

Ledit, Olivier; Wolf, Michael

Working Paper

Markowitz portfolios under transaction costs

Working Paper, No. 420

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Ledit, Olivier; Wolf, Michael (2022) : Markowitz portfolios under transaction costs, Working Paper, No. 420, University of Zurich, Department of Economics, Zurich, <https://doi.org/10.5167/uzh-221804>

This Version is available at:

<https://hdl.handle.net/10419/266244>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich** ^{UZH}

University of Zurich
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 420

Markowitz Portfolios under Transaction Costs

Olivier Ledoit and Michael Wolf

October 2022

Markowitz Portfolios under Transaction Costs

Olivier Ledoit*

Department of Economics

University of Zurich

CH-8032 Zurich, Switzerland

olivier.ledoit@econ.uzh.ch

Michael Wolf[†]

Department of Economics

University of Zurich

CH-8032 Zurich, Switzerland

michael.wolf@econ.uzh.ch

October 2022

Abstract

Markowitz portfolio selection is a cornerstone in finance, both in academia and in the industry. Most academic studies either ignore transaction costs or account for them in a way that is both unrealistic and suboptimal by (i) assuming transaction costs to be constant across stocks and (ii) ignoring them at the portfolio-selection state and simply paying them ‘after the fact’. Our paper proposes a method to fix both shortcomings.. As we show, if transaction costs are accounted for (properly) at the portfolio-selection stage, net performance in terms of the Sharpe ratio increases, particularly so for high-turnover strategies.

KEY WORDS: Covariance matrix estimation, mean-variance efficiency, multivariate GARCH, portfolio selection, transaction costs.

JEL CLASSIFICATION NOS: C13, G11.

*Second affiliation: AlphaCrest Capital, 120 W 45th St, New York NY 10036, USA.

[†]Corresponding author. Mailing address: Zürichbergstrasse 14, CH-8032 Zurich, Switzerland.

1 Introduction

[Markowitz \(1952\)](#) portfolios are a cornerstone of finance, both in the academic literature and in the asset-management world. Since the two main input parameters, the vector of expected returns and the covariance matrix of the returns in a universe of assets, are unknown in practice, the true (or ideal) portfolios are infeasible and have to be estimated instead. To this end one most commonly applies the plug-in method: use the estimated vector of expected returns and the estimated covariance matrix in place of the true quantities.

When several methods to estimate a given portfolio under consideration are available, it is generally of interest to determine which method is ‘best’. In particular when researchers come up with a new method, they generally want to demonstrate that their method is ‘better’ than existing methods. One way to answer this question is via Monte Carlo studies. Unlike in real life, in a Monte Carlo study the true portfolio is known and so one can determine how close (on average) a given method gets to the truth. But any Monte Carlo study is based on a data generating process (DGP) chosen by the researchers, which leaves room to tweak this process in their favor. Therefore, the more common way to answer the question is via backtest exercises: One applies the various methods to a set of real data in a realistic manner (meaning that at any given point only data prior to that point are used to estimate the portfolio) over a sufficiently long period, which results in a series of (pseudo) out-of-sample (oos) returns for each method. Then the various oos return series are used for performance evaluation.

Which criterion to use in the evaluation depends on the portfolio under consideration. In the case of the Global Minimum Variance (GMV) portfolio, arguably, the most relevant criterion is “risk” measured by the sample variance or, equivalently, the sample standard deviation of the oos return series. The respective sample numbers then provide a ranking of the various methods (with smaller numbers resulting in a higher ranking). In addition, one can use hypothesis testing to examine whether one method significantly outperforms another one; for example, see [Ledoit and Wolf \(2011\)](#). On the other hand, a ‘general’ Markowitz portfolio not only seeks a small “risk” but also a high “reward”. Whereas the former is quantified by the standard deviation of the portfolio return, the latter is quantified by its expectation. In such cases, the preferred criterion is the reward-to-risk ratio, that is, the expected return divided by the standard deviation. When the returns are

taken in excess of the risk-free rate, this ratio is known as the Sharpe ratio; otherwise, the ratio is often called the information ratio. In a backtest exercise one then uses the sample average of the oos return series divided by the sample standard deviation. Again, one can use hypothesis tests to study whether one method significantly outperforms another one; for example, see [Ledoit and Wolf \(2008\)](#).

When the Sharpe ratio is used as the performance criterion, the question becomes whether to account for transaction costs or not. In our view, the answer is “it depends”. Although this distinction is rarely made in the literature, there can be two goals in performance evaluation.

The first goal, which one could term ‘accuracy’, is to evaluate the quality of estimated inputs in plug-in methods, namely of the estimated vector of expected returns and of the estimated covariance matrix; the idea being that the higher the quality of the estimated inputs, the higher will be the oos Sharpe ratio. Therefore, in this context, a backtest serves as an ‘indirect’ way of evaluating the quality of estimated inputs. (The ‘direct’ way, by comparing estimated inputs to true inputs, is not feasible, since the true inputs are not observable.) For this goal, transaction costs should *not* be taken into account, since doing so would muddy the waters. For example, consider the case where one wants to compare two estimators of the covariance matrix (holding the estimator of the vector of expected returns fixed). It could well be that one of the estimators, although less precise, leads to lower turnover and thus to a higher Sharpe ratio *after* transaction costs.

The second goal, which one could term ‘efficiency’ (shortcut for the mean-variance efficiency of a dynamically-rebalanced portfolio), is to evaluate actual investment strategies. Of course, for this goal, transaction costs should be taken into account. Arguably, an asset manager does not care about the quality of estimated inputs *per se* as long as they result in superior performance, which for him/her always means after transaction costs. Many academic studies take transaction costs into account, be it as the leading objective or as a robustness check. The problem is that the majority of such studies do so in a non-realistic and suboptimal way: (i) they assume transaction costs to be constant across stocks and (ii) they do not take transaction costs into account at the portfolio-selection stage.

Concerning (i), assuming that stocks have the same transaction costs makes life easier but contradicts reality. Therefore, we incorporate this feature, using a realistic off-the-shelf transaction-cost model from the literature. Concerning (ii), few portfolio managers in the real world ignore transaction costs at the portfolio-selection stage and simply

pay them ‘after the fact’. Doing so can be expected to result in worse net performance (that is, performance after transaction costs). Therefore, we propose a way to account for transaction costs at the portfolio-selection stage and to demonstrate that, indeed, doing so results in superior net performance.

The remainder of the paper is organized as follows. Section 2 specifies the portfolio formulations of interest and how we account for transaction costs at the portfolio-selection stage. Section 3 presents in great detail our backtesting technology including transaction costs. Section 4 presents the results from real stock return data. Section 5 concludes. An appendix contains some additional details and tables.

2 Portfolios

2.1 The Starting Point

Our starting point is the standard Markowitz portfolio-selection problem in its textbook version. There are N assets in the universe with corresponding return vector $r := (r_1, \dots, r_N)'$. Define the expected value and the covariance matrix of r as

$$\mu := \mathbb{E}(r) \quad \text{and} \quad \Sigma := \text{Cov}(r) .$$

The portfolio-selection problem in its textbook version is then formulated as

$$\min_w w' \Sigma w \tag{2.1}$$

$$\text{subject to } w' \mu \geq b , \text{ and} \tag{2.2}$$

$$w' \mathbb{1} = 1 , \tag{2.3}$$

where b is a selected target expected return and $\mathbb{1}$ denotes a conformable vector of ones.¹

The problem has the following analytical solution:

$$w = c_1 \Sigma^{-1} \mathbb{1} + c_2 \Sigma^{-1} \mu , \tag{2.4}$$

$$\text{where } c_1 := \frac{C - bB}{AC - B^2} \quad \text{and} \quad c_2 := \frac{bA - B}{AC - B^2} , \tag{2.5}$$

$$\text{with } A := \mathbb{1}' \Sigma^{-1} \mathbb{1} , \quad B := \mathbb{1}' \Sigma^{-1} \mu , \quad \text{and} \quad C := \mu' \Sigma^{-1} \mu . \tag{2.6}$$

¹One can also formulate the problem in its dual form: maximizing the expected return subject to an upper bound on the variance of the portfolio.

Since μ and Σ are unknown in practice, this solution is not feasible. But a feasible solution can be obtained by replacing μ with an estimator $\hat{\mu}$ and Σ with an estimator $\hat{\Sigma}$ in the analytical formulas: this is the plug-in method.

Although of academic interest, the plug-in version of the solution (2.4)–(2.6) is rarely used in real life. On the one hand, portfolio managers generally are interested in further constraints in addition to (2.2) and (2.3), such as gross-exposure, maximum((-absolute)-weight, industry-exposure, and factor-exposure constraints; of these, as a leading case, we will consider a gross-exposure constraint in our analysis below. On the other hand, the formulation (2.1)–(2.2) completely ignores transaction costs. In real life, the portfolio manager already holds a current portfolio and the question is how to ‘best’ update (or turn over) the portfolio. But updating (or turning over) a portfolio involves transaction costs, which are stock-specific. Of course, one can simply ignore transaction costs at the portfolio-selection stage, find the ‘best’ portfolio given one’s list of constraints, and then pay the resulting transaction costs ‘after the fact’. As can be expected theoretically, and as we will confirm empirically below, doing so is sub-optimal. It is generally advantageous to take transaction costs into account at the portfolio-selection stage already.

2.2 Accounting for Transaction Costs

There is no unique way to take transaction costs into account. One way would be to include an upper bound on such costs as an additional constraint; the question then becomes what that amount should be. Instead, we opt for adding a penalty term, defined as a multiple of the transaction cost, to the objective function (2.1). The question then becomes what the multiplier should be. We feel that the latter question is easier (or more natural) to address than the former and will offer a proposal below.

To start out, the more general problem formulation needed now must address the fact that one is moving from a ‘holding portfolio’ to a ‘new portfolio’, so the concept of ‘time’ needs to be introduced. In our empirical analysis below, we will use daily stock data, which is the norm in literature; but, as also is the norm, we do not update the portfolio on a daily basis but only every 21 trading days, which corresponds to our notion of a ‘month’.² As a result, the notation must distinguish between ‘return time’ and ‘portfolio-selection (or rebalancing) time’. Return days are indexed by $t \in \{1, \dots, T\}$, where T denotes the

²So a ‘month’ generally does not correspond to a calendar month.

sample size; portfolio-selection dates are indexed by $h \in \{1, \dots, H\}$, where H denotes the number of ‘months’. The connection is that t_h denotes the (return) day corresponding to portfolio-selection date h ; in particular, $t_{h+1} = t_h + 21$.

Consider a portfolio-selection date $h > 1$. The current vector of holding-portfolio weights is given by

$$w_{h-1}^* := (w_{h-1,1}^*, \dots, w_{h-1,N_h}^*)'.$$

This vector specifies the portfolio at the end of the holding period $h - 1$ and differs from w_{h-1} , which specifies the portfolio at the beginning of the holding period. The two vectors differ because of the different price evolutions of the various stocks during the holding period; see (3.4) for a ‘formula’ to obtain w_h^* for a general holding period h . Finally, N_h denotes the size of the ‘combined’ investment universe at portfolio-selection date h . This universe is given as the union of all stocks in the current holding portfolio and all stocks to be included in the new (or upcoming) portfolio: At any given date $h > 1$, one has to allow for some stocks leaving the portfolio and others entering the portfolio. So even if the number of stocks in the holding portfolio remains constant (denoted by N), as will be the case for us, $N_h \geq N$ generally varies over time. Of course, in case $N_h > N$, there will be (at least) $N_h - N$ positions in w_{h-1} , w_{h-1}^* , and w_h that are equal to zero.

The next ingredient is a vector of stock-specific transaction costs at portfolio-selection date h :

$$c_h := (c_{h,1}, \dots, c_{h,N_h})'.$$

The ‘unit’ of the transaction costs must match the ‘unit’ of the stock returns. Say a given transaction cost is 10 basis points (bps). If stock returns are ‘raw’ returns, then this number should be expressed as 0.001; on the other hand, if the stock returns are in percent, then this number should be expressed as 0.1 instead.³ When updating (or turning over) the portfolio from w_{h-1}^* to w_h the total transaction cost ‘per dollar’ is then given by

$$\tau_h := c_h' |w_h - w_{h-1}^*| := \sum_{i=1}^{N_h} c_{h,i} |w_{h,i} - w_{h-1,i}^*|. \quad (2.7)$$

Therefore, τ_h is a relative rather than an absolute cost. The absolute cost is obtained by multiplying the relative cost with the current net-asset value (NAV) of the portfolio. As an example, take case where τ_h equals 5 bps. If the NAV equals one hundred million

³Note that 100 bps correspond to one percent.

dollars, then the cost of turning over the portfolio is fifty thousand dollars; if the NAV equals one billion dollars, the cost is half a million dollars.

The advantage of expressing the transaction cost ‘per dollar’ instead of in absolute terms is that the penalty factor on the transaction cost, which is about to be introduced, can also be chosen ‘per dollar’ and, thus, does not have to take into account (or ‘worry about’) the current value of the portfolio.

To facilitate the upcoming methodology it will be convenient to partition the combined investment universe at portfolio-selection date h into two sets: the one containing the stocks in the new portfolio (of size N) and the one containing the stocks leaving the holding portfolio (of size M_h) with $M_h \geq 0$ and $N_h = N + M_h$. For further convenience, the stocks will be (re-)ordered such that the first N will effectively constitute the new portfolio and the last M_h will leave the holding portfolio. In this way, the total transaction cost (per dollar) can be decomposed in the sum of a variable cost (corresponding to the first N stocks) and a fixed cost (corresponding to the last M_h stocks):

$$\tau_h = \tau_h^{\text{var}} + \tau_h^{\text{fix}} := \sum_{i=1}^N c_{h,i} |w_{h,i} - w_{h-1,i}^*| + \sum_{i=N+1}^{N_h} c_{h,i} |w_{h-1,i}^*|. \quad (2.8)$$

The latter cost (which is zero in case $N_h = N$ and positive otherwise) is obviously fixed, since it does not depend on how the weights for the stocks in the new portfolio are selected. On the other hand, this selection affects the former cost, which is therefore variable.

The two final ingredients are $\hat{\mu}_h$ and $\hat{\Sigma}_h$, the estimators of μ_h and Σ_h , which, in slight abuse of notation, are defined as

$$\mu_h := \mathbb{E}(r_{t_h}^{\text{new}}) \quad \text{and} \quad \Sigma_h := \text{Cov}(r_{t_h}^{\text{new}}) \quad \text{and} \quad r_{t_h}^{\text{new}} := (r_{t_h,1}^{\text{new}}, \dots, r_{t_h,N}^{\text{new}})'$$

Clearly only the returns $r_{t_h}^{\text{new}}$ of the stocks in the ‘new’ investment universe are relevant, in the sense that we need estimators of the expectation and the covariance matrix of the vector of returns; the stocks leaving the holding portfolio are irrelevant. At this point, we assume nothing about the nature of $\hat{\mu}_h$ and $\hat{\Sigma}_h$; specific choices will be discussed in the empirical analysis below.

With the definition of τ_h^{fix} as in (2.8), our formulation of the portfolio-selection problem, both accounting for transaction costs and incorporating a gross-exposure constraint κ , is then given by:

$$\min_{w \in \mathbb{R}^N} w' \hat{\Sigma}_h w + \lambda \cdot \left(\sum_{i=1}^N c_{h,i} |w_i - w_{h-1,i}^*| + \tau_h^{\text{fix}} \right) \quad (2.9)$$

$$\text{subject to } w' \hat{\mu}_h \geq b_h, \quad (2.10)$$

$$w' \mathbb{1} = 1, \text{ and} \quad (2.11)$$

$$\|w\|_1 \leq \kappa. \quad (2.12)$$

The constant λ in (2.9) is the penalty parameter on the transaction cost ‘per dollar’; its choice will be important in practice. Furthermore, our formulation allows for the bound b_h on the expected return in (2.10) to vary with time.⁴ Finally, $\|\cdot\|_1$ in (2.12) denotes the L_1 norm of a vector, that is,

$$\|w\|_1 := \sum_{i=1}^N |w_i|.$$

Therefore, (2.12) represents a gross-exposure constraint where $\kappa \geq 1$ expresses the upper bound.⁵ In the extreme case $\kappa = 1$, the portfolio is long only, meaning that $w_i \geq 0$ for all i . The choice $\kappa = 1.6$ corresponds to a 130-30 portfolio whereas the choice $\kappa = 2$ corresponds to a 150-50 portfolios; such portfolios are becoming ever more popular in the financial industry. The choice $\kappa = \infty$ effectively removes the constraint (2.12) from the portfolio formulation.

Denote the solution of (2.9)–(2.12) by \hat{w} . Then w_h is given by $w_h := (\hat{w}_1, \dots, \hat{w}_N, 0, \dots, 0)'$, where the last $M_h \geq 0$ positions equal zero. To be careful, when portfolio-selection date $h + 1$ comes, w_h^* shall be re-ordered relative to w_h such that the last $M_{h+1} \geq 0$ positions correspond to the stocks that will leave the holding portfolio at date $h + 1$.

It should be pointed out that the problem (2.9)–(2.12) cannot be solved analytically, that is, there is no formula for \hat{w} (unless $\lambda = 0$ and $\kappa = \infty$). However, the problem is of convex nature and can be solved with off-the-shelf numerical-optimization software coming under the header “convex solver”. Even for $N = 1,000$, solving the problem only takes a matter of seconds (in the single digits) with modern computers.

⁴This is to allow for some variation in expected market conditions.

⁵One could, in principle, also allow κ to vary with time but doing so is uncommon.

2.3 Variations on the Theme

Many other portfolio formulations based on the plug-in method can be considered, where the formulations can be either ‘smaller’ or ‘larger’.

To go ‘smaller’, one can drop the expected-return constraint (2.10). Doing so has the advantage that one does not require an estimator of the expected return vector μ_h , which is not trivial to come by with. Note that by now, an entire ‘academic industry’, if the pun may be forgiven, is dedicated to finding such estimators, which sometimes are called “factors” or “(return-predictive) signals”; for example, see Green et al. (2013) and Harvey et al. (2016). If in addition the gross-exposure constraint (2.12) is dropped one arrives at the global minimum variance (GMV) portfolio accounting for transaction costs; setting $\lambda = 0$ in (2.9) gives the standard GMV portfolio (in its feasible version based on the plug-in method). But, again, the standard version is not of much interest in real life. On the one hand, the gross exposure can be anything and vary quite a bit from one portfolio-selection date to another, which would not be permitted for many portfolio managers. On the other hand, transaction costs are not accounted for. As for the general Markowitz portfolio including an expected-return constraint, in the empirical study, just like the GMV portfolio, we will compare the performance of the textbook version with versions accounting for transactions costs and/or incorporating a gross-exposure constraint.

To go ‘larger’, as one typical would in real life, one can add further constraints after (2.12), such as maximum-(absolute-)weight constraints (putting a common, say, upper bound on $|w_i|$ to avoid overexposure to any given stock), industry-exposure constraints, and factor-exposure constraints. Importantly, adding any such constraints (in their common forms) preserves the convexity of the optimization problem, so that \hat{w} can still be easily found with off-the-shelf numerical optimization software.

2.4 Getting Improved Performance

The feasible portfolio formulation (2.9)–(2.12) requires estimates of two input parameters, μ_h and Σ_h . The formulation is high-level in the sense that nothing is said about the nature of the estimators $\hat{\mu}_h$ and $\hat{\Sigma}_h$, which must be based on a (finite) history of past observations.

It is common knowledge which estimators *not* to use (without further modifications to the resulting solution \hat{w}): namely, the sample analogs, meaning the vector of sample

means (for $\hat{\mu}_h$) and the sample covariance matrix (for $\hat{\Sigma}_h$). Although unbiased, and maximum likelihood under a normality assumption, in settings where the number of stocks and the number of past observations are of similar magnitude (which is the setting most portfolio managers face) the sample analogs contain too much estimation error. Consequently, portfolios \hat{w} based on them perform poorly out of sample and are sometimes called “estimation error maximizers”, a term coined by [Michaud \(1989\)](#).

2.4.1 First Avenue: Markowitz with Smart Inputs

Broadly speaking, there are two avenues for obtaining portfolios that have improved oos performance. The first avenue is to use ‘sophisticated’ estimated input parameters $\hat{\mu}_h$ and $\hat{\Sigma}_h$, after which the solution \hat{w} is obtained in the standard way and not modified further. The second avenue is to stick with the sample analogs for $\hat{\mu}_h$ and $\hat{\Sigma}$ but then not to use the standard solution \hat{w} .

The first avenue is characterized by a ‘division of labor’. Some people aim to find improved estimators of the vector of expected returns; again, for example, see [Green et al. \(2013\)](#) and [Harvey et al. \(2016\)](#). Other people aim to find improved estimators of the covariance matrix; for example, see [Ledoit and Wolf \(2022a\)](#) and the references therein.⁶ At any rate, this avenue can be considered ‘generic’ in the sense that both the problem formulation and the method of solving it are common to all. The only way in which specific proposals differ is the way in which they estimate the input parameters.

2.4.2 Second Avenue: Non-Markowitz Alternatives

The second avenue is characterized by ‘custom-tailored’ approaches based on the particular problem formulation at hand. Some proposals start with the standard solution \hat{w} but then modify it further. Other proposals use the sample analogs $\hat{\mu}_h$ and $\hat{\Sigma}_h$ but use another way to derive portfolio weights from them compared to the standard solution. This is not the place to give a comprehensive overview over methods following the second avenue. But to illustrate ideas, we provide an example for either type of proposal in [Appendix A](#). The bottom line is that the non-Markowitz avenue does not seem amenable to cases

⁶Improved estimation of the covariance matrix is huge field in the academic literature and this not the place to give a comprehensive overview. Our point is that improved estimators, relative to the sample analog, exist and we take the liberty of highlighting our own proposals in this regard without necessarily claiming that they are (strictly) better than any other such proposals.

where universe dimension exceeds sample size (a frequent occurrence in practice), and where additional constraints or transaction costs have to be factored in. For these two reasons, we remain focused here on the first avenue of Markowitz with smart inputs.

3 Backtesting Methodology

3.1 Data and Universe Rules

We download daily stock return data from the Center for Research in Security Prices (CRSP) starting on 01-Jan-1982 and ending on 31-Dec-2021. We restrict attention to stocks from the NYSE, AMEX, and NASDAQ exchanges. Daily risk-free rates come from the supplemental series of the CRSP US Treasury Database via the Wharton Research Data Services portal.⁷

For simplicity, we adopt the common convention that 21 consecutive trading days constitute one ‘month’. The out-of-sample period ranges from 05-Jan-1988 through 31-Dec-2021, resulting in a total of $H = 408$ months (or $T = 8,568$ days). All portfolios are updated monthly.⁸ We index the portfolio-selection dates by $h = 1, \dots, 408$. At any date h , a covariance matrix is estimated based on the most recent 1,260 daily returns in the past, which roughly corresponds to using five years of past data.⁹

We now detail our rules for determining the investment universe at any given portfolio-selection date h . For a stock to be ‘eligible’, it must satisfy five criteria. The first criterion is not feasible in practice but common in academic analyses (and, basically, never kicks in anyhow for large stocks): The stock must have a complete return future over the next 21 days. Second, the day when the stock ‘started’, that is, the day where we see for the first time a non-missing value for the return of the stock must be at least 1,260 days in the past. Third, over the history of the past 1,260 days, there must be a most 2.5% missing values; any such missing value is then replaced with the return of the S&P 500 index on that day. Fourth, there must be no pairs of stock with respective return series over the

⁷The unique series identifier is 2000061, corresponding to “RISKFREE2 (MTH/DLY)”. The maturity/rebalancing label is “CRSP Risk Free – 4 week (Nominal)”.

⁸Monthly updating is common practice to avoid an unreasonable amount of turnover and thus transaction costs. During a month, from one day to the next, we hold number of shares fixed rather than portfolio weights; in this way, there are no transactions during a month.

⁹This rule implies our starting date of January 1987 for oos investing, since we need a history of about five years of data and NASDAQ stocks become available in CRSP from late 1982 onward only.

past 1,260 days that have a sample correlation exceeding 0.95. To ensure this criterion, we compute the sample correlation matrix of all the stock satisfying the first three criteria. If the largest correlation does not exceed 0.95, we keep all stocks. Otherwise, we identify the pair with the largest correlation and remove the stock with the smaller market value of the two at date h ; then we repeat the exercise until there are no more correlations exceeding 0.95. There is a final, fifth, criterion, which is not based on daily returns. For reasons that will become obvious below we also need price data in open, high, low, and close (OHLC) form. For a stock to be eligible, it needs to have all four pieces of information available for at least 240 out of the past 252 days. Then, at any date $h = 1$, the universe is comprised of the $N = 1,000$ largest eligible stocks in terms of their market value.

3.2 Modeling Transaction Costs

The goal of our paper is to develop a strategy that delivers desirable portfolio returns net of transaction costs. Importantly, we are interested in a strategy that will work well ‘now’, that is, from here on going forward. We are not interested in a strategy that would have worked well ‘in the past’, that is, over the last previous thirty years, say.

Consequently, we will use ‘historical’ stock returns in combination with ‘current’ transaction costs, the idea being that historical stock returns are (hopefully) representative of future ones and that current transactions are also representative of future ones, at least in the short- and medium-term future.

Apart from wanting to use ‘current’ transaction costs, we also want to be realistic and use stock-specific transactions costs. To this end, we use the model of [Briere et al. \(2020\)](#), which works by modeling stock-specific bid-ask spreads and then taking transaction costs to be a constant fraction (across stocks) of these spreads. In particular, they propose to model the log bid-ask spread of stock i at day t as a linear function of past log volatility. The problem is that volatility is unobserved even in hindsight, so in an initial step a feasible ‘version’ of volatility is taken to be the [Garman and Klass \(1980\)](#) intraday estimate using past OHLC price data; see Equation [11.15] of [Briere et al. \(2020\)](#):

$$\sigma_{t,i}^{\text{GK}} := \sqrt{\frac{1}{252} \sum_{s=1}^{252} \frac{1}{2} \log \left(\frac{H_{t-s,i}}{L_{t-s,i}} \right)^2 - (2 \log 2 - 1) \log \left(\frac{C_{t-s,i}}{O_{t-s,i}} \right)^2}. \quad (3.1)$$

If the complete OHLC information is not available for all past 252 days, we base the estimate on the corresponding formula applied only to the days when the complete

information is available, of which there must be at least 240 days by our rules above.

Based on a regression analysis the authors then propose the following linear approximation for the log bid-ask spread of stock i on day t (see their Table 11.3):

$$\log bas_{t,i} \approx -4.137 + 0.777 \log \sigma_{t,i}^{\text{GK}} ,$$

which results in the approximated transaction cost of stock i on day t as follows:

$$c_{t,i} := \frac{1}{2} \exp(-4.137 + 0.777 \log \sigma_{t,i}^{\text{GK}}) . \quad (3.2)$$

Here, using a constant multiplier across stocks, we equate transaction cost to one half of the bid-ask spread, which is a common convention. [Briere et al. \(2020\)](#) actually propose 0.4 instead of 0.5 as the multiplier (see their Table 11.1) but we prefer to stick to the commonly used and somewhat more conservative value of 0.5.

Finally, with some abuse of notation, the approximation we use in the problem formulation (2.9)–(2.12) is given by $c_{h,i} := c_{t_h,i}$, where it should be recalled that t_h denotes the day t that corresponds to portfolio-selection date h . Although an approximation only itself, it is certainly more realistic than assuming a constant trading costs across stocks, which is still the norm in the related academic literature. In our experience, the transaction-cost model of [Briere et al. \(2020\)](#) is a realistic place-holder by default, but any applied researcher who already holds a strong preference for some alternative model is welcome to use it instead.

To get an idea what ‘typical’ trading costs look like using approximation (3.2), we carry out the following exercise: For every portfolio-selection date $h = 1, \dots, 408$, take the vector of the $N = 1,000$ approximate trading costs corresponding to the stocks in the ‘new’ portfolio. Based on this vector, compute the minimum, the first decile, the first quartile, the median, the third quartile, the ninth decile, and the maximum. Then, for each of these seven quantiles $q \in \{0, 0.1, 0.25, 0.5, 0.75, 0.9, 1\}$, compute the average over the 408 portfolio-selection dates h ; note that, with some abuse of terminology, we call the minimum the 0 quantile and the maximum the 1 quantile. The results are reported in Table 3.1. As can be seen, for example, the median transaction cost is 3.9 bps on average. It can also be seen that there is considerable variation across transaction cost, as the 0.9 quantile is, in terms of the average, almost twice as large as the 0.1 quantile; therefore, assuming a constant transaction cost across stocks is unrealistic.

Quantile	0.0	0.1	0.25	0.5	0.75	0.9	1.0
Average	1.7	2.9	3.3	3.9	4.7	5.7	12.2

Table 3.1: Average quantiles of approximate trading costs; the unit is 1 basis point. At each portfolio-selection date $h = 1, \dots, 408$, the seven listed quantiles are computed from the vector of $N = 1,000$ approximate trading costs (of the stocks in the ‘new’ investment universe) according to formula (3.2). Then, for each quantile, the average over the 408 portfolio-selection dates is reported.

3.3 Computing Returns Net of Transaction Costs

We need to be careful about how we compute oos-ntc returns, (where the acronym oos-ntc stands for “out-of-sample and net-of-transaction-costs”), and we need to state our method clearly. The problem is that we have daily portfolio returns but update the portfolio only every ‘month’, that is, every 21 days. In terms of the average oos-ntc return it would not matter if we paid the transaction costs in full ‘at once’. But doing so would unduly affect the standard deviation of the returns (which also enters our performance criteria below), since only one out of twenty-one returns would take the ‘full hit’; so instead we should spread out evenly the incurred transaction cost over all 21 days in the upcoming holding period, that is, over the stretch $\{t_h, \dots, t_h + 20\}$. The cleanest way to do this is in terms of the net-asset value (NAV) series of the portfolio.

At the first portfolio-selection date $h = 1$, we (somewhat cavalierly) ignore transaction costs, since at this date the portfolio is not ‘updated’ but formed for the first time.¹⁰ Based on an (arbitrary) starting value X_0 , one invests amount X_0 according to weight vector w_1 at the beginning of day 1.

Since the value of X_0 is irrelevant for our purposes, we set it to $X_0 := 1$ for convenience. The initial stretch of the NAV series, namely $\{X_t\}_{t=1}^{21}$, is determined by letting the portfolio run for 21 days, keeping the number of shares (rather than the portfolio weights) fixed over time:

$$\forall t = 1, \dots, 21 \quad X_t = w_1' \Pi_{s=1}^t (1 + r_s) ,$$

¹⁰We are not investigating how to build a portfolio once from scratch but how to update a portfolio repeatedly through time.

where multiplication of two vectors is understood to be element-wise, that is,

$$(a_1, \dots, a_N)' \cdot (b_1, \dots, b_N)' := (a_1 b_1, \dots, a_N b_N)' .$$

By the principle of recursion, to define the NAV series over the entire oos period, it now is sufficient to specify our ‘recipe’ for a generic date $h > 1$. Denote by X_t the NAV of the portfolio at the end of day t where, of course, “net” also means “net of transaction costs”. The transaction cost incurred at portfolio-selection date h is given by $\tau_h \cdot X_{t_h-1}$ with τ_h given by (2.7). In an initial step, we derive a no-transaction-cost asset-value series $\{\tilde{X}_{t_h}, \dots, \tilde{X}_{t_h+20}\}$ as follows: Invest amount X_{t_h} according to weight vector w_h at the beginning of day t_h and let the portfolio run for 21 days, holding the number of shares (rather than the portfolio weights) fixed:

$$\forall j = 0, \dots, 20 \quad \tilde{X}_{t_h+j} = X_{t_h-1} \cdot w_h' \Pi_{s=0}^j (1 + r_{t_h+s}) .$$

The NAV series is then defined as

$$\forall j = 0, \dots, 20 \quad X_{t_h+j} := \tilde{X}_{t_h+j} - (\tau_h \cdot X_{t_h-1}) \frac{j+1}{21} .$$

This convention corresponds to paying off the transaction cost incurred evenly over the 21 days in the upcoming holding period.¹¹

In this way, a NAV series $\{X_t\}_{t=0}^T$ is obtained, with $X_0 = 1$ by our above convention. From the NAV series one then backs out the oos-ntc returns as follows:

$$\forall t = 1, \dots, T \quad x_t := \frac{X_t}{X_{t-1}} - 1 . \quad (3.3)$$

The series $\{x_t\}_{t=1}^T$ then forms the basis for performance evaluation.

Last but not least, in this context we can also detail how w_h^* is obtained, the vector of portfolio weights at the end of holding period h :

$$w_h^* \propto w_h \cdot \Pi_{s=0}^{20} (1 + r_{t_h+s}) , \quad (3.4)$$

where the symbol \propto stands for “proportional to”; since the weights need to some up to one, specifying the proportional weights is obviously sufficient.

¹¹It is tacitly assumed here that no interest will be charged by paying off the cost over the 21 upcoming days instead of paying it off in full ‘at once’ on date h .

4 Empirical Study

4.1 Investment Strategies

We consider four portfolio formulations:

- **GMV:** (2.9) and (2.11).
- **GMV-130-30:** in addition (2.12) with $\kappa = 1.6$.
- **Marko:** (2.9)–(2.11).
- **Marko-130-30:** in addition (2.12) with $\kappa = 1.6$.

To judge the benefit of accounting for transaction-costs at the portfolio-selection stage, the portfolios GMV and Marko are of leading interest us. Since portfolio managers in the industry often face additional constraints, we also include the portfolios GMV-130-30 and Marko-130-30 as a ‘real-life’ robustness check.

4.2 Estimators in the Horse Race

We consider two estimators $\hat{\Sigma}_h$ of the covariance matrix Σ_h :

- **NL:** the static nonlinear shrinkage estimator of [Ledoit and Wolf \(2022b\)](#), Section 4.5), termed QIS (Quadratic-Inverse Shrinkage) estimator.
- **DCC-NL:** the dynamic multivariate GARCH estimator of [Engle et al. \(2019\)](#). Note that we use the QIS estimator to estimate the correlation targeting matrix C instead of the QuEST function; see their Section 3.3.

This is not to imply that they are ‘state of the art’, although the underlying papers do provide evidence in this respect, but also because NL and DCC-NL are intrinsically compatible for the apples-to-apples comparison of going from a static to a dynamic estimation method *ceteris paribus*. In both cases we use the returns of the previous 1,260 days for the estimation.

As a simple-minded benchmark for the GMV portfolios, we include the equally-weighted portfolio, called $\mathbf{1}/N$, which *de facto* assumes that the covariance matrix is proportional to the identity matrix.

The portfolios Marko and Marko-130-30 need an estimator $\hat{\mu}_h$ in addition. For simplicity and reproducibility, we use the well-known momentum factor (or simply momentum for short) of [Jegadeesh and Titman \(1993\)](#). For a given portfolio-selection date h and a given stock, the momentum is the geometric average of the previous 252 returns on the stock but excluding the most recent 21 returns; in other words, one uses the geometric average over the previous ‘year’ but excluding the previous ‘month’. Collecting the individual momentums of all the N stocks contained in the ‘new’ universe then yields $\hat{\mu}_h$.

As a simple-minded benchmark for the Marko portfolios we include the equally-weighted portfolio of the top-quintile stocks (according to momentum). This portfolio is obtained by sorting the stocks, from lowest to highest, according to their momentum and then putting equal weight on all the stocks in the top 20%, that is, in the top quintile. We call this portfolio **EW-TQ**. We then use the value of b_h implied for the EW-TQ portfolio as input in the constraint (2.11), that is, b_h is the average momentum of the stocks in the EW-TQ portfolio.¹²

4.3 Performance Measurement

We report the following three out-of-sample performance measures for each scenario, where the acronym oos-ntc stands for “out-of-sample and net-of-transaction-costs” and the corresponding returns are defined in (3.3):

- **AV:** We compute the average of the 8,568 oos-ntc returns in excess of the risk-free rate and then multiply by 252 to annualize.
- **SD:** We compute the standard deviation of the 8,568 oos-ntc returns in excess of the risk-free rate and then multiply by $\sqrt{252}$ to annualize.
- **SR:** We compute the (annualized) Sharpe ratio as the ratio AV/SD.

As stated in the introduction, there are two possible goals for a backtest exercise: (i) the evaluation of the quality of estimated input parameters (like $\hat{\Sigma}_h$) and (ii) the evaluation of actual investment strategies. The goal in this paper is the second one and, therefore, for all four portfolio formulations considered, the leading performance criterion of interest is SR.¹³

¹²In this way, b_h varies over time, as it makes sense to use a lower value for b_h after a bear market than after a bull market.

¹³As an example for the first goal, if one wants to evaluate the quality of a covariance matrix estimator $\hat{\Sigma}_h$,

In addition, we report the following two summary statistics pertaining to the weight vectors:

- **TO:** Average monthly turnover defined as $\frac{1}{407} \sum_{h=2}^{408} \|w_{h-1}^* - w_h\|_1$.
- **GE:** Average gross exposure defined as $\frac{1}{408} \sum_{h=1}^{408} \|w_h\|_1$.

4.4 Transaction-Cost Penalty

Our problem formulations account for transaction costs via the penalty parameter λ in (2.9). Unfortunately, there is no theory dictating an ‘optimal’ choice of λ . Instead we try out various values in a certain grid and report the various measures for each choice. The grid considered is

$$\lambda \in \{0, 2.5, 5, 7.5, 10, 15, 20, 50, 100\} ,$$

where the choice $\lambda = 0$ corresponds to the *status quo* in the academic literature: Ignore transaction costs at the portfolio-selection stage and simply pay them ‘after the fact’; as we aim to show, doing so is sub-optimal with respect to SR. Although we cannot know *ex ante* which value of λ will work well in a given scenario, we expect that *ceteris paribus* NL will require a lower value of λ compared to DCC-NL, since NL incurs lower turnover; for example, see De Nard et al. (2021, Tables A.2 and A.4).

Detailed results can be found in Tables B.1–B.4 in the appendix. In order to mitigate the effects of “cherry picking” and “Monday-morning quarterbacking”, we limit ourselves to picking a common ‘good’ value of λ across the four different portfolios considered for each of the two covariance matrix estimators: NL and DCC-NL. The following general findings can be observed.

- Increasing the penalty parameter λ reduces TO. This is a strict monotonic relation which holds in every scenario, as can be anticipated from theory.
- Although the relation is not strictly monotonic, increasing λ tends to increase AV. This makes sense, since penalizing transaction costs more reduces them, which benefits the average return net of transaction costs.

the canonical method is as follows: Focus on the GMV portfolio without transaction costs, that is, portfolio formulation (2.9) with $\lambda = 0$ and (2.11); use oos instead of oos-ntc returns, that is, ignore transaction costs; then use SD as the leading performance criterion of interest.

- Although the relationship is not strictly monotonic, increasing λ tends to increase SD. This makes sense, since the objective is to minimize the variance of the portfolio return subject to a penalty on transaction costs; and the higher the penalty, the less ‘leeway’ one has in achieving the objective of minimizing the variance.
- The previous two relations imply, on balance, an inverse-U-shaped (and thus concave) relation between λ and SR: As λ increases, first SR rises and then it falls again. The location of the ‘peak’ depends on the scenario and, in particular, on the choice of the estimator of the covariance matrix. As expected, ‘good’ values of λ are larger for DCC-NL compared to NL.

Based on these findings, we select $\lambda = 5$ for NL and $\lambda = 10$ for DCC-NL.

4.5 Results

Armed with these choices, for any portfolio formulation we now present the results for five strategies: the simple-minded benchmark ($1/N$ respectively EW-TQ), NL with penalty parameters $\lambda \in \{0, 5\}$, and DCC-NL with penalty parameters $\lambda \in \{0, 10\}$. Further, to keep the column labeling in the tables compact, we use the following conventions: NL_λ stands for using NL with penalty parameter λ , whereas $DCC-NL_\lambda$ stands for using DCC-NL with penalty parameter λ . The results are presented in Tables 4.1 and 4.2.

Based on these tables, the following conclusions can be drawn:

- With a single exception (namely: NL-based GMV-130-30), penalizing the transaction costs (that is, using $\lambda > 0$) results in a higher SR.
- As expected, the improvements are larger for DCC-NL, since this method results in higher turnover, which ‘eats into’ the net returns more severely compared to NL if one simply pays the transactions costs ‘after he fact’.
- In both cases, GMV and Marko, the ‘raw’ portfolios deliver better performance compared to their 130-30 counterparts.¹⁴ But even for these counterparts choosing $\lambda > 0$ is generally beneficial. Arguably, the ‘raw’ portfolios are not suitable for

¹⁴DeMiguel et al. (2009a) show that constraining portfolio norms (such as imposing a gross-exposure constraint) improves performance when one uses the sample covariance matrix as the estimator of Σ_h . But if one uses a sophisticated estimator instead, such as NL or DCC-NL, this is obviously no longer the case.

	EW-TQ	NL ₀	NL ₅	DCC-NL ₀	DCC-NL ₁₀
GMV Portfolio					
AV	11.70	7.89	8.02	8.37	8.85
SD	18.55	8.44	8.51	7.05	6.38
SR	0.63	0.93	0.94	1.19	1.39
TO	0.10	1.00	0.30	2.46	0.76
GE	1.00	4.07	3.89	2.88	3.28
GMV-130-30 Portfolio					
AV	11.70	8.28	8.16	7.98	8.45
SD	18.55	9.21	9.32	6.86	6.54
SR	0.63	0.90	0.88	1.16	1.30
TO	0.10	0.34	0.17	1.63	0.82
GE	1.00	1.60	1.60	1.60	1.60

Table 4.1: Annualized performance measures (in percent) for various estimators of the GMV and GMV-130-30 portfolios. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 8,568 daily out-of-sample returns net of transaction costs and are in excess of the risk-free rate. Furthermore, TO stands for average monthly turnover and GE stands for average gross exposure.

mainstream players in the industry, such as mutual fund managers, because of their high (average) gross exposure, but they might still be suitable for specialized players who run big leverage ratios, such as hedge-fund managers.

- Although even for $\lambda = 0$ DCC-NL outperforms NL already, the improvement generally increases for $\lambda > 0$. For the GMV portfolio the improvement increases from 0.26 to 0.45 percentage points, whereas for the Marko portfolio it increases from 0.17 to 0.25 percentage points. There are similar improvements as one moves from the ‘raw’ portfolios to their 130-30 counterparts: from 0.26 to 0.42 percentage points for the GMV-130-30 portfolio and from 0.25 to 0.27 percentage points for the Marko-130-30 portfolio.
- For all four portfolio formulations, the winner is always the DCC-NL method with $\lambda = 10$. In particular, for the GMV and Marko formulations, this method more than doubles the SR of the simple-minded benchmark $1/N$ respectively EW-TQ.

	EW-TQ	NL ₀	NL ₅	DCC-NL ₀	DCC-NL ₁₀
Marko Portfolio					
AV	13.70	9.82	10.47	9.39	10.06
SD	21.72	9.69	9.85	7.98	7.65
SR	0.63	1.01	1.06	1.18	1.31
TO	0.56	1.44	0.61	2.80	1.04
GE	1.00	4.47	4.39	3.40	3.81
Marko-130-30 Portfolio					
AV	13.70	9.79	9.99	9.26	9.47
SD	21.72	11.57	11.63	8.39	8.41
SR	0.63	0.85	0.86	1.10	1.13
TO	0.56	0.80	0.60	1.67	1.10
GE	1.00	1.60	1.60	1.60	1.60

Table 4.2: Analogous to Table 4.1 but now for the Marko and Marko-130-30 portfolios.

DeMiguel et al. (2009b) study various portfolio-selection strategies based on estimated input parameters and conclude that “none is consistently better than the $1/N$ rule [net of transaction costs]”. We beg to differ.¹⁵

Remark 4.1 (Addressing statistical significance). Practitioners may wonder whether it is worth it to upgrade from the static estimator NL to the dynamic estimator DCC-NL. In addition to looking at the sample-based SR numbers, which are always higher for DCC-NL, one can also use formal hypothesis testing to address this question. For each of the four problem formulations, we can compute a p -value for the null hypothesis of equal (population) Sharpe ratios for two comparisons: comparing the two methods with $\lambda = 0$ and comparing the two methods with their respective ‘good’ value of λ , that is, $\lambda = 5$ respectively $\lambda = 10$. In other words, does taking transaction costs into account at the portfolio-selection stage ($\lambda > 0$) lead to increased statistical significance compared to simply paying them ‘after the fact’ ($\lambda = 0$)? To this end, for each of the eight comparisons, we compute a two-sided p -value with the HAC method of Ledoit and

¹⁵Part of the reason is that they consider an unduly large transaction cost of 50 bps (constant across stocks); such a cost may have been realistic decades ago but it is no longer anywhere near the costs investors face today. Another part of the reason is the particular list of strategies the authors consider, for reasons not clear, leaves out some promising contenders.

Wolf (2008, Section 3.1).¹⁶

The results are presented in Table 4.3. One can see that for each of the four portfolio formulations, there is indeed increased statistical significance as one moves from $\lambda = 0$ to $\lambda > 0$; in particular, there is one formulation (Marko) where one moves from “no significance” to “significance at the 5% level”. ■

Portfolio	$\lambda = 0$	$\lambda > 0$
GMV	0.04	0.00
GMV-130-30	0.01	0.00
Marko	0.12	0.02
Marko-130-30	0.02	0.01

Table 4.3: p -values for the null hypothesis of equal Sharpe ratios of the respective NL and DCC-NL methods. The column labeled “ $\lambda = 0$ ” compares NL and DCC-NL both with $\lambda = 0$, whereas the column labeled “ $\lambda > 0$ ” compares NL with $\lambda = 5$ and DCC-NL with $\lambda = 10$.

Remark 4.2 (Selecting λ for other strategies). It should be clear that a ‘good’ value of the transaction-cost penalty parameter λ is important. If, for a given scenario, one considers SR as a function of λ , then the good news is that based on our (necessarily limited) results, the function is rather flat near its ‘peak’ so that the exact choice of λ is not overly critical, as long as one gets the ‘neighborhood’ right. A partial backtest based on old data should provide enough guidance to pick a useful λ . Generally speaking, the more dynamic the covariance matrix estimator, the more we want to throttle down turnover by boosting up the λ penalty parameter, and the numbers we use above can serve as guidance. Alternatively, in practical applications, the portfolio manager can work backwards from an average turnover rate that he/she is comfortable with, and reverse-engineer through trial-and-error the value of λ that delivers it. ■

5 Conclusion

In this paper, we have proposed a method to account for transaction costs at the portfolio-selection stage in the context of Markowitz portfolio selection. Doing so increases the

¹⁶In particular, we use the pre-whitened version based on the QS kernel.

Sharpe ratio of the returns net of transaction costs compared to the *status quo* in the literature which amounts to ignoring the transaction costs at the portfolio-selection stage and simply paying them ‘after the fact’. Importantly, and also deviating from the *status quo* in the literature, we incorporate the fact that transaction costs are stock-specific instead of taking the easy route of assuming them to be equal. The benefit of our proposal is especially pronounced if dynamic estimators of the covariance matrix are used as inputs in the Markowitz formulations. The reason is that dynamic estimators, although being more accurate, typically generate higher turnover compared to static estimators and, therefore, investment strategies based upon them suffer more from transaction costs. We hope that our proposal will be of future use to anyone interested in devising Markowitz portfolio-selection strategies that have desirable performance net of transaction costs.

Even though the focus of this paper has been more on the covariance matrix, an obvious avenue for future research is to use the technology introduced here to further ‘tame the factor zoo’ (Feng et al., 2020) by weeding out those alphas that are over-dependent on high-t-cost stocks and turn over too much to pay for themselves — but only those. In this context, ignoring t-costs is too lenient, whereas charging a constant t-cost to every stock is unrealistic, so our method is the suitable ‘scalpel’.

References

- Ao, M., Yingying, L., and Zheng, X. (2019). Approaching mean-variance efficiency for large portfolios. *Review of Financial Studies*, 32(7):2890–2919.
- Briere, M., Lehalle, C.-A., Nefedova, T., and Raboun, A. (2020). Modeling transaction costs when trades may be crowded: A Bayesian network using partially observable orders imbalance. In Jurczenko, E., editor, *Machine Learning for Asset Management: New Developments and Financial Applications*, pages 387–430. John Wiley, Hoboken.
- 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., Nogales, F. J., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? *Review of Financial Studies*, 22:1915–1953.
- Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.
- Feng, G., Giglio, S., and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *Journal of Finance*, 75(3):1327–1370.
- Frahm, G. and Memmel, C. (2010). Dominating estimators for minimum-variance portfolios. *Journal of Econometrics*, 159(2):289–302.
- Garman, M. B. and Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 53(1):67–78.
- Green, J., Hand, J. R. M., and Zhang, X. F. (2013). The superview of return predictive signals. *Review of Accounting Studies*, 18:692–730.
- Harvey, C. R., Liu, Y., and Zhu, H. (2016). ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68.

- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.
- Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15:850–859.
- Ledoit, O. and Wolf, M. (2011). Robust performance hypothesis testing with the variance. *Wilmott Magazine*, September:86–89.
- 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. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.
- Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45:31–42.

A Non-Markowitz Approaches

A.1 Post-Processing the Standard Markowitz Solution

As a representative example of this category, [Frahm and Memmel \(2010\)](#) estimate the unconstrained GMV portfolio, which is the portfolio given by (2.1) and (2.3), dropping the constraint (2.2). This problem has the following analytical solution:

$$w^{\text{GMV}} = \frac{\mathbb{1}'\Sigma^{-1}}{\mathbb{1}'\Sigma^{-1}\mathbb{1}} .$$

Denote the sample covariance matrix by S . Then the corresponding plug-in solution is given by

$$\hat{w} = \frac{\mathbb{1}'S^{-1}}{\mathbb{1}'S^{-1}\mathbb{1}} .$$

(Here we tacitly assume that the number of observations exceeds the number of stocks, since otherwise S is not invertible.) Unless the number of stocks is very small relative to the number of observations, this solution is known to have poor oos performance. [Frahm and Memmel \(2010\)](#) propose to modify it as follows:

$$\hat{w}^{\text{mod}} := \alpha w^{\text{EW}} + (1 - \alpha)\hat{w} \quad \text{where} \quad w^{\text{EW}} := (1/N, \dots, 1/N)' .$$

Therefore, the proposed modification is a convex combination of the equally-weighted (or $1/N$) portfolio and the sample-based solution \hat{w} ; another way to look at it is that the sample-based solution \hat{w} is linearly ‘shrunk’ towards the $1/N$ portfolio. In practice, one has to work out an ‘optimal’ choice of the shrinkage intensity $\alpha \in [0, 1]$ and [Frahm and Memmel \(2010\)](#) offer a corresponding solution.

A.2 Specifying a Different Optimization Program

As a representative example of this category, consider the following problem formulation:

$$\max_{w \in \mathbb{R}^N} w'\mu \tag{A.1}$$

$$\text{subject to} \quad w'\Sigma w \leq \sigma^2 , \tag{A.2}$$

where σ^2 is an upper bound on the portfolio variance. This problem has the following analytical solution:

$$w^{\text{SR}} := \frac{\sigma}{\sqrt{\theta}} \Sigma^{-1} \mu \quad \text{where} \quad \theta := \mu' \Sigma^{-1} \mu . \tag{A.3}$$

(When the returns are expressed in excess of the risk-free rate, (A.3) is also known as the maximum Sharpe ratio (MSR) portfolio.) The standard feasible solution would be to plug in the sample mean vector for μ and the sample covariance matrix for Σ in (A.3); but, as discussed before, the resulting portfolio performs poorly out of sample. Instead the proposal of [Ao et al. \(2019\)](#) works as follows. First, the problem (A.1)–(A.2) can be re-expressed as an unconstrained population regression problem. Second, the population regression problem can be approximated by a sample analog (based on past observations). Third, the regression problem involves the unknown parameter θ which (under normality) can be estimated in an unbiased fashion as

$$\hat{\theta} := \frac{(T - N - 2)\hat{\theta}_s - N}{T}, \quad (\text{A.4})$$

where T denotes the number of observations, N denotes the number of stocks and $\hat{\theta}_s$ is the sample analog of θ . Fourth, the sample analog of the regression problem is estimated via the LASSO instead of OLS. A crucial feature is that at the end of the day no sophisticated estimators of μ and Σ are needed.

A.3 Overall Appraisal

Empirical studies have shown that Non-Markowitz avenues also enjoy improved oos performance compared to standard plug-in solutions based on sample analogs. But there are three (potential) issues. First, many such proposals do not work when $N > T$, that is, when there are more stocks than past observations. Second, it is not always clear how to incorporate additional constraints. Third, it is not always clear how to account for transaction costs. Let us illustrate these problems using the two examples from above:

- The method of [Frahm and Memmel \(2010\)](#) cannot be used when $N > T$, since it requires the sample covariance matrix to be of full rank. Also it does not seem possible to incorporate additional constraints or to account for transaction costs in any meaningful way.
- The method of [Ao et al. \(2019\)](#) also cannot be used when $N > T$, since then the formula (A.4) results in a negative estimator $\hat{\theta}$. Incorporating additional constraints seems possible ‘mechanically’ in the final step of their method (estimating a sample-based regression problem via the LASSO) but it is not clear whether the original

constraint (A.2) on the portfolio variance will be preserved in this way; a similar concern applies to accounting for transaction costs.

By contrast, the avenue described in Section 2.4.1 (Markowitz with smart inputs) can (i) be used when $N > T$ (as long as the estimator of Σ is positive definite, which all sophisticated estimators achieve) and (ii) incorporate additional constraints and account for transaction costs while preserving the original constraints. Therefore, the non-Markowitz avenue seems less promising for realistic dynamic portfolio management strategies.

B Additional Tables

λ	0	2.5	5.0	7.5	10	15	20	50	100
GMV Portfolio: NL									
AV	7.89	8.02	8.02	7.98	7.97	7.97	7.86	7.90	8.42
SD	8.44	8.47	8.51	8.58	8.64	8.79	8.91	9.59	10.29
SR	0.93	0.95	0.94	0.93	0.92	0.91	0.88	0.82	0.82
TO	1.00	0.43	0.30	0.24	0.20	0.16	0.13	0.07	0.05
GE	4.07	3.98	3.89	3.79	3.70	3.53	3.41	3.00	2.71
GMV Portfolio: DCC-NL									
AV	8.37	8.60	8.78	8.84	8.85	9.01	9.29	9.14	8.97
SD	7.05	6.57	6.39	6.36	6.38	6.47	6.56	7.13	7.76
SR	1.19	1.31	1.37	1.39	1.39	1.39	1.42	1.28	1.16
TO	2.46	1.41	1.05	0.87	0.76	0.62	0.54	0.34	0.25
GE	2.88	2.93	3.07	3.18	3.28	3.41	3.49	3.73	3.68

Table B.1: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 8,568 daily out-of-sample returns net of transaction costs and in excess of the risk-free rate. Furthermore, TO stands for average monthly turnover and GE stands for average gross exposure.

λ	0	2.5	5.0	7.5	10	15	20	50	100
GMV-130-30 Portfolio: NL									
AV	8.28	8.18	8.16	8.13	8.16	8.08	8.01	8.25	8.50
SD	9.21	9.25	9.29	9.32	9.36	9.43	9.49	9.86	10.36
SR	0.90	0.88	0.88	0.87	0.87	0.86	0.84	0.84	0.82
TO	0.34	0.22	0.17	0.15	0.13	0.11	0.10	0.07	0.05
GE	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.59	1.57
GMV-130-30 Portfolio: DCC-NL									
AV	7.98	8.29	8.40	8.45	8.45	8.43	8.47	8.37	8.28
SD	6.86	6.70	6.61	6.54	6.52	6.52	6.56	6.93	7.43
SR	1.16	1.24	1.27	1.29	1.30	1.29	1.29	1.21	1.12
TO	1.63	1.26	1.05	0.91	0.82	0.70	0.61	0.39	0.27
GE	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.59

Table B.2: Annualized performance measures (in percent) for various estimators of the GMV-130-30 portfolio. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 8,568 daily out-of-sample returns net of transaction costs and in excess of the risk-free rate. Furthermore, TO stands for average monthly turnover and GE stands for average gross exposure.

λ	0	2.5	5.0	7.5	10	15	20	50	100
Marko Portfolio: NL									
AV	9.82	10.37	10.47	10.51	10.57	10.64	10.72	11	10.62
SD	9.69	9.74	9.85	9.95	10.07	10.31	10.55	11.74	13.14
SR	1.01	1.06	1.06	1.06	1.05	1.03	1.02	0.94	0.81
TO	1.44	0.81	0.61	0.51	0.45	0.38	0.34	0.25	0.21
GE	4.47	4.42	4.39	4.38	4.36	4.35	4.34	4.44	4.73
Marko Portfolio: DCC-NL									
AV	9.39	9.79	10.05	10.03	10.06	10.14	10.31	10.92	11.11
SD	7.98	7.74	7.65	7.66	7.70	7.91	8.14	9.02	10.18
SR	1.18	1.27	1.31	1.31	1.31	1.28	1.27	1.21	1.09
TO	2.80	1.76	1.37	1.17	1.04	0.88	0.79	0.56	0.45
GE	3.40	3.44	3.57	3.70	3.81	4.00	4.15	4.70	5.17

Table B.3: Annualized performance measures (in percent) for various estimators of the Marko portfolio. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 8,568 daily out-of-sample returns net of transaction costs and in excess of the risk-free rate. Furthermore, TO stands for average monthly turnover and GE stands for average gross exposure.

λ	0	2.5	5.0	7.5	10	15	20	50	100
Marko-130-30 Portfolio: NL									
AV	9.79	9.88	9.99	10.09	10.14	10.28	10.35	11.11	11.55
SD	11.57	11.59	11.63	11.67	11.73	11.83	11.94	12.56	13.48
SR	0.85	0.85	0.86	0.86	0.87	0.87	0.87	0.88	0.86
TO	0.80	0.68	0.60	0.54	0.49	0.43	0.39	0.29	0.23
GE	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.59	1.57
Marko-130-30 Portfolio: DCC-NL									
AV	9.26	9.33	9.45	9.46	9.47	9.49	9.53	9.59	10.31
SD	8.39	8.38	8.38	8.40	8.41	8.46	8.51	9.34	10.28
SR	1.10	1.11	1.13	1.13	1.13	1.12	1.12	1.03	1.00
TO	1.67	1.45	1.30	1.19	1.10	0.98	0.90	0.65	0.51
GE	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.59	1.58

Table B.4: Annualized performance measures (in percent) for various estimators of the Marko-130-30 portfolio. AV stands for average; SD stands for standard deviation; and SR stands for Sharpe ratio. All measures are based on 8,568 daily out-of-sample returns net of transaction costs and in excess of the risk-free rate. TO stands for average turnover and GE stands for average gross exposure.