching the estimation of the precision matrix through neighborhood selection, employing Lasso regression to regularize the matrix by adding a penalty term to the absolute values of the coefficients. This is in contrast to other regularization methods such as Ridge or Elastic net regularization. Ridge regularization penalizes the square of the coefficients, leading to a shrinkage of coefficients towards zero. Lasso regression, on the other hand, can completely nullify coefficients, thus facilitating variable selection.

Callot et al. (2021) referred to the method introduced by Stevens (1998) as Nodewise regression. Callot et al. (2021) conducted an empirical analysis comparing the effectiveness of Nodewise regressions with

factor models and shrinkage methods. Their findings suggest that direct covariance estimation via Nodewise regression is competitive, if not superior, to these alternative methods.

While Callot et al. (2021) assume sparsity in the hedging coefficients, our contribution to literature is that we extend upon Stevens (1998) concept of deriving the precision matrix directly through a Reduced-Rank regression, incorporating a penalty on the nuclear norm, as detailed in the works of Negahban and Wainwright (2011), Chen et al. (2013), and Yuan et al. (2007). In this method, reducing dimensions is achieved by limiting the rank of the coefficient matrix. Earlier studies focused on the traditional small-p case and maximum likelihood estimation under rank constraints. Bunea et al. (2011) introduced a criterion for choosing the rank that is applicable in high-dimensional scenarios. This shows that rank-constrained estimation can be seen as a penalized regression technique, where the penalty is related to the rank of matrix Γ. The penalty can also be interpreted as an l_0 penalty, based on the count of non-zero singular values of Γ, expressed as $P_\lambda(\Gamma) = \lambda r(\Gamma) = \lambda \sum_{i=1}^{p \wedge q} I\{d_i(\Gamma) \neq 0\}$, with $I(\cdot)$ as the indicator function and $d_i(\cdot)$ representing the ith largest singular value of a matrix. This leads to an estimator formed through hard thresholding in singular value decomposition. Yuan et al. (2007) proposed a least squares criterion penalized by the nuclear norm, defined as $P_\lambda(\Gamma) = \lambda ||\Gamma||* = \lambda \sum_{i=1}^{p \wedge q} d_i(\Gamma)$, where $||\cdot||_*$ is the nuclear norm. This l_1 penalty promotes sparsity in the singular values and facilitates simultaneous rank reduction and shrinkage estimation (Bunea et al., 2011; Negahban and Wainwright, 2011). Reduced-Rank regression methods relate to various popular techniques like principal component analysis and canonical correlation analysis and are widely explored in matrix completion challenges (Candes and Recht, 2012).

3 Methodology

This section introduces the penalized Reduced-Rank regression, a concept developed by Negahban and Wainwright (2011), Chen et al. (2013), and Yuan et al. (2007). We present our approach for deriving the precision matrix using penalized Reduced-Rank regression. Next, we will introduce Nodewise regression as described in Callot et al. (2021), along with the 1/N portfolio strategy. Subsequently, we will delineate three distinct methods for portfolio allocation. Before proceeding, we introduce the following notation. Let $R_{1t} = [R_{1t}, \cdots, R_{Nt}]'$ denote N asset excess returns observed at time $t = 1, \cdots, T$ with $\mathbb{E}[R_t] = 0$ without loss of generality, and $\mathbb{V}\text{ar}[R_t] = \Sigma \succ 0$.

3.1 Reduced Rank Regression

3.1.1 Derivation of the Precision Matrix Formula

The ability to forecast future stock returns and variances is integral to effective portfolio management. The goal of this paper is to find optimal portfolios by calculating the precision matrix directly and thereby minimizing the portfolio variance. To do so, we start with the following regression,

$$R = R\Gamma + E, \tag{3.1}$$

with R is a $T \times N$ data matrix of excess returns, E denotes error terms and Γ is a $N \times N$ coefficient matrix,

$$\Gamma := \begin{bmatrix} 0 & \gamma_{1,2} & \cdots & \gamma_{1,N} \\ \gamma_{2,1} & 0 & \cdots & \gamma_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N,1} & \gamma_{N,2} & \cdots & 0 \end{bmatrix}. \tag{3.2}$$

Stevens (1998) proves that the precision matrix can then be calculated directly by,

$$\Sigma^{-1} = \Theta = \Psi(I - \Gamma), \tag{3.3}$$

with I the identity matrix and $\Psi = diag(\mathbb{V}ar(E_i))^{-1}$. We can interpret the elements $\gamma_{i,j}/\psi_{i,i}$ as the optimal hedging position in asset j hedging risk in asset i within the $N-1$ assets.

3.1.2 Estimation of the Precision Matrix

This subsection borrows from Yuan et al. (2007), Chen et al. (2013), and Negahban and Wainwright (2011) to approach the estimation of Θ by means of a penalized Reduced Rank Regression

$$\hat{\Gamma} := \underset{\Gamma \in \mathbb{R}^{N \times N}}{\operatorname{argmin}} \left(\frac{1}{2} ||R - R\Gamma||_F^2 + \lambda ||\Gamma||_* : \operatorname{diag}(\Gamma) = 0, \lambda \geq 0 \right), \tag{3.4}$$

where $||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$, is the frobenius norm, and $||A||_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i$, is the nuclear norm. The Frobenius norm provides a way to measure the "size" of a matrix in a manner analogous to how the Euclidean norm measures the length of a vector in Euclidean space. To prevent overfitting, we include the nuclear norm, also known as the trace norm, which is the sum of the singular values of the matrix Γ. Adding the nuclear norm as a penalty encourages a low-rank solution, effectively reducing the complexity of the model. Furthermore, we impose a zero constraint for the main diagonal elements of Γ. By setting the diagonal elements to zero, we explicitly remove the effect of variables on themselves.

In contrast to Chen et al. (2013), we adopt an unweighted penalization on the nuclear norm, assigning equal weights to all assets. This approach is chosen for its simplicity, eliminating the need for complex weight adjustments, and for its computational efficiency. Moreover, the use of an unweighted nuclear norm preserves convexity. In comparison, a weighted nuclear norm is in general not convex, potentially complicating the optimization process. The unweighted norm is also more robust to overfitting.

The penalty parameter λ controls the trade-off between the fit of the model and the complexity. λ multiplies the nuclear norm term, allowing for different penalties on rank and overall size of coefficients.

3.1.3 Generalized Cross Validation

To evaluate predictive performance, we advocate for the use of the Generalized Cross-Validation (GCV) criterion. This criterion offers a method for choosing model parameters that strike an optimal balance between the model's fit and its complexity, eliminating the necessity for dividing the data into distinct training and validation sets. The GCV Criterion is defined as:

$$GCV(\lambda) = \frac{||R - R\Gamma||_F^2}{(1 - \frac{d(\Gamma)}{T})^2}, \tag{3.5}$$

here, the numerator represents the Frobenius norm from the objective function 3.4. A lower value indicates a better fit of the model to the data. The denominator consists of $d(\Gamma)$, the degrees of freedom, which is approximated by the sum of singular values of the matrix Γ, and T is the number of observations. The denominator adjusts the residual by a penalty for model complexity. As the model becomes more complex (using more degrees of freedom), the penalty increases, inflating the GCV score. This discourages overly complex models that might fit the noise in the data rather than underlying trends.

In assessing the GCV score, we systematically explore a sequence of 15[1] lambda (λ) values, distributed logarithmically, spanning from 10^{-4} to 10^{-1}. This range is chosen to encompass a broad spectrum of model complexities, ensuring a thorough investigation into the trade-off between model fidelity and regularization. The λ value that yields the lowest GCV score is selected as optimal for our analysis. Consequently, the Γ matrix associated with this optimal λ is adopted, guiding the structure and parameters of our model.

3.1.4 Projected Subgradient Method

To optimize the objective function $f(\Gamma) = \frac{1}{2}||R - R\Gamma||_F^2 + \lambda||\Gamma||_*$ from (3.4), we implement the projected subgradient method described by Boyd et al. (2003). This method is a versatile and effective optimization tool for complex problems where the objective function includes non-differentiable terms, and/or the problem involves constraints that can be easily projected. The methodology unfolds through a series of steps, within which G_k represents an element of the subgradient of the objective function. The left singular vectors are denoted by U_k, and the right singular vectors by V_k. The projection onto the space of matrices characterized by zero diagonal elements is executed by P_H, which is accomplished by directly setting the diagonal elements to zero.

Algorithm 1: Subgradient Method for Hollow Matrix Projection

1: Initialize $k \leftarrow 0$

2: **while** not converged **do**

3: Compute the subgradient G_k at Γ_k, an element of $\delta f(\Gamma_k)$:

$$G_k = \lambda U_k V_k' - R'(R - R\Gamma_k),$$

where $\Gamma_k = U_k \Sigma_k V_k'$.

4: Update the solution with a step in the direction of the subgradient:

$$\Gamma_{\text{temp}} = \Gamma_k - t_k G_k.$$

5: Project the updated solution back onto the space of hollow matrices H:

$$\Gamma_{k+1} = P_H(\Gamma_{\text{temp}}).$$

6: $k \leftarrow k + 1$

7: **end while**

8: $\Gamma^* \leftarrow \Gamma_{k+1}$

In selecting the step size t_k, various strategies can be employed. For the purposes of this thesis, we adopt a summable but diminishing step size approach, wherein t_k systematically decreases as iterations progress.

$$t_k = \frac{t_0}{\sqrt{k+1}}, \tag{3.6}$$

with

$$t_0 = \frac{2}{\sigma_1^2 + \sigma_N^2}, \tag{3.7}$$

here, σ_1 and σ_N denote the maximum and minimum of the singular values $\sigma_1 \geq \cdots \geq \sigma_N$ of the return matrix R. In implementing the projected subgradient method, we have selected 100 iterations. While this figure may seem modest, and there is a case to be made for increasing the number to 500 or even 1000 to potentially improve results, our choice is constrained by the consideration of computational efficiency.

[1] In our study, we selected a sequence of 15 lambda values to optimize computational efficiency. However, it is worth noting that employing an expanded sequence, such as 100 lambda values, has the potential to refine the identification of the optimal lambda and enhance the corresponding Generalized Cross-Validation (GCV) score.

3.2 Nodewise Regression

This section introduces the Nodewise Regression approach of Callot et al. (2021). Nodewise regression aims to estimate the precision matrix by regressing each variable independently against all others, employing regularization techniques to ensure sparsity. This sparsity indicates conditional independence among variables. Conversely, Reduced Rank regression addresses multicollinearity and dimensionality reduction by applying a penalty to the regression coefficients. This penalty encourages a lower-rank coefficient matrix, thus simplifying the model while preserving its predictive abilities. Although both methods tackle high-dimensional challenges, their focuses differ: Nodewise regression is geared towards revealing network structures, whereas Reduced Rank regression focuses on efficient prediction among correlated variables. We outline the derivation of the precision matrix formula and its practical estimation using Nodewise regressions as detailed in section 2 in Callot et al. (2021).

3.2.1 Derivation of the Precision Matrix

Denote $\Sigma_{-i,-i}$ as the $N-1 \times N-1$ submatrix obtained by excluding the ith row and column from Σ. Let $\Sigma_{i,-i}$ represent the row vector corresponding to the ith row of Σ, excluding the ith element, and let $\Sigma_{-i,i}$ signify the column vector corresponding to the ith column of Σ, also excluding the ith element. Employing the formula for the inverse of block matrices, we can determine the ith principal diagonal element as

$$\Theta_{i,i} = (\Sigma_{i,i} - \Sigma_{i,-i}\Sigma_{-i,-i}^{-1}\Sigma_{-i,i})^{-1}, \tag{3.8}$$

and the ith row of Θ with the ith element removed is

$$\Theta_{i,-i} = -(\Sigma_{i,i} - \Sigma_{i,-i}\Sigma_{-i,-i}^{-1}\Sigma_{-i,i})^{-1}\Sigma_{i,-i}\Sigma_{-i,-i}^{-1} = -\Theta_{i,i}\Sigma_{i,-i}\Sigma_{-i,-i}^{-1}. \tag{3.9}$$

We demonstrate the connection between (3.8) and (3.9) to a linear regression framework. Consider $R_{-i,t}$ as the vector containing all demeaned returns, excluding the ith return.

Let γ_i be the coefficient vector of length $N-1$ that minimizes $\mathbb{E}\left[R_{i,t} - (R_{-i,t})'\gamma\right]^2$ across all $t = 1, \ldots, T$. The solution is obtained as follows

$$\gamma_i = \Sigma_{-i,-i}^{-1}\Sigma_{-i,i}, \tag{3.10}$$

by using strict stationary of the data. Using the symmetry of Σ and (3.10), we can re-write (3.9) as

$$\Theta_{i,-i} = -\Theta_{i,i}\gamma_i'. \tag{3.11}$$

Define $E_{i,t} := R_{i,t} - (R_{-i,t})'\gamma_i$. By (3.10), we can verify that

$$\begin{aligned} \mathbb{E}\Big[R_{-i,t}E_{i,t}\Big] &= \mathbb{E}\Big[R_{-i,t}R_{i,t}\Big] - \mathbb{E}\Big[R_{-i,t}(R_{-i,t})'\Big]\gamma_i \\ &= \Sigma_{-i,i} - \Sigma_{-i,-i}\Sigma_{-i,-i}^{-1}\Sigma_{-i,i} = 0. \end{aligned} \tag{3.12}$$

Compiling the findings discussed previously, we formulate a regression model where the covariates are orthogonal to the error terms, as follows:

$$R_{i,t} = (R_{-i,t})'\gamma_i + E_{i,t}. \tag{3.13}$$

Based on (3.11) and (3.13), it is evident that $\Theta_{i,-i}$, and consequently the entire row Θ_i, exhibits sparsity precisely when γ_i is sparse.

Leveraging the outcomes described above, we proceed to deduce a formula for Θ. Incorporating (3.10), (3.12), and (3.1) into our calculations, we arrive at the following expression

$$\begin{aligned}\Sigma_{i,i} = \mathbb{E}\Big[R_{i,t}\Big]^2 &= \gamma_i' \Sigma_{-i,-i}\gamma_i + \mathbb{E}\Big[E_{i,t}^2\Big] \\ &= \Sigma_{i,-i}\Sigma_{-i,-i}^{-1}\Sigma_{-i,i} + \mathbb{E}\Big[E_{i,t}\Big]^2.\end{aligned} \tag{3.14}$$

Define $\psi_i^2 := \mathbb{E}\Big[E_{i,t}^2\Big]$. By (3.14),

$$\psi_i^2 = \Sigma_{i,i} - \Sigma_{i,-i}\Sigma_{-i,-i}^{-1}\Sigma_{-i,i} = \frac{1}{\Theta_{i,i}}, \tag{3.15}$$

where we use (3.8) for the second equality. Next, define a $N \times N$ matrix

$$\Gamma_1 := (I - \Gamma) := \begin{bmatrix} 1 & -\gamma_{1,2} & \dots & -\gamma_{1,N} \\ -\gamma_{2,1} & 1 & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots \\ -\gamma_{N,1} & -\gamma_{N,2} & \dots & 1 \end{bmatrix},$$

and define the diagonal matrix $\Psi = diag(\psi_1^2, \cdots, \psi_N^2)^{-1}$. Finally, we get

$$\Theta = \Psi\Gamma_1, \tag{3.16}$$

since (3.15) establishes that $\Theta_{i,i} = \frac{1}{\psi_i^2}$, and by (3.11), $\Theta_{i,-i} = \Theta_{i,i}\gamma_i' = \frac{-\gamma_i'}{\psi_i^2}$.

3.2.2 Nodewise Regression Estimation of the Precision Matrix

Drawing on the work of Van de Geer et al. (2014) and Chan et al. (1999), this paragraph presents the Nodewise regression estimation algorithm. The algorithm is specified for each $i = 1, \cdots, N$, as

$$\hat{\gamma}_i = \underset{\gamma \in \mathbb{R}^{N-1}}{\operatorname{argmin}} \left(||R_i - \mathbf{R}_{-i}\gamma||_2^2/T + 2\lambda_i||\gamma||_1 \right), \tag{3.17}$$

where $\hat{\gamma}_i = \{\hat{\gamma}_{i,k}; k = 1, \cdots, N; k \neq i\}$ is the vector of length $(N-1)$ of regression coefficient estimates which are utilized in formulating the precision matrix estimate. The parameter λ_i, a positive scalar, functions as a tuning parameter influencing the magnitude of the penalty applied to the coefficients. We denote S_i as the collection of indices corresponding to nonzero coefficient estimates within row γ_i (as shown in (3.11)), with $s_i := |S_i|$ indicating the number of elements in S_i (its cardinality).

To construct the nodewise regression estimator $\hat{\Theta}$ for the precision matrix Θ, we begin by defining

$$\hat{\Gamma}_1 := \begin{pmatrix} 1 & -\hat{\gamma}_{1,2} & \dots & -\hat{\gamma}_{1,N} \\ -\hat{\gamma}_{2,1} & 1 & \dots & -\hat{\gamma}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ -\hat{\gamma}_{N,1} & -\hat{\gamma}_{N,2} & \dots & 1 \end{pmatrix},$$

and write $\hat{\Psi} = diag(\hat{\psi}_1^2, \cdots, \hat{\psi}_N^2)^{-1}$, the $N \times N$ diagonal matrix, with entries

$$\hat{\psi}_i^2 := \frac{||R_i - \mathbf{R}_{-i}\hat{\gamma}_i||_2^2}{T} + 2\lambda_i||\hat{\gamma}_i||_1. \tag{3.18}$$

Next, we establish the approximate inverse $\hat{\Theta} := \hat{\Psi}\hat{\Gamma}_1$ to serve as the estimator for (3.16). It's important to note that although $\hat{\Sigma}$ is self-adjoint, $\hat{\Theta}$ lacks this property.

The selection of the penalty parameter λ_i, as detailed in (3.17), is determined through the optimization of the Generalized Information Criterion (GIC), as proposed by Fan and Tang (2013):

$$GIC(\lambda_i) := \log\left(\hat{\sigma}_{\lambda_i}^2\right) + \left|\hat{S}_i\left(\lambda_i\right)\right| \frac{\log(N)}{T} \log\left(\log(T)\right), \tag{3.19}$$

where $\hat{\sigma}_{\lambda i}^2 := ||R_i - \mathbf{R}_{-i}\hat{\gamma}_i||_2^2/T$ as the residual variance associated with asset i, with $|\hat{S}_i(\lambda_i)|$ denoting the estimated count of nonzero parameters within the vector $\hat{\gamma}_i$. The Generalized Information Criterion (GIC) is employed to determine the optimal value of λ_i. As demonstrated in Corollary 1 and Theorem 2 by Fan and Tang (2013), the GIC is effective in accurately identifying the true model with a probability that approaches unity, applicable in scenarios where $N > T$ as well as $N \leq T$.

The following algorithm outlines the procedure for calculating $\hat{\Theta}$:

Algorithm 2: Computation of the Nodewise Estimator $\hat{\Theta}$

1: **for** $i = 1$ to N **do**
2: Estimate $\hat{\gamma}_i$ for a given λ_i by solving the optimization problem defined in (3.17).
3: Choose λ_i using the Generalized Information Criterion (GIC) as defined in (3.19).
4: **end for**
5: Compute $\hat{\Psi}$ using the estimates obtained from the previous steps.
6: Compute $\hat{\Gamma}_1$ using the selected λ_i for each asset.
7: Return the nodewise estimator of the precision matrix $\hat{\Theta} = \hat{\Psi}\hat{\Gamma}_1$.

Upon computing the optimal $\hat{\Theta}$, we employ it to calculate portfolio weights in accordance with the procedures detailed in Section 3.4.

3.3 Equally Weighted Portfolio

For our third (benchmark) method, we adopt the equally weighted (1/N) portfolio approach. Often referred to as the 'naive' portfolio, this strategy assigns equal weights to all stocks, eliminating the need for complex covariance matrix computations. The appeal of the equally weighted portfolio lies in its simplicity and inherent diversification benefits. By allocating equal portions of the portfolio to each asset, it naturally mitigates unsystematic risk associated with individual stocks. Moreover, this strategy can be particularly effective in markets where less information is available or when predicting future returns and relationships between assets is challenging.

DeMiguel et al. (2009) assessed the out-of-sample performance of the sample-based mean-variance model and its adaptations intended to minimize estimation errors, comparing these against the baseline 1/N portfolio strategy. Their extensive evaluation, encompassing 14 models across seven empirical datasets, failed to identify any model that consistently surpassed the 1/N rule with respect to Sharpe Ratio, certainty-equivalent return, or turnover. This outcome suggests that the theoretical benefits of optimal diversification are significantly counterbalanced by estimation errors when applied to out-of-sample data.

Implementing the equally weighted portfolio strategy is remarkably straightforward. By allocating equal weights to all assets and applying these weights to the out-of-sample asset return data, one can readily compute the corresponding portfolio returns. The equation for the weights in the 1/N portfolio is:

$$w_i = \frac{1}{N} \text{ with } \sum_{i=1}^{N} w_i = 1, \tag{3.20}$$

where w_i is the weight of the ith asset in the portfolio.

3.4 Optimal Portfolio Allocation and Risk Assessment

A portfolio is defined by a vector of weights $w = (w_1, \cdots, w_N)' \in \mathbb{R}^N$, which denotes the proportion of total wealth allocated to each asset. To satisfy the full investment constraint, the sum of these weights must equal 1, expressed mathematically as $w'1_N = 1$ where $1_N = (1, \cdots, 1)'$. Throughout this paper, we permit short-selling, allowing the weights to take on negative values.

3.4.1 Global Minimum Variance Portfolio Weights

In this subsection, we use the optimal Γ derived from the objective function detailed in equation (3.4) from the preceding subsection. By substituting this optimal Γ in (3.3), we directly calculate the precision matrix. Subsequently, this precision matrix is utilized to ascertain the weights for both the GMV portfolio and the Markowitz Mean-ariance portfolio.

The GMV portfolio is characterized by the set of weights w_gmv that minimizes the portfolio's variance $w'\Sigma w$. The GMV weights are given by:

$$\hat{w}_{gmv} = \underset{w}{\text{argmin}}(w'\Sigma w), \text{ such that } w'1_N = 1. \tag{3.21}$$

Define $A = 1'_N \Theta 1_N/N$, with $\Theta = \Sigma^{-1}$ The estimation of A is represented by $\hat{A} = 1'_N \hat{\Theta}^{-1} 1_N/N$. According to Fan et al. (2008), the global minimum variance $\Phi_{gmv} = w'_{gmv}\Theta w = (NA)^{-1}$ is estimated by $\hat{\Phi}_{gmv} = (N\hat{A})^{-1}$. The variance estimation error for the GMV portfolio is calculated as:

$$|\frac{\hat{\Phi}_{gmv}}{\Phi_{gmv}} - 1|. \tag{3.22}$$

The weights of the GMV portfolio are given by:

$$w_{gmv} = \frac{\Theta 1_N}{1'_N \Theta 1_N} = \frac{\Theta 1_N/N}{A}, \tag{3.23}$$

and are estimated by:

$$\hat{w}_{gmv} = \frac{\hat{\Theta} 1_N}{1'_N \hat{\Theta} 1_N} = \frac{\hat{\Theta} 1_N/N}{\hat{A}}. \tag{3.24}$$

The weight estimation error is defined as $||\hat{w}_{gmv} - w_{gmv}||$ and the portfolio risk estimation error is calculated as $|\hat{w}_{gmv}(\hat{\Sigma} - \Sigma)\hat{w}_{gmv}|$.

3.4.2 Markowitz Portfolio Weights

The portfolio selection problem outlined by Markowitz (1952), entails identifying the portfolio with the minimal variance for a specified expected return, denoted as $\rho_1 > 0$. At any given time t, investors aim to ascertain the optimal portfolio weights by minimizing the mean-variance objective function:

$$\hat{w}_{mv} = \underset{w}{\text{argmin}}(w'\Sigma w), \text{ subject to } w'1_N = 1 \text{ and } w'\mu = \rho_1. \tag{3.25}$$

Define the terms $B = 1'_N \Theta \mu / N$, and $D = \mu' \Theta \mu / N$ and their estimates $\hat{B} = 1'_N \hat{\Theta} \hat{\mu} / N$, and $\hat{D} = \hat{\mu}' \hat{\Theta} \hat{\mu} / N$, with $\hat{\mu} := \bar{R} = n^{-1} \sum_{t=1}^{n} R_t$.

The solution to the constrained optimization problem outlined in 3.25 is:

$$w_{mv} \Sigma w_{mv} = N^{-1} \left[\frac{A\rho_1^2 - 2B\rho_1 + D}{AD - B^2} \right], \tag{3.26}$$

with the estimate for the above optimal portfolio variance is

$$N^{-1} \left[\frac{\hat{A}\rho_1^2 - 2\hat{B}\rho_1 + \hat{D}}{\hat{A}\hat{D} - \hat{B}^2} \right]. \tag{3.27}$$

Further, define the optimal portfolio variance as $\Psi_{mv} := N^{-1} \left[\frac{A\rho_1^2 - 2B\rho_1 + D}{AD - B^2} \right]$ and its estimate $\hat{\Psi}_{mv} := N^{-1} \left[\frac{\hat{A}\rho_1^2 - 2\hat{B}\rho_1 + \hat{D}}{\hat{A}\hat{D} - \hat{B}^2} \right]$. The variance estimation error for the Markowitz portfolio is calculated as:

$$|\frac{\hat{\Psi}_{mv}}{\Psi_{mv}} - 1|. \tag{3.28}$$

The solution of the Markowitz (1952) portfolio is:

$$w_{mv} = \frac{D - \rho_1 B}{AD - B^2} \left(\Theta 1_N / N \right) + \frac{\rho_1 A - B}{AD - B^2} \left(\Theta \mu / N \right). \tag{3.29}$$

Given that Σ is positive-definite, we have $A > 0$ and $D > 0$. By the Cauchy-Schwarz inequality, a solution exists for the system when $AD - B^2 > 0$. The estimation of the optimal weight vector w^* is accomplished through

$$\hat{w}_{mv} = \frac{\hat{D} - \rho_1 \hat{B}}{\hat{A}\hat{D} - \hat{B}^2} \left(\hat{\Theta} 1_N / N \right) + \frac{\rho_1 \hat{A} - \hat{B}}{\hat{A}\hat{D} - \hat{B}^2} \left(\hat{\Theta} \hat{\mu} / N \right). \tag{3.30}$$

Again, the weight estimation error is defined as $||\hat{w}_{mv} - w_{mv}||$ and the portfolio risk estimation error is calculated as $|\hat{w}_{mv}(\hat{\Sigma} - \Sigma)\hat{w}_{mv}|$.

3.4.3 Second Method for Minimum Variance Weights

We present an alternative approach for computing minimum variance weights, as utilized in our Empirical Analysis (Section 6.3). Leveraging Ψ, derived from Sections 3.1 and 3.2, we outline the calculation of the weights w_{mv2} in the following manner.

Algorithm 3: Second method for Minimum Variance Weights

1: **for** each optimal Γ **do**
2: Calculate weights:
$$\text{weights} \leftarrow \text{colSums}(\text{diag}(N) - \Gamma) \times \Psi$$
3: Sum of weights:
$$\text{weights\_sum} \leftarrow \text{sum}(\text{weights})$$
4: **if** weights_sum $\neq 0$ **then**
5: Normalize weights:
$$w_{mv2} \leftarrow \frac{\text{weights}}{\text{weights\_sum}}$$
6: **end if**
7: **end for**

4 Data

To facilitate a direct comparison of methods and build on existing research, this thesis employs the same dataset as used in Callot et al. (2021). In this study, we utilize both monthly and daily returns data from The Standard and Poor's 500 index (S&P 500), with the 3-month Treasury bill rate serving as the benchmark for risk-free returns. Regarding the monthly returns, our analysis period spans from January 1994 to May 2018. Similar to our approach with daily returns, we included only those stocks without missing values, yielding a dataset of 293 observations (T) and 304 stocks (N).

Upon reviewing the monthly data, we observed unusually high values for certain stocks, which skewed the performance metrics significantly. To ensure the accuracy of our study, we cross-referenced these outliers with historical data from Yahoo Finance (2021). This verification process revealed discrepancies, prompting us to exclude two specific stocks from our dataset: DXC Technology Co (DXC) and Welltower Inc. (WELL). Additionally, we rectified the dataset by nullifying two observations from BorgWarner Inc. (BWA), dated 1995-08-01 and 2016-11-01. Implementing these adjustments yields a dataset of 293 observations (T) and 302 stocks (N).

For the analysis of daily returns, we selected stocks that have complete data points from July 2, 2013, to April 30, 2018. This criterion resulted in a dataset comprising 1,216 observations (T) across 452 stocks (N).

After encountering high computational demands encountered in our analysis, we opted to focus on a subset of our daily dataset. Specifically, we limited our observations to the period between December 10, 2015, and April 30, 2018, resulting in a dataset encompassing 600 observations (T) and 452 stocks (N).

In the case of monthly data, we have $T < N$. Conversely, for daily data, the situation is reversed with $T > N$. We assess performance over two distinct out-of-sample periods for both daily and monthly data sets.

1. **Monthly Data:** January 1994 to May 2018 with $T = 293$ and $N = 302$.
   (a) In-Sample 1: January 1994-March 2010 ($T_I = 195$), Out-Of-Sample 1: April 2010-May 2018 ($T - T_I = 98$).
   (b) In-Sample 2: January 1994-March 2008 ($T_I = 173$), Out-Of-Sample 2: June 2008-May 2018 ($T - T_I = 120$).
2. **Daily Data:** December 10, 2015 to April 30, 2018 with $T = 600$ and $N = 452$.
   (a) In-Sample 1: July 2, 2013, to April 30, 2017, ($T_I = 348$), Out-Of-Sample 1: 1 May, 2017-April 30, 2018 ($T - T_I = 252$).
   (b) In-Sample 2: July 2, 2013-January 29, 2018 ($T_I = 537$), Out-Of-Sample 2: January 30, 2018-April 30, 2018 ($T - T_I = 63$).

Portfolios undergo rebalancing at the start of each new training window, during which both the expected return vector and the covariance matrices are recalculated. Consider, for example, a five-year ($T - T_I = 60$) rolling window forecast with a monthly interval. Here, we perform the estimation of expected returns and covariance matrices and construct both the GMV and Markowitz portfolios a total of 60 times. The portfolios are held for a month, followed by a rebalancing at the commencement of each subsequent month. For setting the return target in the mean-variance portfolio, we adopt a monthly goal of 0.7974% and a daily target of 0.0378% like in Callot et al. (2021). These targets are set to align with an annualized return rate of 10% when compounded.

5 Simulation Analysis

In this section, we start by presenting the performance metrics used to evaluate our simulation results. Subsequently, we delve into the specifics of the data-generating processes selected for our simulations, highlighting the dimensions under consideration. Finally, we analyze the results of the simulations.

5.1 Methods and Metrics

We present the outcomes of simulations conducted to evaluate the performance of Reduced Rank Regression-based portfolios against those constructed using two alternative methods: Nodewise regression and the equal weight (1/N) strategy. For each set of experiments, we analyze the portfolios based on three critical metrics: variance estimation error, weight estimation error, and risk estimation error.

The variance estimation error is calculated the absolute difference between the ratio of the estimated portfolio variance ($\hat{\Phi}_{gmv}$, $\hat{\Psi}_{mkw}$) to the actual portfolio variance (Φ_{gmv}, Ψ_{mkw}) and 1, as in 3.22, and 3.28. This metric quantifies the discrepancy in variance predictions, offering insight into the precision of our estimation process. A value of 0 would indicate a perfect estimation, while any deviation from 0 shows the level of estimation error.

The weight estimation error is determined by $||\hat{w} - w||$, which denotes the L1 norm, or the sum of the absolute differences, between the estimated weights ($\hat{w}_{gmv}$, $\hat{w}_{mkw}$) and the actual weights (w_{gmv}, w_{mkw}) of the portfolio. This measure assesses the accuracy with which the portfolio weights are estimated, reflecting the efficacy of the underlying covariance estimation. The closer this value is to 0, the more accurate the weight estimation.

The risk estimation error is calculated using $|\hat{w}(\hat{\Sigma} - \Sigma)\hat{w}|$. This formula involves the difference between the estimated covariance matrix ($\hat{\Sigma}$) and the actual covariance matrix (Σ), weighted by the estimated portfolio weights ($\hat{w}_{gmv}$, $\hat{w}_{mkw}$). It quantifies the error in risk estimation due to inaccuracies in the covariance matrix estimation. The absolute value of this calculation indicates the magnitude of risk estimation error, with a lower value signifying a more accurate risk estimation. It is important to clarify that the estimated covariance matrix ($\hat{\Sigma}$) is not derived as the inverse of the estimated precision matrix obtained through methods such as Reduced-Rank regression or Nodewise regression. Instead, this covariance matrix is meticulously estimated by calculating the covariance of the simulated returns for each sample. This approach ensures a direct estimation from the empirical data, providing a nuanced and accurate representation of the return dynamics, rather than relying on inverse operations from precision matrix estimations.

These metrics serve as the primary indicators of portfolio performance and are denoted as "Variance," "Weight," and "Risk" in the corresponding tables. The first column of each table identifies the forecasting method being evaluated: Reduced-Rank regression (RRR), our primary focus in this study; Nodewise regression (NDW), which employs an inverse-covariance matrix approach as detailed in subsection 3.2; and 1/N, the equally weighted portfolio approach outlined in subsection 3.3. It is worth noting that the 1/N portfolio strategy does not differentiate between the GMV and Markowitz portfolios. However, to ensure that readers can easily assess the performance outcomes of all methods side by side, we have included the results for the 1/N portfolio in both the GMV and Markowitz portfolio tables. Furthermore, it is not applicable to calculate the weight estimation error for the 1/N portfolio, as this approach does not involve the estimation of weights; rather, it uniformly allocates to each asset.

5.2 DGP

For the simulation, we employ two sets of data-generating processes (DGPs).

The initial set features a covariance matrix for returns, demonstrating a Toeplitz pattern. For every time period $t = 1, 2, \cdots, T$, it is presumed that the $N \times 1$ vector of excess returns R_t adheres to a multivariate normal distribution with a covariance matrix Ω. Within Ω, the entry at row $k \in 1, \cdots, N$, and column $l \in 1, \cdots, N$ is indicated by $\omega^{|k-l|}$, where ω set to 0.15 like in Callot et al. (2021). The first DGP within this set assumes a mean vector of excess returns μ equal to a zero vector 0_N and it is utilized in conjunction with the GMV portfolio.

For the second DGP in this set, the mean vector μ is generated from a normal distribution with a mean of zero and a variance of 0.0001. This particular DGP is applied to the Markowitz portfolio, setting the return target ρ_1 to 0.000376. This target corresponds to an annual return of 10% over 252 trading days.

In the second set of DGPs, we are leveraging the mean vector μ and covariance matrix Ω derived from the empirical data described in the next section 4. This method also assumes that the returns follow a multivariate normal distribution. By utilizing μ and Ω as parameters for our simulation, we can create realistic scenarios of future asset returns. The first DGP within this set is predicated on monthly data, whereas the second DGP focuses on daily data.

5.3 Dimensions

We carry out simulations for the four Data Generating Processes (DGPs) and the three methods - Reduced-Rank regression, Nodewise Regression, and the 1/N portfolio - outlined in previous sections, within a framework where $N = T/2$. This setup guarantees that the number of assets grows proportionally with the sample size, maintaining a condition where $N < T$. Simulations are performed for sample sizes $n = 100, 200, 400$. Considering the extensive computational demands of assessing the three methods across four distinct DGPs, we limit our analysis to 100 replications to balance thoroughness with feasibility.

5.4 Simulation Results

In the analysis of the results corresponding to the Toeplitz GDP, it is observable that the weight and risk metrics for the RRR estimator exhibit convergence toward zero with increasing sample sizes, highlighting an improvement in estimation accuracy and reliability in larger datasets. This trend, however, does not extend to the variance metric, which instead shows a positive correlation with the sample size. For example, as illustrated in Table 5.1, the variance estimation error increases from $0,29290$ with $T = 100$, $N = 50$ to $0,30336$ with $T = 200$, $N = 100$.

In contrast, all the performance metrics for the NDW estimator consistently converge to zero as the sample size expands, aligning with theoretical predictions and empirical evidence presented in Callot et al. (2021).

The 1/N portfolio also shows a trend toward zero with increasing sample sizes. Notably, the variance associated with the equal-weight (1/N) portfolio is significantly lower—more than twice as small—compared to those calculated for both the RRR and NDW estimators. The significantly lower variance estimation error of the 1/N portfolio compared to those constructed using the RRR or NDW methods reflects the trade-offs between simplicity and diversification against the sophistication and estimation risks of model-based strategies. While the 1/N strategy may lack the potential for optimization based on predictive signals, its robustness against model assumption errors and its inherent diversification benefit make it a competitive alternative, especially in uncertain or volatile market conditions DeMiguel et al. (2009).

In the analysis presented in Tables 5.3, 5.4, 5.5, and 5.6, which detail the GDP data derived from monthly and daily population parameters, it can be observed that the errors in estimating variance and risk for all evaluated methods tend toward zero as the sample size expands. However, an intriguing discrepancy is noted in the behavior of weight estimation risk, which exhibits a positive correlation

	T = 100, N = 50			T = 200, N = 100			T = 400, N = 200		
	Variance	Weight	Risk	Variance	Weight	Risk	Variance	Weight	Risk
RRR	0,29290	0,13428	0,00372	0,29614	0,10293	0,00142	0,30336	0,10264	0,00071
NDW	0,27659	0,12049	0,00316	0,26582	0,08468	0,00115	0,25691	0,06314	0,00040
1/N	0,12153	-	0,00401	0,08577	-	0,00142	0,05695	-	0,00047

Table 5.1: TOEPLITZ DGP: Global Minimum Variance Portfolio, N = T/2

	T = 100, N = 50			T = 200, N = 100			T = 400, N = 200		
	Variance	Weight	Risk	Variance	Weight	Risk	Variance	Weight	Risk
RRR	0,28442	0,20970	0,00368	0,29216	0,15882	0,00159	0,30470	0,13010	0,00071
NDW	0,26807	0,20227	0,00342	0,26054	0,14722	0,00107	0,25473	0,10674	0,00036
1/N	0,12153	-	0,00401	0,08577	-	0,00142	0,05695	-	0,00047

Table 5.2: TOEPLITZ DGP: Markowitz Mean-Variance Portfolio, N = T/2

with sample size. This observation appears counterintuitive, as the literature suggests that estimation errors, including those related to weight estimation, should diminish with larger sample sizes, thereby aligning more closely with the true population parameters. A potential explanation might reside in the possibility that the methods - RRR and NDW - do not fully capture the dynamics of the DGP. In that case, increasing the sample size might introduce more noise into the estimation process. This suggests that while these methods proficiently reduce variance and risk estimation errors with increased data, their capacity to accurately estimate weights under augmented sample sizes is compromised.

Additionally, an interesting observation emerges from the analysis of GMV portfolio, as shown in Table 5.3, and Table 5.5, where the variance estimation risk associated with the RRR estimator is found to be lower than that of the NDW estimator. Conversely, in the context of the Markowitz portfolio, detailed in Table 5.4, and Table 5.6, this relationship is inverted for the initial two sample sizes. A plausible explanation for these shifting dynamics could be attributed to the differences in the construction of the GMV and Markowitz portfolios. The Markowitz portfolio, which seeks to optimize the balance between risk and return, introduces complexities that may initially align well with the strengths of the NDW estimator, particularly its approach to managing covariance relationships between assets. This alignment could explain the NDW estimator's initial outperformance in smaller sample sizes for the Markowitz portfolio.

Furthermore, it is observed that the 1/N portfolio surpasses both methods in terms of these metrics. This outcome highlights the robustness of the 1/N strategy.

	T = 100, N = 50			T = 200, N = 100			T = 400, N = 200		
	Variance	Weight	Risk	Variance	Weight	Risk	Variance	Weight	Risk
RRR	0,27588	2,31922	0,00039	0,20789	3,50658	0,00017	0,18968	7,95864	0,00008
NDW	0,66566	1,60558	0,00017	0,65069	3,10998	0,00010	0,56289	8,07011	0,00010
1/N	0,11351	-	0,00023	0,07374	-	0,00014	0,05883	-	0,00011

Table 5.3: Monthly Population Parameters DGP: Global Minimum Variance Portfolio, N = T/2

	T = 100, N = 50			T = 200, N = 100			T = 400, N = 200		
	Variance	Weight	Risk	Variance	Weight	Risk	Variance	Weight	Risk
RRR	0,70235	3,65574	0,00914	0,54667	6,12173	0,00484	0,28179	9,87482	0,00282
NDW	0,55498	2,71877	0,00157	0,42676	4,01187	0,00135	0,37556	9,24353	0,00156
1/N	0,11351	-	0,00023	0,07374	-	0,00014	0,05883	-	0,00011

Table 5.4: Monthly Population Parameters DGP: Markowitz Mean-Variance Portfolio, N = T/2

	T = 100, N = 50			T = 200, N = 100			T = 400, N = 200		
	Variance	Weight	Risk	Variance	Weight	Risk	Variance	Weight	Risk
RRR	0,44552	2,67668	0,00001	0,35999	4,32142	0,00002	0,21956	5,84675	0,00001
NDW	0,46053	1,95415	0,00001	0,44953	3,23344	0,00000	0,22970	5,71926	0,00000
1/N	0,10672	-	0,00001	0,08417	-	0,00001	0,05354	-	0,00000

Table 5.5: Daily Population Parameters DGP: Global Minimum Variance Portfolio, N = T/2

	T = 100, N = 50			T = 200, N = 100			T = 400, N = 200		
	Variance	Weight	Risk	Variance	Weight	Risk	Variance	Weight	Risk
RRR	0,58405	2,01752	0,00005	0,49963	3,53984	0,00005	0,38141	6,46767	0,00004
NDW	0,37920	1,98713	0,00003	0,35983	3,23409	0,00003	0,0,17634	5,70807	0,00003
1/N	0,10672	-	0,00001	0,08417	-	0,00001	0,05354	-	0,00000

Table 5.6: Daily Population Parameters DGP: Markowitz Mean-Variance Portfolio, N = T/2

6 Empirical Analysis

In this section, we first outline the performance metrics used to evaluate our empirical results. Subsequently, we delve deeper into the data utilized. Finally, we analyze the results of the empirical study.

6.1 Methods and Metrics

In this section, we delve into a comparative analysis of the Reduced-Rank regression method, the NDW approach, and 1/N portfolio within the context of out-of-sample optimization applications. Our evaluation focuses on four key financial metrics: average portfolio return, portfolio variance, the corresponding Sharpe Ratio (hereafter referred to as SR), and portfolio turnover. Additionally, we assess portfolio performance both in scenarios with and without transaction costs taken into account.

We employ a rolling window technique to project out-of-sample returns, utilizing both Monthly and Daily data sets described in Section 4. For the Monthly data, we utilize a rolling window that projects one month into the future and another that forecasts six months ahead, representing a half-year period. Similarly, for the Daily data, we implement one method that predicts portfolio returns one day ahead and another that extends the forecast to five days ahead, effectively covering one trading week. Samples of size n are divided into two segments: an in-sample training sample, indexed by $(1 : T_I)$, and an out-of-sample testing sample, indexed by $(T_I + 1 : T)$.

The rolling window methodology works as follows: initially, the portfolio weights $\hat{w}_{T_I}$ are calculated in-sample for the interval $(1 : T_I)$. These weights are then multiplied by the returns at time $T_I + x$ $(x = 1, 5, 6)$, producing the out-of-sample portfolio returns $\hat{w}'_{T_I} R_{T_I+x}$. Subsequently, the window is advanced by one period to $(2 : T_I + 1)$ and the portfolio weights (w_{T_I+1}) for this new period are calculated. These updated weights are then multiplied by the returns at time $(T_I + (x + 1))$ to obtain

the portfolio returns ($\hat{w}'_{T_I+1}R_{T_I+(x+1)}$) for that period. This process is continuously repeated, moving the window forward one period at a time, until the end of the data sample is reached.

Excluding transaction costs, the average portfolio return and variance in the out-of-sample period are

$$\hat{\mu}_{os} = \frac{1}{T-T_I}\sum_{t=T_I}^{T-1}\hat{w}'_t R_{t+x} \text{ and } \hat{\sigma}^2_{os} = \frac{1}{(T-T_I)-1}\sum_{t=T_I}^{T-1}(\hat{w}'_t R_{t+x} - \hat{\mu}_{os})^2, \tag{6.1}$$

utilizing these two statistics, we calculate the corresponding Sharpe ratio

$$SR = \hat{\mu}_{os}/\hat{\sigma}_{os}. \tag{6.2}$$

The introduction of transaction costs necessitates an adjustment in the metrics' definition. We denote the transaction cost by c, which is consistently set at 10 basis points, in alignment with DeMiguel et al. (2009). According to Ban et al. (2018), the excess portfolio return at time t, factoring in transaction costs, is defined as follows

$$Return_t = \hat{w}'_t R_{t+x} - c(1+\hat{w}'_t R_{t+x})\sum_{j=1}^{N}|\hat{w}_{t+1} - \hat{w}^+_{t,j}|, \tag{6.3}$$

where $\hat{w}^+_{t,j}$ represents the adjusted weight of the j^{th} asset at time t, calculated as $\hat{w}_{t,j}(1+R^a_{t+1,j})/(1+R^a_{t+1,N})$. Here, $R^a_{t+1,j}$ denotes the excess return of the j^{th} asset over the risk-free rate, and $R^a_{t+1,N}$ signifies the overall portfolio's excess return including the risk-free rate.

With this framework for excess portfolio returns, we proceed to determine the portfolio's mean return and variance.

$$\hat{\mu}_{os,c} = \frac{1}{T-T_I}\sum_{t=T_I}^{T-1} Return_t \text{ and } \hat{\sigma}^2_{os,c} = \frac{1}{T-T_I}\sum_{t=T_I}^{T-1}(Return_t - \hat{\mu}_{os,c})^2, \tag{6.4}$$

the Sharpe Ratio, adjusted for transaction costs, becomes

$$SR_c = \hat{\mu}_{os,c}/\hat{\sigma}_{os,c}. \tag{6.5}$$

Finally, we take into account the portfolio turnover

$$PT = \frac{1}{T-T_I}\sum_{t=T_I}^{T-1}\sum_{j=1}^{N}|\hat{w}_{t+1} - \hat{w}^+_{t,j}|. \tag{6.6}$$

6.2 Data Specification

To calculate the performance metrics for the three methodologies, we utilized the dataset outlined in Section 4. Nevertheless, due to the extended computational durations encountered when processing the entire dataset, we adjusted our data incorporation approach within these methods. As previously noted, in the context of daily data, we encounter a scenario where $T > N$. However, this dynamic is inverted for monthly data, resulting in $T < N$. In such high-dimensional settings, the covariance matrix is typically singular (or nearly singular), meaning it does not have a full rank, and therefore, cannot be directly inverted. This singularity is because there are not enough observations to uniquely determine the relationships among all variables.

While Callot et al. (2021)'s contribution provided robust and consistent estimations for variance, weight, and risk metrics within both GMV and Markowitz portfolios, our attempts to achieve similar consistency

with the RRR method were unsuccessful (see 5.4). To address the complexities and challenges associated with high-dimensional data, we implemented a strategy of employing data subsets generated through sampling techniques. This approach involves randomly selecting 100 out of the 302 variables, with replacement, for each sample. Instead of analyzing the entire dataset, we focus our analysis on smaller subsets. This sampling process is repeated 10 times to ensure a robust representation of the full dataset. The decision to limit the number of samples to 10 was informed by our observation that the results obtained from 10 samples closely mirrored those derived from the entire dataset, thereby providing a balanced approach to comprehensively capture the dataset's characteristics while addressing the issue of high dimensionality.

	Return	Variance	SR	Turnover
With sampling	0,01216	0,00093	0,39933	0,12495
Without sampling	0,01265	0,00096	0,40759	0,11382

Table 6.1: Comparison of Performance Metrics Using Monthly Data: Nodewise Regression Estimator Across the Full Dataset and the Average of 10 Samples

To further alleviate the computational strain for daily data, we implemented a similar strategy of randomly sampling 150 stocks from this dataset, with replacement. We generated 10 distinct subsets in this manner and conducted our analysis on these subsets instead of the entire dataset.

6.3 Results

We present the empirical outcomes for the GMV portfolio, the Markowitz portfolio, and a second Minimum-Variance (MV2) portfolio, detailing their performance both with and without transaction costs (denoted as TC in the tables), utilizing the RRR and NDW estimators. Similar to the approach described in Section 5.4, the equally weighted (1/N) portfolio distributes weights uniformly across all assets, making the frameworks of the GMV and Markowitz portfolios not applicable to it. However, for ease of comparison and convenience, the results for the 1/N portfolio are included alongside those of the GMV and Markowitz portfolios in the same table. Additionally, turnover is utilized in the computation of performance metrics in scenarios where transaction costs are considered. Consequently, transaction costs are displayed in the table labeled 'without TC' and represented by '-' in the table marked 'with TC'.

Before diving into the analysis, it is crucial to acknowledge that the performance metrics of the RRR method, when applied to monthly data for both the GMV and Markowitz portfolios, appear to be less accurate in comparison to those observed for the 'MV2' portfolio. As a result, we will exclude the columns 'GMV' and 'MKW' in our analysis for the Monthly performance metrics.

In Table 6.2, we report the performance metrics for the 1-month ahead forecast of monthly portfolios over the full evaluation period from January 1994 to May 2018. Observing the 'MV2 Portfolio' column, it is evident that the RRR method is outperformed by both the NDW method and the 1/N portfolio in both samples, both in terms of the SR and Turnover. The 1/N portfolio consistently delivers the highest returns. However, upon adjusting for variance, the NDW method emerges as the leader in SR across both samples, also when transaction costs are considered. Notably, the 1/N portfolio excels in minimizing turnover, benefiting from its straightforward rebalancing process, which contrasts with the more frequent and complex adjustments required by other strategies that respond dynamically to market changes. Upon comparing the performance metrics for each method across the two samples, we observe a uniform decline in all metrics. Specifically focusing on the RRR method, there is a notable decrease in the SR, dropping from 0.36718 to 0.21630. Concurrently, there is an increase in turnover, moving from 0.16901 to 0.18915.

	GMV Portfolio				MKW Portfolio				MV2 Portfolio			
	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover
In-Sample: January 1994-March 2010, Out-Of-Sample: April 2010-May 2018, $\mathrm{T}_I = 195, T - T_I = 98$												
without TC												
RRR	0,02609	2,86753	0,05404	43,29120	0,04855	0,94030	0,14556	8,31776	0,01076	0,00086	0,36718	0,16901
NDW	0,01216	0,00093	0,39933	0,12954	0,01167	0,00094	0,38149	0,15932	0,01264	0,00095	0,41108	0,13945
1/N	0,01296	0,00135	0,35323	0,04494	0,01296	0,00135	0,35323	0,04494	0,01296	0,00135	0,35323	0,04494
with TC												
RRR	-0,33613	19,58128	-0,00254	-	0,00815	0,06710	0,11740	-	0,0105	0,00086	0,36164	-
NDW	0,01150	0,00093	0,37815	-	0,01086	0,00093	0,35572	-	0,01193	0,00094	0,38850	-
1/N	0,01274	0,00135	0,34732	-	0,01274	0,00135	0,34732	-	0,01274	0,00135	0,34732	-
In Sample: January 1994-May 2008, Out-Of-Sample: June 2008-May 2018, $\mathrm{T}_I - 173, T - T_I = 120$												
without TC												
RRR	-0,00534	22,66664	-0,01079	111,02030	-0,10010	17,27243	0,06009	54,79360	0,00734	0,00116	0,21630	0,18915
NDW	0,01034	0,00162	0,25773	0,14423	0,01007	0,00176	0,24132	0,17421	0,01071	0,00164	0,26486	0,15478
1/N	0,01159	0,00242	0,23623	0,05044	0,01159	0,00242	0,23623	0,05044	0,01159	0,00242	0,23623	0,05044
with TC												
RRR	-2,35056	1715,40	-0,08019	-	-3,820679	15977,7	0,036315	-	0,00715	0,00116	0,21068	-
NDW	0,00962	0,00162	0,23989	-	0,00920	0,00176	0,22072	-	0,00993	0,00164	0,24588	-
1/N	0,01134	0,00242	0,23114	-	0,01134	0,00242	0,23114	-	0,01134	0,00242	0,23114	-

Table 6.2: Comparative Analysis of Monthly Portfolio Performance Metrics for 1-Month Ahead Forecasting

In Table 6.3, we report the performance metrics for the 6-month (half-year) ahead forecast of monthly portfolios over the full evaluation period from January 1994 to May 2018. It is quite remarkable to observe that the performance metrics for both the NDW method and the 1/N portfolio surpass those reported in 6.2. Contrary to what one might intuitively expect—where forecasting over a longer horizon, such as 6 months instead of 1 month, typically leads to diminished results—we find an improvement. Notably, the return has increased and the variance has decreased, resulting in a significantly higher SR. For instance, for the NDW method, in the absence of transaction costs within the first sample, the SR escalates from 0.41108 to 0.47749, while the turnover also sees a marginal reduction, moving from 0.13945 to 0.13864. In contrast, for the RRR method, the performance metrics exhibit a slight decline for the 6-month forecast compared to the 1-month forecast. Furthermore, mirroring the observations in 6.2, the RRR method is again outperformed by the NDW and 1/N methods across all performance metrics in both samples.

In comparing the outcomes presented in Tables 6.2 and 6.3, it is observed that the variance increases for the larger out-of-sample (OOS) dataset. Theoretically, this is coherent, as augmenting the dataset (thereby diminishing the OOS dataset size) is expected to decrease the variance in the estimates. Simultaneously, a comparison between the first and second samples reveals a decline in returns. Nonetheless, an increase in variance, which indicates heightened uncertainty or risk in the forecasted returns, does not inherently correspond to a reduction in returns.

	GMV Portfolio				MKW Portfolio				MV2 Portfolio			
	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover
In-Sample: January 1994-March 2010, Out-Of-Sample: April 2010-May 2018, $\mathrm{T}_I = 195, T - T_I = 98$												
without TC												
RRR	0,03095	1,47207	0,08679	47,63061	0,01628	0,18719	0,15663	9,23645	0,01049	0,00089	0,35403	0,16612
NDW	0,01334	0,00083	0,46414	0,12965	0,01268	0,00081	0,44526	0,15518	0,01391	0,00085	0,47749	0,13864
1/N	0,01445	0,00118	0,42154	0,04333	0,01445	0,00118	0,42154	0,04333	0,01445	0,00118	0,42154	0,04333
with TC												
RRR	-0,06583	1,23946	-0,02522	-	-0,01090	0,05507	0,14394	-	0,01032	0,00089	0,34828	-
NDW	0,01268	0,00083	0,44132	-	0,01189	0,00081	0,41771	-	0,01321	0,00085	0,45328	-
1/N	0,01423	0,00118	0,41525	-	0,01423	0,00118	0,41525	-	0,01423	0,00118	0,41525	-
In Sample: January 1994-May 2008, Out-Of-Sample: June 2008-May 2018, $\mathrm{T}_I - 173, T - n_I = 120$												
without TC												
RRR	-0,08695	3,14309	-0,03689	83,59610	-1,87091	24,49309	0,08796	70,85446	0,00926	0,00114	0,27552	0,18431
NDW	0,01319	0,00136	0,35813	0,14363	0,01285	0,00152	0,33088	0,17036	0,01362	0,00137	0,36949	0,15343
1/N	0,01482	0,00199	0,33344	0,04845	0,01482	0,00199	0,33344	0,04845	0,01482	0,00199	0,33344	0,04845
with TC												
RRR	-0,65193	62,18214	-0,09293	-	-5,97528	36457,1	4,88923	-	0,00907	0,00114	0,27001	-
NDW	0,01247	0,00136	0,33888		0,01200	0,00152	0,30925		0,01285	0,00136	0,34890	
1/N	0,01457	0,00199	0,32807	-	0,01457	0,00199	0,32807	-	0,01457	0,00199	0,32807	-

Table 6.3: Comparative Analysis of Monthly Portfolio Performance Metrics for 6-Month Ahead Forecasting

In contrast to the monthly performance metrics presented in Tables 6.2 and 6.3, we include the columns 'GMV' and 'MKW' in our analysis for the daily performance metrics.

In Table 6.4, we present the performance metrics for the 1-day ahead forecast of daily portfolios, spanning the full evaluation period from December 10, 2015 to April 30, 2018. For the first sample, without TC, the RRR method secures the most favorable SR for the 'MV2 portfolio' among the 3 methods evaluated, $0,07726 > 0,06785 > 0,04629$. Regarding turnover, the 1/N portfolio consistently outperforms the others. Considering transaction costs, the 1/N portfolio stands out for its superior SR performance. This advantage stems from the 1/N method's minimal turnover, which ensures that returns are less diminished by the transaction costs incurred from portfolio rebalancing. Analysis reveals that, under the RRR method, returns fall from 0.00044 to 0.00031, whereas for the 1/N portfolio, the reduction is marginal, from 0.00050 to 0.00049. Meanwhile, the variances of both methods remain consistently stable. As a result, the SR for the 1/N portfolio notably exceeds that of the RRR method, demonstrating a comparison of 0.06662 to 0.05411.

Upon examining the outcomes for the 'GMV Portfolio' and 'MKW Portfolio', we notice that the RRR and NDW methods yield comparable results in terms of the SR. Yet, when transaction costs are considered, the NDW method surpasses the RRR method in SR efficiency, attributed to its lower turnover rate. Notably, the 1/N portfolio exceeds the performance of both the RRR and NDW methods across the 'GMV' and 'MKW' portfolios, irrespective of the inclusion or exclusion of transaction costs.

For the second sample, it is observed that the returns generated by all three methods across all three portfolios are negative. This phenomenon can be attributed to potential overfitting arising from the utilization of a larger in-sample dataset. Additionally, changing market conditions could also explain this outcome. Upon reviewing the historical returns of the S&P 500, as detailed in the appendix (A.1), for the relevant period, we observe notable volatility commencing from the beginning of 2018. This heightened volatility presents a plausible explanation for the markedly poorer performance of the three methods in the second sample compared to their performance in the first sample.

The observations from Tables 6.2 and 6.3 typically suggest that variance tends to increase with an expanded OOS dataset. However, in Tables 6.4 and 6.5, we encounter a contrasting scenario where the variance actually rises as the OOS dataset becomes smaller. This anomaly highlights a deviation from the expected pattern, indicating that the relationship between dataset size and variance may be influenced by additional factors not captured by a straightforward increase or decrease in dataset size.

	GMV Portfolio				MKW Portfolio				MV2 Portfolio			
	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover
In-Sample: December 10, 2015-April 28, 2017, Out-Of-Sample: April 29, 2017-April 30, 2018, $T_I = 348, T - T_I = 252$												
without TC												
RRR	0,00094	0,00028	0,06260	0,45464	0,00052	0,00013	0,04866	0,38984	0,00044	0,00003	0,07726	0,13046
NDW	0,00041	0,00005	0,05970	0,11861	0,00034	0,00005	0,05044	0,13335	0,00033	0,00005	0,04629	0,17044
1/N	0,00050	0,00005	0,06785	0,00895	0,00050	0,00005	0,06785	0,00895	0,00050	0,00005	0,06785	0,00895
with TC												
RRR	0,00049	0,00030	0,03925	-	0,00013	0,00013	0,01421	-	0,00031	0,00003	0,05411	-
NDW	0,00029	0,00005	0,04233	-	0,00021	0,00005	0,03080		0,00016	0,00005	0,02235	-
1/N	0,00049	0,00005	0,06662	-	0,00049	0,00005	0,06662	-	0,00049	0,00005	0,06662	-
In-Sample: December 10, 2015-Decmeber 1, 2018, Out-Of-Sample: December 2, 2018-April 30, 2018, $T_I - 527, T - T_I = 63$												
without TC												
RRR	-0,00124	0,00022	-0,08437	0,08078	-0,00123	0,00018	-0,09052	0,20814	-0,00031	0,00008	-0,03558	0,17311
NDW	-0,00089	0,00013	-0,07956	0,10208	-0,00093	0,00012	-0,08434	0,11887	-0,00080	0,00013	-0,07033	0,27622
1/N	-0,00090	0,00015	-0,07353	0,00996	-0,00090	0,00015	-0,07353	0,00996	-0,00090	0,00015	-0,07353	0,00996
with TC												
RRR	-0,00132	0,00022	-0,08981	-	-0,00144	0,00018	-0,10584	-	-0,00049	0,00008	-0,05491	-
NDW	-0,00140	0,00013	-0,12403		-0,00151	0,00012	-0,13738		-0,00215	0,00013	-0,18954	
1/N	-0,00091	0,00015	-0,07434	-	-0,00091	0,00015	-0,07434	-	-0,00091	0,00015	-0,07434	-

Table 6.4: Comparative Analysis of Daily Portfolio Performance Metrics for 1-Day Ahead Forecasting

In Table 6.5, we present the performance metrics for the 5-day (1 trading week) ahead forecast of daily portfolios, spanning the full evaluation period from December 10, 2015 to April 30, 2018. Contrasting with the monthly performance metrics, a noticeable decline in returns is observed when comparing the 1-day ahead forecasts (see 6.4) with the 5-day ahead forecast, as is expected.

For the 'MV2 Portfolio' in the first sample, without TC, we observe again that RRR method performs best in terms of SR, but now it is very close with the 1/N portfolio $0,06968$ to $0,06927$, the NDW method again falls a bit behind with a SR of $0,04833$. When incorporating transaction costs, the 1/N portfolio is again affected less by transaction costs corresponding to rebalancing due to the lowest turnover. Upon examining the outcomes for the 'GMV Portfolio' and 'MKW Portfolio,' it is noted that the RRR method generates high returns compared to the NDW method and the 1/N portfolio (specifically for the 'MKW Portfolio'). However, the increased variance associated with the RRR method results in a lower SR compared to the NDW method and the 1/N portfolio. When considering transaction costs, a similar trend is observed, with the SR of the 1/N portfolio experiencing a smaller decline compared to the RRR and NDW methods due to the small turnover. Particularly noteworthy is the substantial turnover associated with the RRR method, which leads to a more pronounced reduction in returns when contrasted with the NDW method.

For the second sample, the results are negative, mirroring the outcomes presented in Table 6.4.

	GMV Portfolio				MKW Portfolio				MV2 Portfolio			
	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover
In-Sample: December 10, 2015-April 28, 2017, Out-Of-Sample: April 29, 2017-April 30, 2018, $T_I = 348, T - T_I = 252$												
without TC												
RRR	0,00068	0,00028	0,05308	0,46230	0,00044	0,00013	0,04142	0,39117	0,00040	0,00003	0,06968	0,13178
NDW	0,00041	0,00005	0,05969	0,11830	0,00034	0,00005	0,05010	0,01321	0,00035	0,00005	0,04833	0,16994
1/N	0,00051	0,00005	0,06927	0,00894	0,00051	0,00005	0,06927	0,00894	0,00051	0,00005	0,06927	0,00894
with TC												
RRR	0,00023	0,00031	0,03051	-	0,00005	0,00013	0,00707	-	0,00027	0,00003	0,04665	-
NDW	0,00029	0,00005	0,04253	-	0,00021	0,00005	0,03071	-	0,00018	0,00005	0,02471	-
1/N	0,00050	0,00005	0,06806	-	0,00050	0,00005	0,06806	-	0,00050	0,00005	0,06806	-
In-Sample: December 10, 2015-Decmeber 1, 2018, Out-Of-Sample: December 2, 2018-April 30, 2018, $T_I - 527, T - T_I = 63$												
without TC												
RRR	-0,00073	0,00022	-0,04893	0,08240	-0,00064	0,00018	-0,04756	0,20732	-0,00021	0,00008	0,02422	0,18261
NDW	0,00012	0,00011	0,01115	0,10254	0,00012	0,00010	0,01144	0,11926	0,00021	0,00012	0,01928	0,27672
1/N	-0,00041	0,00015	-0,03342	0,00984	-0,00041	0,00015	-0,03342	0,00984	-0,00041	0,00015	-0,03342	0,00984
with TC												
RRR	-0,00081	0,00022	-0,05439	-	-0,00085	0,00018	-0,06316	-	-0,00039	0,00008	-0,04404	-
NDW	-0,00039	0,00011	-0,03712		-0,00047	0,00010	-0,04722		-0,00115	0,00012	-0,10692	
1/N	-0,00042	0,00015	-0,03421	-	-0,00042	0,00015	-0,03421	-	-0,00042	0,00015	-0,03421	-

Table 6.5: Comparative Analysis of Daily Portfolio Performance Metrics for 5-Day Ahead Forecasting

7 Discussion

This study is dedicated to the estimation of the precision matrix through Reduced Rank Regression (RRR) with the aim of constructing portfolio weights. Subsequently, it evaluates the performance of these portfolios using both daily and monthly historical returns from the S&P 500. To provide a comprehensive assessment of the RRR method, we benchmark its performance against that of the Nodewise regression (NDW) method, as introduced by Callot et al. (2021), and the equally weighted (1/N) portfolio strategy outlined in DeMiguel et al. (2009). In conducting this empirical analysis, we employ all three methods; however, the high computational demands of the RRR method within the R programming environment necessitated the imposition of certain limitations. These restrictions may have influenced the evaluated performance of our method.

Initially, to determine the optimal coefficient matrix, Gamma (Γ), which minimizes the objective function, I employed Generalized Cross Validation (GCV). The GCV Criterion assigns a score to each

lambda (λ) used. In this research, a series of 15 lambdas, distributed on a logarithmic scale, was selected. However, with access to more advanced computational resources, it would be advantageous to expand this series to include 100 or even 200 lambdas. By doing so, the likelihood of identifying a lambda that yields an even lower GCV score increases significantly. Such an enhancement in the selection process would lead to a more accurate estimation of the coefficient matrix Gamma (Γ) and, consequently, to improved portfolio performance.

Secondly, due to the objective function's non-differentiable nature, I employed a Projected Subgradient Method (PSM) to approximate the optimal Gamma (Γ) that minimizes this function. However, I limited the number of iterations in this PSM to 100. Increasing the iteration count to 1000, 2000, or even more would likely yield a more precise estimation of Gamma, thereby enhancing portfolio performance. For clarity and to illustrate potential improvements, I conducted a single run of the RRR method with 1000 iterations using monthly data. The outcomes of this run hint at promising enhancements in performance.

Thirdly, due to the significant computational demands, I utilized smaller subsets of the entire dataset to calculate both monthly and daily performance metrics. This approach was uniformly applied across all three methods (RRR, NDW, and 1/N), and it was demonstrated that the results were consistent across these subsets. Despite this consistency, directly testing these methods on the entire dataset would be preferable. Such comprehensive testing could potentially offer a more detailed understanding of the methods' performance characteristics.

Fourthly, to mitigate overfitting, we have incorporated the nuclear norm of the coefficient matrix Gamma $||\Gamma||_*$ as a penalizing term within the objective function. Future studies could explore the implementation of alternative penalty terms, such as the nuclear norm of the product of returns and Gamma $||R\Gamma||_*$ or, possibly in conjunction with, the squared Frobenius norm of Gamma $||\Gamma||_F^2$. It is important to consider that alternative tuning criteria, beyond GCV, should be utilized in the application of different penalty terms to identify the optimal lambda λ and its associated coefficient matrix Gamma Γ.

Moreover, the findings are derived from an empirical analysis utilizing daily and monthly returns spanning January 1994 to May 2018 for daily data, and December 10, 2015, to April 30, 2018, for monthly data. The period covered by the daily data encompasses significant market events, including the peak of the dot-com bubble in 2000 and the height of the Great Recession in 2008. In contrast, the timeframe for the monthly data lacks notable financial crises, which could impact the comparative performance of the portfolios. Consequently, it would be beneficial for future research to extend the analysis to more recent data and over a longer duration, potentially capturing the effects of crises on the efficacy of these methods.

Upon reviewing the results, it is evident that the NDW method delivers superior monthly performance across both samples, regardless of the inclusion or exclusion of transaction costs. Nevertheless, the performance gap between this method and our RRR approach is relatively narrow. Therefore, applying the three(/four) suggestions outlined earlier holds the potential to enhance the RRR's efficacy to a degree where it could surpass the NDW method in performance. In analyzing the daily data, it is noted that the 1/N portfolio exhibits robust performance, particularly when transaction costs are accounted for. Similar to the observations made regarding monthly performance, the potential for substantial enhancement of our RRR method exists. Such improvement could be realized through the adoption of a longer sequence of lambda values, increased iterations in the PSM, and the avoidance of reliance on sampled data.

8 Conclusion

This research aims to assess portfolio performance utilizing weights derived from the direct estimation of the precision matrix via Reduced-Rank Regression. In an empirical study, we examine this approach's

effectiveness using both daily and monthly historical returns, comparing it against the performance of the Nodewise regression method introduced by Callot et al. (2021) and the equally weighted (1/N) portfolio. Our findings indicate that, in the context of monthly data, Reduced-Rank Regression is outperformed by the other two methods, with Nodewise regression emerging as the superior method in terms of the Sharpe Ratio, even when accounting for transaction costs.

In the analysis of daily data without transaction costs, Reduced-Rank Regression demonstrates commendable performance relative to Nodewise regression. However, once transaction costs are factored in, the 1/N portfolio exhibits a stronger Sharpe Ratio performance across both out-of-sample datasets examined. This observation aligns with the conclusions drawn by DeMiguel et al. (2009), suggesting the challenge of surpassing the simplicity and effectiveness of a 'naive' allocation strategy in empirical settings.

Moreover, it is worth noting that the performance of Reduced-Rank Regression could be further enhanced by relaxing some of the constraints imposed in this study.

References

Ait-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399.

Ao, M., Y. L. and Zheng, X. (2017). Heteroskedasticity in asset returns and high dimensional portfolio selection.

Ban, G.-Y., El Karoui, N., and Lim, A. E. (2018). Machine learning and portfolio optimization. *Management Science*, 64(3):1136–1154.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate garch models: a survey. *Journal of applied econometrics*, 21(1):79–109.

Boyd, S., Xiao, L., and Mutapcic, A. (2003). Subgradient methods. *lecture notes of EE392o, Stanford University, Autumn Quarter*, 2004:2004–2005.

Bunea, F., She, Y., and Wegkamp, M. H. (2011). Optimal selection of reduced rank estimators of high-dimensional matrices.

Callot, L., Caner, M., Önder, A. Ö., and Ulaşan, E. (2021). A nodewise regression approach to estimating large portfolios. *Journal of Business & Economic Statistics*, 39(2):520–531.

Candes, E. and Recht, B. (2012). Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119.

Chan, L. K., Karceski, J., and Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *The review of Financial studies*, 12(5):937–974.

Chen, K., Dong, H., and Chan, K.-S. (2013). Reduced rank regression via adaptive nuclear norm penalization. *Biometrika*, 100(4):901–920.

De Nard, G., Ledoit, O., and Wolf, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*, 19(2):236–257.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953.

Ding, Y. (2017). *High dimensional minimum variance portfolio under factor model*. PhD thesis.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007.

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

Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197.

Fan, J., Furger, A., and Xiu, D. (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):489–503.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(4):603–680.

Fan, J., Liao, Y., and Shi, X. (2015). Risks of large portfolios. *Journal of Econometrics*, 186(2):367–387.

Fan, Y. and Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(3):531–552.

Green, J., Hand, J. R., and Zhang, X. F. (2013). The supraview of return predictive signals. *Review of Accounting Studies*, 18:692–730.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621.

Ledoit, O. and Wolf, M. (2004a). Honey, i shrunk the sample covariance matrix. *UPF economics and business working paper*, (691).

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

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices.

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and pca in large dimensions. *Journal of Multivariate Analysis*, 139:360–384.

Ledoit, O. and Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices.

Ledoit, O. and Wolf, M. (2022a). The power of (non-) linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*, 20(1):187–218.

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

Markowitz, H. M. (1952). Portfolio selection. *Journal of finance*, 7(1):71–91.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso.

Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling.

Stevens, G. V. (1998). On the inverse of the covariance matrix in portfolio analysis. *The Journal of Finance*, 53(5):1821–1827.

Van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models.

Yahoo Finance (2021). Yahoo finance - stock market live, quotes, business & finance news. `https://finance.yahoo.com`.

Yuan, M., Ekici, A., Lu, Z., and Monteiro, R. (2007). Dimension reduction and coefficient estimation in multivariate linear regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(3):329–346.

A Appendix

Figure A.1: S&P500 returns for the period December 10, 2015 - April 30, 2018