Prof. Dr. Martin Biewen

Department of Economics, University of Tübingen

Finding the minimum variance portfolio using LASSO regularization revisited using developing countries' data

Seminar "Master seminar in Econometrics"

Summer Term 2022

Stefan Fauth

Fichtenweg 3 / 1208

72074 Tuebingen

stefan.fauth@student.uni-tuebingen.de

Student number: 5706097

Field of study: Quantitative Data Science

Date: 08.07.2022

Contents

1 Introduction **1**

2 Main part **3**

2.1 Theoretical framework 3

2.2 Data Selection 6

2.3 Experimental setup 9

2.4 Results 11

2.5 Discussion 13

3 Conclusion **16**

4 Appendix **18**

List of Figures

1 Visualization of the performance metrics presented in table 2. Each data points illustrates the 10 % value at a risk, the Sharpe ratio, the shorting fraction and the fraction of active assets for the corresponding regularization strength. 12

List of Tables

1 Summary statistics of the final dataset 9

2 Test set performances of Asian equity index portfolios 12

1 Introduction

Recently, in several countries, the two main economic actors, central banks, as well as governments, significantly changed their economic behavior, or announced their intention, with great repercussions on the world-wide economic situation. With central banks, slowly, going back to a more restrictive monetary policy (FED (2022a), FED (2022b)), mainly forced by inflation, policy-makers tend to restrict the potential for further economic growth by imposing economic sanctions, as well as increasing protectionism (e.g. the EU-Russia sanctions). Combined with the absence of long-term economic growth (real GDP growth), over-aged societies, and low birth-rates, lately, investors start to acknowledge the problematic economic future of the developed countries. Added to that, affecting all economies over the world, more frequent natural disasters, the COVID pandemic and military conflicts pose a big challenge for investors, regardless of origin, suggesting that the world economy is not, merely, in a short term recession. This potential absence of positive real GDP growth rates is manifested by financial markets, as they display investors' expectations of the future. In particular, this can be noted by observing the great volatility in financial markets. Altogether, due to this increase in uncertainty, one could conclude that the success of a long and hold strategy of developed countries' assets, like in the last decades has, very likely, come to an end. In particular, due to excessive, expansive monetary policies in the past and the caused increase of money across monetary systems, one could also come to the conclusion that current asset prices might be, artificially, overvalued. The aforementioned theoretical arguments of this breakdown have been inspired by Homm (2020).

For investors, who want to invest in stocks or products that are composed of stocks, e.g. index funds, we want to outline two solutions which are opened up by this, which do not exclude each other. One alternative might be to stay inside the developed markets, but to adapt to the economic situation by high-frequent, time-consuming and high risk trading, including short-selling.

Compared to the developed countries, the emerging markets, especially some Asian and African countries have higher birth rates and higher real economic growth rates (Roser (2014), United Nations and Social Affairs (2019) and The World Bank (2022a)). Indicated by low GDP per capita levels, average living standards are, in some countries, still on a level, where a majority of the populations desire a change of their current situation (The World Bank, 2022b). Thus, future economic growth can be regarded as an inevitable outcome. Acknowledging the potential presence of a lack of rule of law and corruption, for investors, this, still, can be interpreted as a low-risk alternative to developed countries' markets, as they could profit from a natural increase in basic good's demand, such as drinking water.

Yet, when following the purpose of building an optimal portfolio that consists of emerging markets' assets, it is, first, necessary to investigate the validity of current research to emerging market's data. Due to the validity of current research for, mainly US data, this paper adapts the LASSO penalty in the context of portfolio optimization and applies it onto the data of interest.

The remaining parts of this paper are organized in the following way: after introducing the reader to the underlying nature of the basic, unpenalized optimization problem and its associated issues, several regularization terms will be presented that seek to solve these. Among these, we implemented the, widely acknowledged, LASSO and compared it, using, commonly known, performance metrics (Sharpe ratio, Value at risk), to an unregularized portfolio. As an additional section, the appendix, also, features a didactic step by step breakdown of a concrete implementation of cross-validation in the context of portfolio selection.

2 Main part

2.1 Theoretical framework

Also in the emerging markets, there is a large amount of assets and asset classes available, which require a flexible investment strategy. The components of a portfolio require following theory-based principles, i.e. choosing portfolio weights that minimize risk, while delivering the highest expected return. The most prominent theoretical framework that formalized this principle was provided by Markowitz (1952). Mathematically, an investor's decision problem was formalized by describing the following optimization problem:

$$\min_{\boldsymbol{w}^*} \quad \boldsymbol{w}^* \boldsymbol{\Sigma} \boldsymbol{w}^* \tag{1}$$

$$\text{subject to} \quad \max_{\boldsymbol{w}^*} \boldsymbol{w}^* \boldsymbol{\mu} \tag{2}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{3}$$

The investor finds the optimal weight vector $\boldsymbol{w}^* = \begin{bmatrix} w_1 & w_2 & \cdots & w_N \end{bmatrix}$, consisting of one component for each asset that minimizes the covariance $\boldsymbol{w}^* \boldsymbol{\Sigma} \boldsymbol{w}^*$, while maximizing the expected return $\boldsymbol{w}^* \boldsymbol{\mu}$, under the constraint that the individual weights of each asset i, w_i, sum up to 1, corresponding to the constraint that the whole budget has to be invested.

In practice, the parameters that are utilized are replaced by their estimates, for example, by the sample mean of past returns (equation 4) and the sample covariance matrix (equation 5), yielding:

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^{T} r_{1t} & \frac{1}{T} \sum_{t=1}^{T} r_{2t} & \cdots & \frac{1}{T} \sum_{t=1}^{T} r_{it} \end{bmatrix} \tag{4}$$

$$\hat{\boldsymbol{\Sigma}_{ij}} = \frac{1}{T-1} \sum_{t=1}^{T} (r_i - \overline{r_i})(r_j - \overline{r_j}) \tag{5}$$

However, based on the mathematical implications of the optimization problem, previous research has shown that the procedure illustrated in equation 1 to 5 yields naive, optimal weights (Fastrich et al., 2015). Due to the nature of the problem, the optimizer is greedily overly confident in the implemented sample estimators, allocating the whole budget to the "best" few assets that, allegedly, maximize the ratio between chance and risk. From an investor's point of view, one might think that a maximally sparse solution is an advantageous solution, as transaction costs are, relatively, lower, if only a small amount of assets is part of the portfolio. Yet, it can also be considered problematic due to the following reasons: on the one hand, this means that the portfolio only includes few assets, indicating higher risk, even if they feature minimal correlation, which is not desirable, as the portfolio is then very susceptible to exogenous shocks, such as financial statement fraud (e.g. the German payment service provider Wirecard). On the other hand, the statistical evidence, i.e. the estimators, on which the construction portfolio was based on, have to be even more reliable, as risk is not as well diversified.

However, past research has proven that not to be the case, in particular, for the expected mean return: e.g. using linear regression, Jorion (1985) has shown that the sample means of past data can not be considered, economically, useful to predict future returns. Added to that, Best and Grauer (1991) also concluded that only marginal changes in the estimated mean return can lead to an entirely different set of weights. The negative consequence of this estimation error was, then, observable in an unsatisfying out-of-sample performance of portfolios.

Consequently, researchers tried to implement different solutions to overcome this issue. The following overview is based on our main reference Fastrich et al. (2015).

With regard to the insufficient precision of the mean estimator, certain parts of the literature, focused on, merely, finding the minimum-variance portfolio (MVP), i.e. neglecting the mean and simplifying the optimization problem, which led to advantageous out-of-sample performances (e.g. DeMiguel and Nogales (2009)). Yet, this still posed certain problems: the tendency to end up weighing a few assets extremely, due to the model's confidence in an naive estimator, did not vanish.

In order to overcome this pitfall, authors began implementing regularization methods to shrink the asset allocation weights, which were proven to feature a higher out-of-sample Sharpe-ratio (e.g. Brodie et al. (2009)):

$$SR = \frac{\bar{r}}{s}, \tag{6}$$

in which is the portfolio"s return $\bar{r}$ per standard deviation s, i.e. risk.

Mathematically, this inclusion can be described by including a term to the objective function, which is a function f of the weight vector w^*. Disregarding the mean further simplifies the optimization problem, in which λ describes the regularization strength:

$$\min_{\boldsymbol{w}^*} \quad \boldsymbol{w}^{*\boldsymbol{T}}\boldsymbol{\Sigma}\boldsymbol{w}^* + \lambda \sum_{i=1}^{N} f(w_i) \tag{7}$$

$$\text{subject to} \quad \sum_{i=1}^{N} w_i = 1. \tag{8}$$

Now, the optimization also has to take the second term, in equation 7, into account, which, proportionally, increases the loss, the higher the weights.

First researchers applying this idea of regularization then implemented the least absolute shrinkage and selection operator (LASSO), in which the function f is, merely, utilized to take the absolute value of its argument (Tibshirani (1996), Brodie et al. (2009)):

$$\lambda \sum_{i=1}^{N} f(w_i) = \lambda(|w_1| + |w_2| + ... + |w_N|) \tag{9}$$

Except for the penalization, this penalty also is constructed in a way that it can, also, shrink individual weights to zero. In economic terms, this means that certain assets can be, entirely, removed from the portfolio, which can be interpreted as a way of, automatically, selecting the portfolio's components, which is, again, desirable to some extent, since it corresponds to lower transaction costs.

Zou (2006) further developed this idea by moving away from using the same regularization strength λ for all assets. His suggestion was to, individually, reweigh every weight by another parameter p_i (w8Las):

$$\lambda \sum_{i=1}^{N} f(w_i, p_i) = \lambda(|w_1|\, p_1 + |w_2|\, p_2 + ... + |w_N|\, p_N)$$

(10)

Intuitively, this can be interpreted as quantifying an investor's individual opinion / confidence on / in the quality of an asset's estimated moments. If an investor concludes that an asset's sample data is a good proxy to learn about the stochastic process that generated the asset's time series at hand, then the individual weighting p_i should be, relatively, lower than the generic LASSO, which implicitly assumes the same individual weighting p_i for every asset. Mathematically, like the LASSO, one has to note that the optimization problem remains convex. In addition to the LASSO, Fastrich et al. (2015) presented other penalties that caused the optimization function to be non-convex. The content of this contribution will be a mere focus on the plain LASSO regularization method, as its validity has, to our knowledge, not been confirmed for South and South-East Asian data.

2.2 Data Selection

As a data source, this paper used Reuter's Refinitiv Eikon Datastream platform, which provides a user-friendly Microsoft Excel Add-in, which has been utilized to extract this analysis' data.

We limited the amount of asset classes to be considered in the analysis, excluding physical real estate. First of all, tracking prices of individual buildings is infeasible and might be misleading, as a lot of actors on the market might be outliers, e.g. a buyer that does not not pay the optimal market rate. Secondly, this asset class is not available for every investor, since investing in a far-distant country might be only possible with very high transaction cost, e.g. caused by language and cultural barriers. Added to that, an investment in real estate is, relatively, more expensive than in other asset classes, making it not accessible for certain types of investors. Besides, we concluded to disregard physical raw metals. Since silver is also an industrial raw material, due to differences in local supply and demand, its price can vary from country to country, making comparability harder. Distortions in the local market might also cause prices that are too high, as there is no global market that is, easily, accessible, which calls for digital products, whose prices can be, without great effort, compared, using the total expense ratio (TER) or the tracking error (TE). The aforementioned assets also do not feature a high diversification potential for investors that face a high budget restriction. Even for investors that want to invest a higher amount, acquiring more assets features high costs.

Taking these arguments into account, we decided to limit the analysis to financial products that are based on equity indices, as they can be considered from the investor's direct perspective, especially from a private investor's perspective, advantageous, because of higher accessibility, diversification with, relatively, low transaction costs. Statistically, they also do not suffer from missing value problems, as individual shares do. An index can be computed since the first day, it has been market-listed, since if a company drops out, another just replaces it, e.g. in June 2022, Beiersdorf joined the German stock index (DAX), after Delivery hero dropped out. Added to that, on average, it also allows for using more data to perform the optimization. One company might have just been founded a few years ago, which leaves investors with a lot of uncertainty about the stochastic process' properties that generated the stock price's time series that can be observed (e.g. Tesla). In contrast, an index, such as the S&P500, has been existing since 1957, which provides analysts with more observations.

In order to maximize the amount of data and flexibility, we computed daily returns of all assets that were left-over that are then utilized to estimate MVPs. More specifically, we chose to compute discrete daily returns:

$$r_{t-1,t,i} = \frac{p_{t,i} - p_{t-1,i}}{p_{t-1,i}}, \tag{11}$$

in which $p_{t,i}$ represents a price measure of asset i in period t. As we are using daily returns, In contrast to continuous returns, the portfolio's total return can be just calculated as a weighted sum of each asset's individual return (Schmid and Trede, 2006).

Starting with all equity indices from Bangladesh, India, Pakistan, Thailand, Vietnam, certain international Asian indices and Sub-Saharan Africa, the initial dataset included 1219 assets.[1] The central criterion to include an asset in the list of potential investments was the length of its price index' time series, which, unfortunately, excluded all Sub-Saharan African assets.

Consequently, the dataset, to estimate the portfolios, ultimately, consisted of assets from Bangladesh, India, Pakistan, India, Thailand, Vietnam and international South-East Asian indices.

Again, following the criterion that the time series have a sufficient enough length to produce, somewhat, reliable estimates, the pre-selected assets contained dates ranging from 1988-12-30 until 2022-06-07, which created, by definition, many missing values for all assets that are shorter.

There are different strategies to handle missing values, such as imputation and, brutally, dropping out a column that contains a fraction that surpasses a defined threshold. In this

[1] Due to the political risk present in the People' Republic of China, we decided to exclude it from the analysis.

case, we decided not to proceed with imputation, as the quality of the estimator is, as described above, the major problem of Markowitz portfolios. Decreasing the data quality would, further decrease the quality of the estimator, which is counterproductive.

To sum up, when creating the dataset, we faced the tradeoff between the number of assets and the number of observations. The more assets we wanted to consider, to increase the portfolio's variety to include more unique assets, the less information we can include about each component's stochastic process. Intuitively speaking, more assets reduce the credibility of the estimator. We ended up using, relatively, more assets and less observations. Fastrich et al. (2015) justified this decision with the presence of, arguably, outdated observations, given the time series are too long. To ensure a comparability between their and our experiments, the length was chosen to be close to theirs, namely roughly, 1400 observations. After computing daily returns, the final dataset featured the following summary statistics:

Table 1: Summary statistics of the final dataset

T	Date range	N	$\bar{r}$	$\bar{s}$	$\bar{S}$	$\bar{KU}$
1407	2017-01-13 - 2022-06-06	1179	0.00037	0.019	-0.099	20.44

Table 1 displays, in which data range the N assets with a sample size of T lie. It also displays the average return $\bar{r}$, the average standard deviation $\bar{s}$, the average skewness $\bar{S}$ and the average kurtosis $\bar{KU}$.

The return distributions are, negatively, skewed and feature fat tails (leptopkurtic). High positive returns are less frequent than high negative returns. One can also come to the conclusion of the presence of fat tails, when considering the average standard deviation, which is, rather, high, undermining a high fluctuation in daily returns.

2.3 Experimental setup

Fastrich et al. (2015) use a variable time window in order to test the out-of-sample performance. In order to determine the optimal portfolio weights, 250 days, which corresponds to one trading year are used. Considering the availability of 1407 time points, this would have left us with 55 portfolios to be created and evaluated. We did not end up following their approach to tune the regularization strength λ and to validate the model. Nevertheless, we provide interested readers with an intuitive step by step breakdown of an improved, alternative cross-validation procedure in the appendix, which is based on our main reference.

The main reason to not include the aforementioned validation / tuning procedure was that fitting a high amount of models is, computationally, expensive, if the number of assets is high, as in our case. Therefore, being aware of the superiority of nested cross-validation, due to the presence of limited computational capacities, we ended up shuffling the data and splitting the whole data set into training (80%) and test data (20%) to obtain an impression of the out of sample performance. We then found the optimal weighting vector for the training data using trust-region constrained optimization (Conn et al., 2000). Subsequently, the evaluation of the optimal solution was conducted, utilizing the test set.

As a measure of success, standard benchmarks, such as the Sharpe-ratio and the Value-at-Risk (VaR) are computed for the test days. The VaR is, intuitively speaking, the value for which we can say that in $p\%$ of the cases, the return is even lower than it. Statistically, this is estimated by the $p\%$ quantile of an asset's inverse return distribution:

$$VaR_p = F_r^{-1}(p). \tag{12}$$

They, also, end up computing the fraction of active positions, i.e. the fraction of assets that get assigned a non-negative weight, indicating they are part of the portfolio, and the fraction of short positions, i.e the fraction of assets that got assigned a negative weight to monitor, whether there is a, questionably high, amount of shorting observable.

Given the aforementioned metrics, we are, now, able to compare different Markowitz portfolios that follow the functional form, which was introduced in equation 7. This functional form can also be used to construct a basic portfolio, which results from setting lambda to zero, which causes the penalty term to drop out of equation 7, yielding an unregularized solution.

Summarizing the theory, which was explained, in detail, above, the optimal lambda should not be too high, as this would lead to a too high penalty, which would lead to a very complex portfolio with a large amount of assets. Yet, if the regularization strength is too low, the model is likely to be too confident in the estimator, i.e. it includes few assets only, leading to a poor test performance. In order to find the optimal regularization strength, by tackling the aforementioned trade-off, hyperparameter turning of all kinds could be applied. The criterion to compare different models that hold different regularization strengths will be done, in the optimization, using the Sharpe ratio. When it comes to finding the optimal regularization strength, λ^*, the Sharpe ratio can be considered advantageous, since it, also, takes into account the return and not only the risk. In contrast, the VaR is just an expression of risk, while the fraction of active positions or short positions rather is a feature and not a quantitative performance measure.

2.4 Results

Due to the aforementioned computational limitations and time constraints, we ended up using a linearly spaced parameter grid that contained ten different values from 0 to 1. Table 2 and figure 1 provide a summary of the results.

Table 2: Test set performances of Asian equity index portfolios

Regularization strength λ	$VaR_{10\%}$	$\bar{SR}$	active fraction	shorting fraction
0	-0.001061	0.058997	1.0	0.256997
0.1	-0.002287	0.094352	1.0	0.050891
0.2	-0.004267	0.061815	1.0	0.016964
0.3	-0.004602	0.060380	1.0	0.003393
0.4	-0.004931	0.059828	1.0	0.003393
0.5	-0.005149	0.059453	1.0	0.002545
0.6	-0.005307	0.059253	1.0	0.002545
0.7	-0.005418	0.059035	1.0	0.003393
0.8	-0.005499	0.058909	1.0	0.003393
0.9	-0.005572	0.058806	1.0	0.003393
1	-0.005573	0.058744	1.0	0.003393

For each penalty, 80% of the data were utilized to find the optimal weights. They were then evaluated on a test set, yielding a 10 % value at a risk, the Sharpe ratio, the shorting fraction and the fraction of active assets.

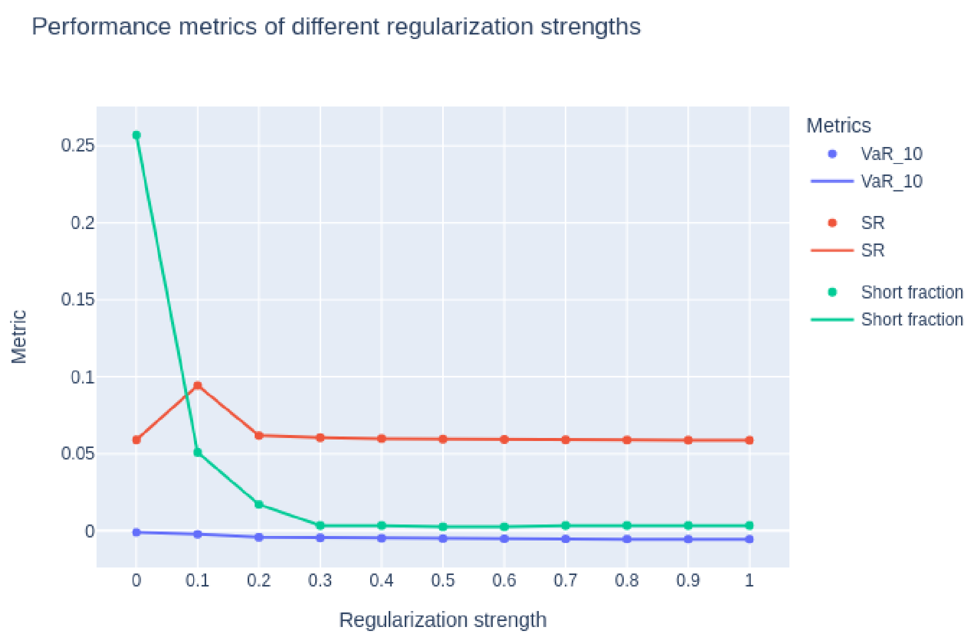

Figure 1: Visualization of the performance metrics presented in table 2. Each data points illustrates the 10 % value at a risk, the Sharpe ratio, the shorting fraction and the fraction of active assets for the corresponding regularization strength.

Among our ten fitted portfolios, the minimum regularization strength of $\lambda = 0.1$ yielded, by far, the highest Sharpe ratio: for this portfolio, the out-of-sample relationship between chance and risk was 0.094. The unregularized portfolio had, surprisingly, the lowest 10 % value at risk: in 10 % of the cases, the daily return is lower than 0.23 %. For all regularized portfolios, it can be stated that the higher the regularization strength, the lower the Sharpe ratio and the higher the 10 % value at risk.

By comparing the metrics of the unregularized portfolio and the best regularized portfolio, we can conclude that LASSO regularization, also for developing countries' data, leads to a superior solution.

For the regularized portfolios, there, also, seems to exist a positive relationship between the fraction of shorting and the performance metrics. The higher the shorting fraction, the higher the observed Sharpe ratio and the lower the value at risk, indicating a better ratio between chance and risk. It can be stated that if the regularization strength is higher than $\lambda = 0.2$ the amount of shorting, distinctly, decreased, corresponding to a lower performance. Apparently, the regularization did not lead to a sparser portfolio, as the fraction of active assets is 100 % in all models, indicating that, for every penalization strength, all assets were part of the portfolio.

2.5 Discussion

As the lowest regularization strength of $\lambda = 0.1$ clearly, strictly, dominated all other models, future research that contains a similar type of data should feature similar lambdas of rather lower magnitude, than higher magnitude. This, clearly, implies that a certain amount of shorting is, statistically, desirable. As described, in detail, in the introduction, from a macroeconomic perspective as well, long-only portfolios, feature an inferior return-risk ratio. Our best model was able to score similar Sharpe ratios as the best LASSO models in our main reference for a similar amount of assets, which makes developing countries' assets, similarly, attractive for investors. However, due to the societal arguments presented above, they can, even, be considered superior.

Besides, the presence of all assets in all portfolios is, definitely, questionable. The LASSO's advantageous feature over other penalties, also, lies in its capability to shrink weights to zero, corresponding to discarding certain assets completely. In contrast to our analysis, in our main reference, in all regularized portfolios, at least, 45 % of all assets were discarded. Except for its questionability, it is not a desirable result, as well. As the number of assets was $N = 1179$, practically, following the goal of possessing an optimal portfolio, investors would be advised to acquire every single of these assets, requiring their availability, causing infeasibly high transaction costs.

With the exception of potential errors in the optimization procedure or, simply, by the choice of a wrong optimizer that might have caused the results to be biased, by taking into account, previous work in the field, our analysis, still, has the potential to be further improved, in statistical, as well as in economic terms that should be addressed by future research: as we, merely, chose to implement the sample estimator for the covariance matrix, the performance of the approach using a different estimator remains unclear. For example, alternatively, Fastrich et al. (2015) used a three-factor model to estimate the covariance matrix, which yielded different results. Their analysis does not allow for an unambiguous comparison of both estimators, as they used different datasets to compute performance metrics. Nevertheless, the existence of different results hints at the potential existence of a way to further improve the approach, particularly, because estimation errors in the model's input parameter, i.e. the covariance matrix, is considered to be the main problem.

Furthermore, one could also question the chosen, aforementioned tradeoff between the sample size and the number of assets to choose from. Our main reference concludes that Markowitz portfolios, ceteris paribus, perform differently, when the number of available assets differs. Even though the LASSO is supposed to cause automatic asset selection, it failed to do so in our case, meaning that if one dataset version is just a truncated version of another version, e.g. SP200 vs. SP500, it should be, theoretically, able to produce the same results for both datasets, if the optimal solution features assets that are available in both datasets.

Considering this, since this contribution, merely, included the LASSO, future research should be concerned with, also, covering the implementation of other penalties, to check if they are able to produce sparser portfolios for developing countries' data. Regarding this idea, Fastrich et al. (2015) provide an overview over a wide range of alternative penalties. They can be categorized into convex penalties, such as the LASSO and the aforementioned weighted LASSO (w8las), and non-convex penalties, such as the Zhang penalty , which assigns a different penalty for different domains of the weights (Gasso et al., 2009). In other words, if a weight exceeds a threshold, it gets assigned a constant value, which could also be tuned. This additional flexibility could be an additional tool to increase the portfolio's performance, as they were able to outperform convex penalties, such as the LASSO, for a similar amount of assets as ours, approximately 1000 assets.

Moreover, from a statistical perspective, the choice of the size and the ratio between the respective training and test tests could make a difference in the outcome. Assigning a higher number of samples to the training set provides more information about the stochastic process that generated the data, which could lead to a better training fit. Except for these concerns, our whole validation procedure should be improved further. Just shuffling the data, training and testing only once, i.e. not conducting any form of validation, such as cross-validation, is insufficient and limits the internal and external validity of this paper's results. Acknowledging the growing availability of engines that feature high computational power, future research should, therefore, contain a routine, as illustrated in the appendix, which is concerned with tuning λ and validating the performance on every sample. Arbitrary choices of, for example, one trading year as the training size is, definitely, questionable and should be examined in future research. Of course, one can make a similar argument for the length of the test set.

Similarly, with regard to the t-distributed nature of daily returns, being centered around zero, higher returns are very unlikely. From an economic point of view, working with such a small time window also features high transaction costs that may not be attractive to investors, while holding assets for a longer time reduce these, in total. Changing the variable of interest to monthly returns could be an alternative that still features a certain amount of data that can illustrate, somewhat, reliable information about the stochastic process that generated the data, while acknowledging the aforementioned advantages of a longer time period.

3 Conclusion

Leaving aside validation concerns of our methodology, we were able to show the applicability of LASSO regularization in a minimum variance portfolio context for certain developing countries' data. After, theoretically, motivating the transition of developed countries' assets to developing countries, by outlining the effects of current political, social and economic developments, we, theoretically motivated and empirically, proved the potential of regularization methods to improve, also, a portfolio's performance that consists of, less developed, Asian countries' equity indices: our best model was able to outperform an unregularized, naive minimum variance portfolio. A regularization strength of $\lambda = 0.1$ yielded the highest Sharpe ratio, indicating the best ratio between chance and risk.

The high differences in the results for different regularization strengths highlight the importance of the choice of the hyperparameter. Learning the optimal penalization strength can be achieved by hyperparameter tuning approaches, such as cross-validation, also in a portfolio optimization context. In the appendix, we introduced an intuitive implementation that, successfully, can be considered a valid tradeoff between model validation and required computational power. For all regularized portfolios, the amount of shorting positively correlated with their performance, indicating that a certain amount of shorting is, empirically, superior to long only portfolios.

4 Appendix

Fastrich et al. (2015) implemented cross-validation to find the optimal lambda, while it remains unclear, whether they implemented nested cross-validation. One has to note that nested cross-validation would be, from a statistical point of view, be a better alternative to determine the validity of an estimator' optimal hyperparameters, as the parameters are fit and evaluated on all samples. Please note that nested cross-validation would be associated with fitting even more models, making the whole procedure more successful, but also more complicated, from a conceptual, as well as from a computational point of view. The following can be, thus, interpreted as a less complex alternative that still incorporates a certain degree of tuning.

The reader will be, now, provided with a concrete implementation of a cross-validation procedure that incorporates the aforementioned theoretical ideas. The purpose of this breakdown is not to present a, computationally highly efficient, solution, but rather to provide some intuition to readers not having a background in the methods of machine learning, in particular in validation.

After merging different asset's time series, discrete, daily returns can be computed for asset i, according to equation 11, yielding the daily return matrix R, which then can serve as an input for algorithm 1 to find the optimal lambda, λ^*:

In our case, the moving time window procedure corresponds to creating 55 different portfolios. For each step, first, the current window's training and test data are defined (line 1 and 2). After that, the optimal weight vector is found for every lambda, using the training set. After obtaining the optimal set, the performance will be evaluated on the test by computing the return distribution (line 6), which then can be analyzed by the Sharpe ratio (line 7). When we evaluated each λ on every window, we can compute the average Sharpe ratio for every lambda to find the optimal λ.

Algorithm 1 Cross-validation to find λ^*

Require: R = shuffled daily returns, $\vec{\lambda}$ = grid of lambdas to evaluate

1: **for** window in windows **do**
2: training data = R[21 * window : 21 * window + 250] {Get 250 training samples}
3: test data = R[21 * window + 250 + 1 : 21 * window + 250 + 21 + 1] {Follow up with the next 21 samples to evaluate}
4: **for** λ in $\vec{\lambda}$ **do**
5: w^* = optimizer(training data, λ) {Find optimal weights}
6: $\vec{r}$ = test data * w^* {Compute test returns}
7: SR$\{\lambda_{window}\}$ = $\frac{\bar{\vec{r}}}{\sqrt{Var(\vec{r})}}$ {Compute test Sharpe ratio and store it}
8: **end for**
9: **end for**
10: $\lambda^* = \arg\max_{\lambda} \bar{SR}$ {Findλ that had highest average Sharpe ratio}
11: **return** λ^*

Subsequently, this optimal λ can serve as an input for algorithm 2. We then use the same first loop, as in algorithm 1, to fit different portfolios and to evaluate them (line 5 to 9). As a final performance summary of our approach, we calculate the average performance metrics. These can then be compared to any other portfolio.

Algorithm 2 Obtain the performance metrics for every time window

Require: R = newly shuffled daily returns, λ^* = optimal lambda, pth % VaR

1: **for** window in windows **do**
2: training data = R[21 * window : 21 * window + 250] {Get 250 training samples}
3: test data = R[21 * window + 250 + 1 : 21 * window + 250 + 21 + 1] {Follow up with the next 21 samples to evaluate}
4: w^* = optimizer(training data, λ^*) {Find optimal weights}
5: $\vec{r}$ = test data * w^* {Compute test returns}
6: $SR_{window} = \frac{\bar{\vec{r}}}{\sqrt{Var(\vec{r})}}$ {Compute test Sharpe ratio and store it}
7: VaR$_p$ =F$_r^{-1}(p)$ {Compute test Sharpe ratio and store it}
8: $\%$ active =$\frac{active}{N}$ {Calculate active fraction of all N assets}
9: $\%$ short =$\frac{short}{N}$ {Calculate shorting fraction of all N assets}
10: **end for**
11: **return** $\bar{SR}_{window}$, $V\bar{a}R_p$, $\bar{\%}$ active, $\bar{\%}$ short

References

Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *The review of financial studies*, 4(2):315–342.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2009). Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106(30):12267–12272.

Conn, A. R., Gould, N. I., and Toint, P. L. (2000). *Trust region methods*. SIAM.

DeMiguel, V. and Nogales, F. J. (2009). Portfolio selection with robust estimation. *Operations Research*, 57(3):560–577.

Fastrich, B., Paterlini, S., and Winker, P. (2015). Constructing optimal sparse portfolios using regularization methods. *Computational Management Science*, 12(3):417–434.

FED (2022a). Consumer price index: Total all items for the united states [cpaltt01usm657n]. *Federal Reserve Bank of St. Louis, retrieved from FRED*.

FED (2022b). Monetary policy report submitted to the congress on june 17, 2022, pursuant to section 2b of the federal reserve act. *Board of Governors of the Federal Reserve System*.

Gasso, G., Rakotomamonjy, A., and Canu, S. (2009). Recovering sparse signals with a certain family of nonconvex penalties and dc programming. *IEEE Transactions on Signal Processing*, 57(12):4686–4698.

Homm, F. (2022, 2021, 2020). Videos on financial markets and political developments. *YouTube*. https://www.youtube.com/c/FlorianHommOfficial/videos.

Jorion, P. (1985). International portfolio diversification with estimation risk. *Journal of Business*, pages 259–278.

Markowitz, H. (1952). Portfolio selection*. *The Journal of Finance*, 7(1):77–91.

Roser, M. (2014). Fertility rate. *Our World in Data*. https://ourworldindata.org/fertility-rate.

Schmid, F. and Trede, M. M. (2006). *Finanzmarktstatistik*. Springer-Verlag.

The World Bank (2022a). *Global Economic Prospects, January 2022*. The World Bank.

The World Bank (2022b). World development indicators database. https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

United Nations, D. o. E. and Social Affairs, P. D. (2019). World population prospects: The 2019 revision (medium variant).

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.