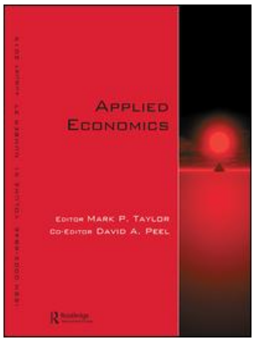

Applied Economics

ISSN: 0003-6846 (Print) 1466-4283 (Online) Journal homepage: https://www.tandfonline.com/loi/raec20

CVaR-LASSO Enhanced Index Replication (CLEIR): outperforming by minimizing downside risk

Brian Gendreau, Yong Jin, Mahendrarajah Nimalendran & Xiaolong Zhong

APPLIED ECONOMICS
https://doi.org/10.1080/00036846.2019.1616072

CVaR-LASSO Enhanced Index Replication (CLEIR): outperforming by minimizing downside risk

Brian Gendreau[a], Yong Jin[b], Mahendrarajah Nimalendran[a] and Xiaolong Zhong[c]

[a]Department of Finance, Insurance & Real Estate, Warrington College of Business Administration, University of Florida, Gainesville, FL, USA; [b]School of Accounting and Finance, The Hong Kong Polytechnic University, Hung Hom, Hong Kong; [c]Amazon

ABSTRACT

Index-funds are one of the most popular investment vehicles among investors, with total assets indexed to the S&P500 exceeding \$8.7 trillion at-the-end of 2016. Recently, enhanced-index-funds, which seek to outperform an index while maintaining a similar risk-profile, have grown in popularity. We propose an enhanced-index-tracking method that uses the linear absolute shrinkage selection operator (LASSO) method to minimize the Conditional Value-at-Risk (CVaR) of the tracking error. This minimizes the large downside tracking-error while keeping the upside. Using historical and simulated data, our CLEIR method outperformed the benchmark with a tracking error of $\sim 1\%$. The effect is more pronounced when the number of the constituents is large. Using 50–80 large stocks in the S&P 500 index, our method closely tracked the benchmark with an alpha 2.55%.

KEYWORDS

Stochastic programming; conditional value-at-risk; LASSO; enhanced indexation

JEL CLASSIFICATION

G11; D81; C63

I. Introduction

Since the introduction of Index tracking funds by Vanguard in 1976, the industry has grown dramatically, accounting for 19.3% of all equity investments at the end of 2016 (Investment Company Institute (US) 2017). Indexing or benchmarking investments is popular because it provides investors with simple passive investment vehicles that have low transactions costs. This has led to a huge demand for the creation and sale of index tracking funds (Arnott, Hsu, and Moore 2005).

Modern portfolio theory recommends that investors hold capitalization-weighted portfolios (the market portfolio) of risky assets, which according to the CAPM model, is mean-variance efficient. However, many papers, both in academia and industry, have shown that capitalization-weighted indices such as the *S&P*500 and Russell 1000 do not lie on the efficient frontier. At the same time, it is, almost impossible to find an index that unambiguously lies on the efficient frontier *ex ante*. The belief among investors nonetheless seems to be that the capitalization-weighted indices are almost efficient, which may explain why investing in such indices is still quite popular. Investors find Exchange-Traded Index Funds (ETFs) and index tracking mutual funds attractive because of their low transaction costs and the perception that they track their underlying indices closely. These expectations make index replication a critical task for the institutions that create these Index Funds and ETFs (Blume and Edelen 2002).

Currently, two main replication strategies exist in the market: full or exact replication, and sampling or partial replication. The full replication strategy is straightforward but difficult to implement. A full replication of the *S&P*500, for example, would require the fund to hold all 500 constituent stocks at their capitalization weights at all times. Market frictions including illiquidity, monitoring costs, redemptions and inflows, taxes, and dividends lead to frequent rebalancing that makes it difficult to implement a full replication strategy. The strategy is even more expensive to implement for an index such as the Russell 2000 that has 2000 smaller, less liquid stocks. Because of these concerns most index funds use sampling strategies that choose a subset of stocks in the benchmark to track the index. Further, if the index is not value weighted – for example equally weighted – then it may require frequent

CONTACT Yong Jin jimmy.jin@polyu.edu.hk School of Accounting and Finance, The Hong Kong Polytechnic University

rebalancing. Hence, using a subset of stocks becomes more attractive as it will reduce transactions costs.

The two basic steps involved in index replication strategies are sampling and optimizing. Two widely used sampling strategies are stratified sampling and large market capitalization stock sampling. In the latter case, the index replicator holds only the largest stocks by market capitalization or by market volume, which are liquid and hence will have lower implicit and explicit transactions costs. In this paper, we will explore large market capitalization stock sampling. The optimization involves obtaining optimal weights for the stocks in the sample by minimizing the tracking error volatility (TEV). Tracking error (TE) is the difference between the index return and benchmark return it is supposed to mimic over a specified time interval.

Traditionally, an index replicator's objective has been to minimize TEV. This is considered a 'passive' strategy. Recently, however, enhanced indexation has gained popularity. In this strategy, the fund managers strive to outperform the index while keeping TE low. One way to enhance the index performance is to control or minimize the negative TE (index fund under-performs the benchmark index) rather than positive errors (index fund outperforms than benchmark index). Hence, an index replicator would prefer to minimize the fund's negative tracking error while maximizing the positive tracking error that occurs when the fund outperforms the benchmark. In this paper we replicate a benchmark index (is the value-weight index return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ, or $S\&P500$) by selecting a full set or subset of firms from the set of the whole portfolio universe or largest 100 firms by market capitalization, and determine the portfolio weights that minimize the Conditional Value-at-Risk (CVaR) of the tracking error. CVaR has several nice properties as a metric for risk management: (1) CVaR is a coherent risk measure (Artzner et al. 1999) and CVaR of a portfolio of returns is a convex function of the weights; (2) CVaR minimizes the left tail of the TE distribution and hence allows the index to outperform the benchmark; (3) CVaR optimization can be reduced to convex programming and in some cases to linear programming (see Rockafellar and Uryasev 2000, 2002).

We solve the stock selection and weighting problem via LASSO (Least Absolute Shrinkage and Selection Operator) algorithm proposed by Tibshirani (1996). LASSO minimizes an objective function such as the residual sum of squares or in our case the CVaR, subject to the sum of the absolute value of the coefficients being less than a constant. Because of the nature of this constraint, it can lead to some of the weights being exactly zero. This is an advantage when we want to select a subset of the stocks to include in a portfolio. A limitation of LASSO algorithm is that it cannot restrict the weights to be nonnegative. However, this is not a serious concern given that some funds that use indexation such as mutual funds are allowed to short sales (Chen, Desai, and Krishnamurthy 2013). In this study, we allow the weights in the replicated portfolio to be negative as well as positive. However, we set the LASSO penalty value low ($s = 1.5$) for the l_1 minimization constraint to reduce the amount of short selling. This constraint limits the amount of short sales or negative weights on the stocks in the fund.[1] The LASSO penalty s can be alternatively selected by cross-validation. Besides the LASSO penalty, we also add one more feasibility constraint that the weights sum to one: $\sum_{i=1}^{p} w_i = 1$. We also theoretically show the oracle property of the estimators, which guarantees that our estimators are best sparse linear approximation of the true values. In the online Appendix, we analyse the behaviour of the objective function and provide theoretical evidence why the portfolio obtained via CLEIR outperforms the benchmark.

The remainder of this paper is organized as follows: In section 2 we briefly review the existing literature. Section 3 describes CVaR-LASSO Enhanced Index Replication (CLEIR) model and the formulation of the problem. We use a transformation developed by Rockafellar and Uryasev (2000) that makes it easier to solve the optimization problem. Section 4 presents empirical results and compares the CLEIR method with the additional weighting methods in different

[1]Chen, Desai, and Krishnamurthy (2013) find that the magnitude of short sales is around 15.65% of net assets on average; thus, we choose a penalty level $s = 1.5$, which limits the short sales up to 25%.

portfolio universes from year 1988 to 2018, with the value-weighted index as the benchmark. Section 4 further examines the out of sample properties of CLEIR performance using daily stock return data from 2003 to 2012 to track *S&P* 500 index. We present our conclusion in Section 5.

II. Literature review

The main strategies that have been proposed to implement index tracking are full replication, partial replication, and enhanced replication. Full replication involves holding all the constituent securities in the exact same proportion as in the benchmark portfolio at all times. However, in many instances market frictions may not permit this strategy to be feasible. In these instances, the index replicator can use a partial replication strategy using a subset of the most liquid firms in the index. An enhanced replication strategy, in contrast, seeks to outperform an index while maintaining a similar risk profile. Our methodology falls into the enhanced replication category.

A comprehensive review of traditional index tracking methodology is provided by Blume and Edelen (2002) and Chavez-Bedoya and Birge (2014). In addition to traditional indexations strategies, researchers and practitioners are also interested in alternative indexations (sometimes called smart-beta indexation). The goal in these strategies is to outperform a benchmark index while keeping close to the risk profile of the index. Heuristic-based weighting methodologies and optimization-based weighting methodologies are two main methods used in alternative indexing strategies. Chow et al. (2011) define heuristic-based weighting methodologies as 'ad hoc weighting schemes established on simple and, arguably, sensible rules.' For example, naive diversification, or the '1/N' rule is one kind of heuristic-based weighting methodology, which assigns equal weight for each stock in the sampling pool. DeMiguel, Garlappi, and Uppal (2009) provide empirical evidence that portfolios based on equal weighting perform well relative to portfolios based on mean-variance optimization. Another heuristic-based weighting methodology is risk-cluster equal weighting, which weights the portfolio by equally weighting stocks in a risk cluster (Chow et al. 2011). Fernholz (1995) employs a diversity weighting design that blends the portfolio based on different weighting schemes. For example, the blending could be based on market capitalization and equal weighting. Arnott, Hsu, and Moore (2005) propose an indexation method where the weights are based on the fundamental characteristics (e.g. sales, earnings, or size). Recently Lejeune and Samatli-Pa (2013) address a sampling replication strategy based on stochastic integer problem to construct risk-averse enhanced index funds. Their innovative method takes the derivation of a deterministic equivalent for the risk constraint and then use a block decomposition technique to provide a well scale and fast convergent solution.

In terms of optimization-based weighting indexation, mean-variance optimization is the traditional way to construct an alternative indexation. Chopra and Ziemba (1993) address the importance of the estimation risk in forming the mean-variance optimal portfolio. Michaud (1989), Kan and Zhou (2007) and others emphasize the estimation loss in covariance matrix estimation. In order to reduce the significant estimation risk, Chopra and Ziemba (1993) suggest a fixed mean vector (that is, the minimum variance portfolio) to improve the portfolio performance. Haugen and Baker (1991) and Clarke, De Silva, and Thorley (2006) propose a minimum-variance strategy, and they provide empirical evidence to support that their strategy will improve the cap-weighted indexation return and reduce the volatility. Goto and Xu (2015) argue the estimation of the covariance matrix is still problematic, and they apply the graphical LASSO method to estimate the sparse inverse covariance matrix. In addition, Basak and Shapiro (2001) analyse the portfolio policies which maximize the utility when the investors manage market risk using Value-at-Risk (VaR). Further, Basak, Shapiro, and Tepla (2006) also discuss a framework, which maximizes the utility under the constraint of risk measures, to evaluate relative performance evaluation. Goel, Sharma, and Mehra (2018) implement the mixed Conditional Value-at-Risk to design portfolios for index tracking and enhanced indexing problems. And further Goel, Sharma, and Mehra (2019) develop the methodology for robust optimization of mixed Conditional Value-at-Risk stable tail-adjusted return ratio using copulas, with the applications in portfolio construction. Besides the literature using risk measures, Bonami and Lejeune (2009) discuss an exact solution approach for the portfolio optimization when there are stochastic and integer constraints. Choueifaty and Coignard (2008) and Amenc et al. (2011) develop several indexation strategies which maximize the Sharpe ratio. Chow et al. (2011) provide a comprehensive analysis of most of the passive investing alternative indexations in their survey paper and Cai et al. (2018) empirical examines several popular alternative indexations' performance in the Chinese A-Share market.

III. CVaR-LASSO Enhanced Index Replication(CLEIR)

Objective

The objective in the enhanced replication strategy we propose is to use a small subset of the most liquid stocks to replicate and track a benchmark index such as $S\&P500$. Instead of the traditional objective of minimizing tracking error volatility (TEV), we minimize the CVaR of the TE. The optimization is achieved via the LASSO optimization methodology proposed by Tibshirani (1996).

The objective function and the constraints are given by

$$\min_{w_i} \ \mathrm{CVaR}_\alpha \left(Y_t - \sum_{i=1}^{p} w_i R_{it} \right),$$

subject to,

$$\begin{cases} \sum_{i=1}^{p} |w_i| \le s, \\ \sum_{i=1}^{p} w_i = 1, \end{cases}$$

where Y_t is the rate of return of the index at time t, $CVaR_\alpha$ is the α percent level of the Conditional Value-at-Risk, R_{it} is the rate of return of the ith candidate stock at time t. There are a total of p candidate stocks and $w = (w_1, \ldots, w_p)$ is the weight of ith stock in the final index replicating portfolio. We allow short position in stocks and hence w_i can be negative.

The CVaR objective function is a new way to replicate an index compared to the usual square error loss or other types of overall mean loss functions. CVaR has the attractive property of it minimizing the likelihood the tracking portfolio will underperform the benchmark by a large amount (i.e. it minimizes the extreme tail risk of the tracking error). In contrast, most other loss functions penalize both under-performance and over-performance. Although the traditional loss functions may efficiently control the tracking errors, they also limit the ability to outperform the benchmark. Our objective of partial penalization using CVaR should be more attractive to portfolio managers who would naturally prefer to minimize underperformance while outperforming the benchmark. There are other partial penalization methods such as VaR. However, VaR does not account for properties of the distribution beyond its specified confidence level. This may lead to undesirable outcomes for skewed or discrete distributions. On the other hand, CVaR controls for the overall performance over a range of possibilities below the specified α level in which a loss actually occurs.

The optimization of CVaR can be achieved using several techniques. We propose to use the LASSO method that yields sparse models, which is an advantage given we would like to replicate the benchmark index with a small subset of stocks. This is because LASSO can shrink some weights to zero. Further, with some mild conditions on the objective function, the results are guaranteed to be the best weights when the sample size is large, as is typically the case for financial data. Also, the LASSO constraint can be easily changed into a collection of linear constraints. This makes it a more feasible convex optimization problem to solve. Finally, the simple linear constraints increase the speed of obtaining a solution when sample size is large, but also make the solutions more robust.

In sum, the TE-CVaR objective and LASSO method can be used to create an index that can track a benchmark with fewer stocks and at the same time has the potential to outperform the benchmark by minimizing the large expected negative TE.

Risk measures: VaR and CVaR

In this section, we briefly discuss the definition and properties of VaR (Value-at-Risk) and CVaR (Conditional Value-at-risk), two widely used measures in risk management. Unlike volatility measures such as the variance of returns that measure the variability in both upside (gain) and downside (loss) of an asset, the VaR and CVaR describe the loss associated with an asset or portfolio and hence are more appropriate for risk management. See Sarykalin, Serraino, and Uryasev (2008) for a discussion of the pros and cons of VaR and CVaR in risk management and optimization. Here we provide a brief summary of some major differences between the two measures and their use in optimization of portfolios.

VaR

Let ξ be a loss random variable such as the loss (negative returns) on an index or portfolio of assets. The $VaR_\alpha(\xi)$ at a confidence level $\alpha \in (0,1)$,

$$\mathrm{VaR}_\alpha = \inf\{\zeta | P(\xi \leq \zeta) \geq \alpha\}.$$

The VaR concept was introduced in 1990 by JP Morgan for risk management following the 1987 market crash. VaR is a very popular measure as it is simple and provides one number to describe the potential loss during a set time period with a certain probability. A major criticism of VaR is that it does not address scenarios in which the VaR is exceeded. In addition, VaR has some undesirable mathematical properties. It is not a coherent measure of risk and is not subadditive. For these reasons we consider CVaR instead as the objective function to minimize in replicating a benchmark portfolio.

CVaR

$$\mathrm{CVaR}_\alpha = \frac{1}{1-\alpha} E\xi I_{[\xi \geq \mathrm{VaR}_\alpha]}.$$

CVaR was proposed as a risk measure by Rockafellar and Uryasev (2000) as an alternative to VaR. CVaR, under certain conditions, equals the average of the loss beyond a specified confidence level, and hence its name conditional value at risk. In this sense it is a measure that summarizes all the potential losses below a specified confidence level. It has some nice mathematical properties. It is a coherent measure and also subadditive. More importantly, Rockafellar and Uryasev (2000, 2002) show that it is superior to VaR in optimization applications.

LASSO method

In this section, we first provide some necessary definitions and outline without proof that LASSO method can asymptotically converge to an optimum with a good choice of a penalty parameter. The detailed proof is given in the Appendix.

We assume all the random variables we consider are continuous, i.e., all the distributions of stocks and the index are continuous. Let α be the significant level for CVaR. Let $\xi_{w,\alpha}$ be the α quantile for the random variable $Y - \sum_{i=1}^{p} w_i R_i$. Let

$$\rho_{w,\alpha}\left(\sum_{i=1}^{p} w_i R_i, Y\right) = \frac{1}{1-\alpha}\left(Y - \sum_{i=1}^{p} w_i R_i\right) I_{[Y-\sum_{i=1}^{p} w_i R_i > \xi_{w_i,\alpha}]},$$

be the loss function where $I_{[Y-\sum_{i=1}^{p} w_i R_i > \xi_{w_i,\alpha}]}$ is an indicator function. The excess shortfall, $\rho_{w_i,\alpha}$, is convex in w_i for all i. Let (Y_t, R_{it}), $t = 1, \ldots, n$ be our sample, and assume they are i .i.d. Let

$$\begin{aligned} E\rho_{w_i,\alpha} &= \frac{1}{n}\sum_{t=1}^{n} E\rho_{w_i,\alpha}\left(\sum_{i=1}^{p} w_i R_{it}, Y_t\right) \\ &= \mathrm{CVaR}_\alpha\left(Y - \sum_{i=1}^{p} w_i R_i\right), \end{aligned}$$

and

$$E_n\rho_{w,\alpha} = \frac{1}{n}\sum_{t=1}^{n} \rho_{w,\alpha}\left(\sum_{i=1}^{p} w_i R_{it}, Y_t\right),$$

which is a sample version of $\mathrm{CVaR_a}$. Define

$$\begin{aligned} w^0 &= \underset{w}{\mathrm{argmin}}\ \ E\rho_{w,\alpha} \\ &= \underset{w}{\mathrm{argmin}}\ \ \mathrm{CVaR}_\alpha\left(Y - \sum_{i=1}^{p} w_i R_i\right), \end{aligned}$$

standing for the best w in theory when we consider $\mathrm{CVaR_a}$. Define

$$\begin{aligned} \mathcal{E}(w) &= E\rho_{w,\alpha} - E\rho_{w^0,\alpha} \\ &= \mathrm{CVaR}_\alpha\left(Y - \sum_{i=1}^{p} w_i R_i\right) \\ &\quad - \mathrm{CVaR}_\alpha\left(Y - \sum_{i=1}^{p} w_i^0 R_i\right), \end{aligned}$$

measuring how far our estimation is from the optimum. Theorem 1 described later shows that the optimization will eventually converge to the optimum weights provided we have enough data. The LASSO estimator in its Lagrange multiplier form is

Definition 1. (Margin Condition) We say the margin condition holds with some strictly convex function G, if there exits $\eta > 0$, for all $\sum_{t=1}^{n}\left(\sum_{i=1}^{p}(w_i - w_i^0)R_{it}\right)^2 \leq \eta$,

$$\mathcal{E}(w_i) > G\left(\sum_{t=1}^{n}\left(\sum_{i=1}^{p}(w_i - w_i^0)R_{it}\right)^2\right).$$

This condition restricts our population error to be exactly larger than zero if w is not from the true model.

$$\hat{w} = \arg\min_{w}\left(E_n\rho_{w.\alpha} + \lambda\sum_{i=1}^{p}|w_i|\right).$$

Our objective in using LASSO is to choose some of the stock weights w_i to be exactly zero. This is an important reason for choosing the LASSO method for optimization. Let $S \subset \{1, \ldots, p\}$, and $S^C = \{1, \ldots, p\}\backslash S$ and $w_{S,i} = w_i I_{[i \in S]}$. Let $|S|$ be the number of elements in S. The next condition sets l_1 norm of the coefficients bounded by some generalized l_2 norm.

Definition 2. (Compatibility Condition) We say the compatibility condition is met for the set S, with constant $\varphi(S) > 0$, if for all w_i, that satisfy $\sum_{i=1}^{p}|w_{S^c,i}| \leq 2\sum_{i=1}^{p}|w_{S,i}|$, it holds that

$$\left(\sum_{i=1}^{p}|w_{S,i}|\right)^2 \leq \sum_{t=1}^{n}\left(\sum_{i=1}^{p}w_iR_{it}\right)^2\frac{|S|}{\varphi^2(S)}.$$

Let $H(v)$ be the conjugate convex for $G(u)$ defined in 3.3, margin condition, and $G(0) = 0$,

$$H(v) = \sup_{u}(uv - G(u)), u \geq 0.$$

Now we can define the concept oracle, which is the true target of LASSO estimator,

Definition 3. (Oracle) w^* is oracle if

$$(w^*, S^*) = \arg\min_{w,S}\left(3\mathcal{E}(w_S) + 2H\left(\frac{4\lambda\sqrt{|S|}}{\varphi(S)}\right)\right),$$

where $S \subset \{1, 2, \ldots, p\}$.

We can view oracle w^* be a best approximation to w^0 with l_0 penalty for number of nonzero entries under CVaR_α loss. In our index replication case, these optimal weights incorporate the trade-off between tracking error and penalty on including too many stocks. Let

$$2\varepsilon^* = 3\mathcal{E}(w^*_{S^*}) + 2H\left(\frac{4\lambda\sqrt{|S^*|}}{\varphi(S^*)}\right).$$

Define empirical process to be

$$v_n(w) = E_n\rho_{w,\alpha} - E\rho_{w,a},$$

which measures the error between sample loss and population loss. Let

$$\mathcal{T} = \left\{(Y, R) : \sup_{\lambda_0\sum_{i=1}^{p}|w_i - w_i^*| \leq \varepsilon^*}|v_n(w) - v_n(w^*)| < \varepsilon^*\right\},$$

and if we choose the proper λ_0 having order $\sqrt{\ln p/n}$, $\mathcal{T}$ will have a large probability by concentration inequality in Bousquet (2002).

(3.1) Theorem 1. (Oracle Inequality for LASSO) Assume compatibility condition holds for all $S \subset \{1, 2, \ldots, p\}$. Assume margin condition holds. $\sum_{t=1}^{n}\left(\sum_{i=1}^{p}(w_i - w_i^0)R_{it}\right)^2 \leq \eta$ for all $\lambda_0\sum_{i=1}^{p}|w_i - w_i^*| \leq \varepsilon^*$ and $\sum_{t=1}^{n}\left(\sum_{i=1}^{p}(w_i^* - w_i^0)R_{it}\right)^2 \leq \eta$. *Suppose* $\lambda > 8\lambda_0$, then on $\mathcal{T}$, we have

$$\mathcal{E}(\hat{w}) + \lambda\sum_{i=1}^{p}|\hat{w}_i - w_i^*| \leq 4\varepsilon^*.$$

Note that if all assumptions hold and $w_i^* = w_i^0$, then $\varepsilon^* = H\left(4\lambda\sqrt{|S^*|}/\varphi(S^*)\right)$, with proper λ_0 having the order $\sqrt{\ln p/n}$, when $n \to \infty, \varepsilon^* \to 0$. In other words, our LASSO estimator $\hat{w}_i \to w_i^0$. This is the asymptotic consistency property of LASSO.

The above theorem guarantees that w^* are the best sparse linear approximation if the relation between Y and R_i is linear. In our application, the weights of stocks $\hat{w}$ we obtain using LASSO penalty will approach the best sparse linear weights w^* when our estimation (training) sample is large enough, even if the underlying relation between Y and R_i is not linear. This justifies the use of the CVAR objective and the LASSO method for selecting the stocks and their weights to include in the index.

Computational technique

We use the techniques outlined by Rockafellar and Uryasev (2002) to transform the stochastic programming problem into a linear programming problem. The standard CVaR minimization problem can be converted to a LP problem by introducing slack variables z_t, for $t = 1, \ldots, n$, where n is the length of time. The resulting LP problem is

$$\min_{\zeta,w,z} \zeta + \frac{1}{1-\alpha}\sum_{t=1}^{n}\frac{1}{n}z_t,$$

subject to constraints,

$$\begin{cases} z_t \geq Y_t - \sum_{i=1}^{p} w_i R_{it} - \zeta, & t = 1, \ldots, n, \\ z_t \geq 0, & t = 1, \ldots, n, \\ \sum_{i=1}^{p} |w_i| \leq s, & \\ \sum_{i=1}^{p} w_i = 1. & \end{cases}$$

Note that $|w_i| = \max\{w_i, -w_i\}$, hence we can change the LASSO type penalty constraint with absolute values for the weights into several linear constraints by introducing dummy variables $u_i, i = 1, \ldots, p$,

$$\begin{cases} \sum_{i=1}^{p} u_i \leq s, & \\ u_i \geq w_i, & i = 1, \ldots, p, \\ u_i \geq -w_i, & i = 1, \ldots, p. \end{cases}$$

So the stochastic programming can be written as a LP problem;

$$\min_{\zeta,w,z,u} \zeta + \frac{1}{n(1-\alpha)}\sum_{t=1}^{n} z_t,$$

subject to constraints,

$$\begin{cases} z_t \geq Y_t - \sum_{i=1}^{p} w_i R_{it} - \zeta, & t = 1, \ldots, n, \\ z_t \geq 0, & t = 1, \ldots, n, \\ \sum_{i=1}^{p} u_i \leq s, & \\ u_i \geq -w_i, & i = 1, \ldots, p, \\ u_i \geq w_i, & i = 1, \ldots, p, \\ \sum_{i=1}^{p} w_i = 1. & \end{cases}$$

IV. Empirical analysis

Data and performance measures

To assess the performance of CLEIR replication methodology, we conducted an out-of-sample analysis of the strategy using the 'rolling horizon approach'. The data are assembled from several sources including Ken French's website, the Center for Research in Security Prices (CRSP) and other stock/portfolio return databases. Following Kan and Zhou (2007), DeMiguel, Garlappi, and Uppal (2009), Goel, Sharma, and Mehra (2018), we consider the following well-diversified portfolio universes: (1) 25 portfolios formed on size and book-to-market ratios, (2) 100 portfolios formed on size and book-to-market ratios, (3) Fama and French (1997) 49 industry portfolios, (4) 100 portfolios formed on size and book-to-market ratios and 49 industry portfolios, (5) simulated 300 individual stocks (using the equal weighted portfolio returns as the benchmark).[2] The benchmark is the value-weighted index return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ.

The data sample is from November 1988 to October 2018. For each year t, we first estimated the weighted for the portfolios/stocks to include in the index using the daily stock return data from year $t-1$. Then we used the estimated stock weights to calculate the enhanced index return and its properties and at year t compare it to the benchmark index. Following Rockafellar and Uryasev (2002) and Bamberg and Wagner (2000), we use risk-tolerance level $\omega = 0.95$ in the CVaR constraint. The LASSO penalty equals to 1.5, which keeps the number of negative

[2]Following Goto and Xu (2015), we randomly select 300 stocks with careful imputation of the missing data using the value-weighted index returns, then calculate the mean vector and variance-covariance matrix as the true parameters to generate 30 years return vectors under the multivariate normal distribution assumptions. Next, we still use the 'rolling horizon approach' as the portfolio universes (1)-(4), and use the equal weighted portfolio returns as the benchmark.

weights low (less short selling). The LASSO penalty s can be alternatively selected by cross-validation.

We compare the CLEIR performance with the several alternative weighting schemes: (1) CLEIR without LASSO constraint (CEIR) (2) value weighted index return (VW), (3) equal weighted index return (EW), (4) plug-in mean variance portfolio (MV), (5) Global Minimum Variance portfolio (GMV), (6) Equal Variance portfolio (EV), (7) Jin and Wang (2016) method (JW). Besides traditional portfolio measures such as holding period returns and Sharpe Ratios, we further implement the following two performance measures to evaluate the index tracking performance of the replicated index.

$$s_{STD} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(e_t - \bar{e})^2}, \tag{4.1}$$

and one robust alternative, standardized Median Absolute Deviation (MAD),

$$s_{MAD} = \frac{1}{\Phi^{-1}(3/4)} med_{t'}(|e_{t'} - median_t(e_t)|), where,\ e_t \text{ is the TE.} \tag{4.2}$$

The measure s_{MAD} is more robust because s_{STD} would be influenced when there are some large absolute tracking error due to extreme event, or outliers. Bamberg and Wagner (2000) provide details on the components of the risk in the tracking index.

Table 1 reports the tracking error performance measures (s_{STD} and s_{MAD}) of different weighting schemes in different portfolio universes. Value weighted portfolios have zero tracking error since they are full replication of the benchmark index, while CLEIR has very robust performance with an annualized tracking error (s_{STD}) around 1.5%. Without the LASSO constraint, the CEIR method performs well in the first three portfolio universes however in the universe of '100 portfolios formed on size and book-to-market ratios and 49 industry portfolios' and 'simulated 300 stocks,' the tracking error becomes much larger and the portfolio based on this method cannot track the index closely. The empirical evidence shows that the LASSO technique helps provide a more robust performance particularly when more constituents are included in replicating the benchmark index. The remaining methods (EW, MV, GMV, EV, JW) have much larger tracking errors than the CLEIR method. Table 1 Panel B reports the robustness check using the alternative measure s_{MAD} performance, and the findings are similar to Panel A.

Table 2 reports the out-of-sample Sharpe Ratios of the different weighting schemes in different stock universes and Table 3 further reports the corresponding out-of-sample returns. When the number of constituents is not large, for example, 25 or 49 constituents shown in the first two rows, the Sharpe Ratios of CLEIR are very close to the benchmark. However, when the number of the constituents is relatively large, for example, more than 100 constituents in the latter three cases, the Sharpe Ratios are much improved compared to the benchmark. The CEIR, CLEIR without LASSO, has relatively poor performance in the universe of '100 portfolios formed on size and

Table 1. Out-of-sample index tracking performance measures from 1988 to 2018.

	CLEIR	CEIR	VW	EW	MV	GMV	EV	JW
Panel A: s_{STD}(%)								
SZBM25	1.000	1.000	0.000	5.628	234.982	13.939	5.692	6.277
IND49	1.893	1.894	0.000	4.313	125.160	15.290	4.344	7.848
SZBM100	1.249	1.533	0.000	5.863	403.277	15.481	5.853	6.546
IND49+ SZBM100	1.587	102.559	0.000	4.926	210.839	16.806	4.961	6.605
Simulation (300 Stocks)	2.611	193.587	–	0.000	$T < N$	$T < N$	0.935	4.563
Panel B: s_{MAD}(%)								
SZBM25	0.476	0.475	0.000	4.525	21.546	10.408	4.466	3.114
IND49	0.569	0.575	0.000	3.022	41.068	11.029	2.560	5.386
SZBM100	0.848	0.888	0.000	4.739	30.495	11.032	4.631	3.995
IND49+ SZBM100	0.963	84.805	0.000	3.895	40.083	12.435	3.730	4.031
Simulation (300 Stocks)	2.558	192.234	-	0.000	$T < N$	$T < N$	0.936	4.632

Table 2. Out-of-sample Sharpe ratios from 1988 to 2018.

Sharp Ratio	CLEIR	CEIR	VW	EW	MV	GMV	EV	JW
SZBM25	0.630	0.630	0.629	0.673	0.082	1.319	0.700	0.858
IND49	0.621	0.621	0.629	0.680	0.009	0.758	0.702	0.763
SZBM100	0.651	0.656	0.629	0.678	−0.074	1.469	0.709	0.872
IND49+ SZBM100	0.648	0.463	0.629	0.684	0.398	1.252	0.714	0.889
Simulation (300 Stocks)	1.127	1.066	-	0.985	$T < N$	$T < N$	1.013	1.196

Table 3. Out-of-sample returns from 1988 to 2018.

Return (%)	CLEIR	CEIR	VW	EW	MV	GMV	EV	JW
SZBM25	10.956	10.957	10.950	12.196	19.188	15.642	12.414	13.462
IND49	10.992	11.000	10.950	11.560	1.060	8.484	11.345	10.001
SZBM100	11.342	11.370	10.950	12.292	−29.807	15.316	12.458	13.173
IND49+ SZBM100	11.368	48.560	10.950	12.051	83.355	13.935	12.123	12.325
Simulation (300 Stocks)	18.140	207.732	–	15.528	$T < N$	$T < N$	15.523	14.930

book-to-market ratios and 49 industry portfolios'. As DeMiguel, Garlappi, and Uppal (2009) documented, the equal weighting strategy (EW) has very good and robust performance, and the Equal Variance and Jin and Wang (2016) methods also have good and robust performance though the tracking errors are relatively large (see Jin and Wang 2016). The plug-in mean variance (MV) and global minimum variance (GMV) performances are not very consistent in different portfolio universes. MV achieves low Sharpe ratios (in Table 2) and big variation in out-of-sample returns (in Table 3). In most of the cases, GMV performs well however the return in '49 industry portfolios' (in Table 3) is relatively low though the whole portfolio based on GMV method has reasonable out-of-sample Sharpe ratio (in Table 2).In the next section, we provide additional out-of-sample tests based on sampling replications in the S&P 500.

Additional test on S&P 500 using 100 largest stocks

To further analyse the performance of CLEIR enhanced index replication, we conducted a comprehensive out-of-sample examination over the period 2003 to 2012. The procedure is described below.

(1) Choose 100 largest stocks listed on the NYSE-AMEX by market value on the last trading day of year $t-2$.
(2) Estimate CLEIR index stock weights using one year of daily data for the largest 100 market capitalization stocks in year $t-1$, some weights could be zero.
(3) Compare the performance of the index portfolio based on the weights chosen in (i) with the benchmark $S\&P500$ index for the year t.
(4) Repeat (i), (ii) and (iii) for years t= 2003 to 2012.

In Table 4 we report the out-of-sample performances for consecutive ten years from 2003–2012. For nine out of ten years, the out-of-sample CLEIR return consistently beats the $S\&P$ 500 return with a minimum of .25% per year to a maximum of 6.67% per year. The only year it underperformed was in 2007 and the under-performance was only 0.4% per year. More strikingly the CLEIR index

Table 4. Out-of-sample performance of CLEIR index from 2003 to 2012.

Year	No. Stocks	CLEIR Ret (Annual %)	$S\&P$500 Ret (Annual %)	(CLEIR-$S\&P$500 (Annual %)	Correlation
2003	76	23.34	22.32	1.02	99.1
2004	74	11.13	9.33	1.79	99.0
2005	72	4.10	3.84	0.25	98.6
2006	73	14.20	11.78	2.42	98.3
2007	70	3.25	3.65	−0.40	99.3
2008	83	−30.92	−37.59	6.67	98.3
2009	77	23.16	19.67	3.49	98.7
2010	54	13.52	11.00	2.52	99.3
2011	73	2.12	−1.12	3.24	99.4
2012	52	13.56	11.68	1.89	98.0

outperformed the $S\&P500$ index by 6.67% during the 2008 financial crisis when the market fell by 37.59%. This supports our contention that by minimizing CVaR we can not only consistently ouperform the benchmark, but can also minimize large negative returns. The average out-of-sample CLEIR excess return is 2.3% (over the period 2003–2012). Also, we document that the average standard deviation is close to the benchmark and the CLEIR index has a high correlation with the benchmark. In the appendix, we report the out-of-sample performance of the daily return and the level of the index and the benchmark. From these figures, we can clearly see that our CLEIR algorithm also passes the out-of-sample robustness check and consistently beats the $S\&P$ 500 benchmark.

We implement the Fama and French (1993) three factor model and Carhart (1997) four-factor Model to study the alpha after controlling for the market risk premium, Small-minus-Big factor (SMB), High-minus-Low factor (HML) and momentum factor (MOM). The result is reported in Table 5. The CLEIR method demonstrates a very robust performance in the annual alpha, with an average of 2.44%. The main exposure of CLEIR is from the market risk, with a coefficient of $Mkt - R_f$ 0.91, with t-statistic 24.31. The coefficients on SMB and HML are insignificant, which shows the outperformance by CLEIR is not from SMB and HML factors but the alpha generated by CLIER. The Fama and French (1993) and Carhart (1997) verify the previous findings and a significant alpha is found by CLEIR.

Table 5. Fama and French (1993) and Carhart (1997) model risk decomposition.

	α	$Mkt - R_f$	SMB	HML	MOM	Adjust R^2
S&P 500	0.00	1.00	0.00	0.00	0.00	1.00
		(–)				
CLEIR	2.55	0.93				0.99
	(4.70)	(28.34)				
	2.31	0.91	0.07			0.99
	(4.15)	(26.26)	(1.25)			
	2.55	0.93		0.00		0.99
	(4.29)	(26.33)		(0.06)		
	2.33	0.91	0.07	−0.01		0.99
	(3.82)	(24.31)	(1.16)	(−0.10)		
	2.35	0.88	0.03	0.02	−0.04	0.99
	(4.90)	(26.79)	(0.66)	(0.32)	(−2.17)	

Figure 1 shows the cumulative CLEIR index performance from 2003 to 2012. The starting point for the CLEIR index and $S\&P500$ is

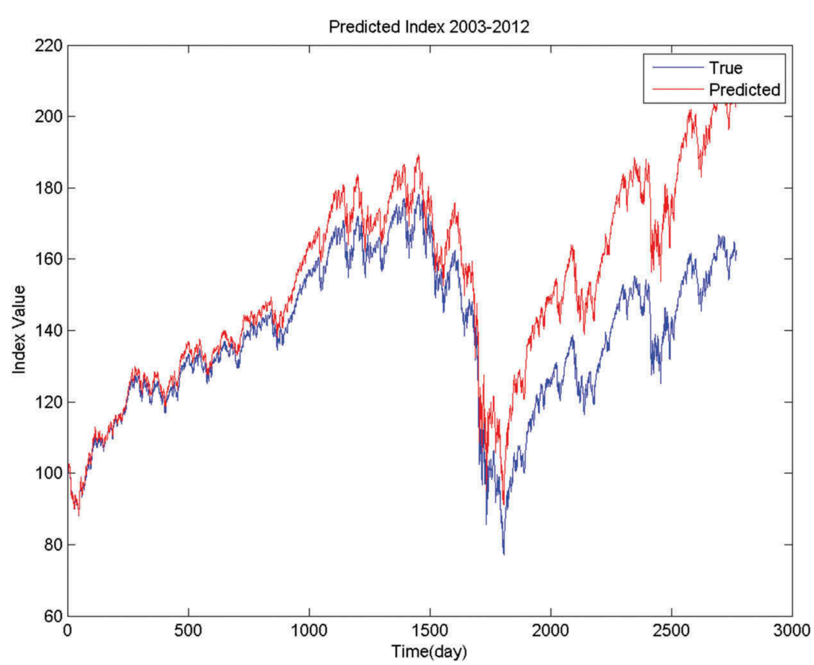

Figure 1. Out-of-sample CLEIR index performance from 2003 to 2012.

normalized to 100 on 1 January 2003. After ten years, our CLEIR index outperforms the *S&P*500 by 60%!

V. Conclusion

In this paper, we propose a new enhanced indexation method, which we call CVaR-LASSO Enhanced Index Replication (CLEIR). This method of enhanced replication minimizes the CVaR of the TE between the index and a benchmark using a LASSO-l_1 minimization constraint. The CVaR objective function is a new way to track a benchmark portfolio compared to the usual square error loss or other type of overall mean loss functions. CVaR has an attractive property in that it minimizes the likelihood of the tracking portfolio under performing the benchmark by a large amount, while most other loss functions penalize both under-performance and over-performance. Although the traditional loss functions may efficiently control the tracking errors, they also limit the ability to outperform the benchmark. We implement a transformation similar to that developed by Rockafellar and Uryasev (2000) and solve a linear programming problem along with the LASSO penalty and portfolio constraints. The CLEIR method has several nice properties: (a) sparsity in stock selection by shrinking some of the weights to zero; (b) sampling with a subset of stocks to ensure liquidity and low transactions costs; and (c) an automatic stock selection process l_1 constraint. In out-of-sample 'rolling horizon approach,' our CLEIR method outperformed the benchmark and other methods with low tracking error deviation (standard deviation or median absolute deviation), and the effect is more pronounced when the number of assets in the portfolio is large. Further over a ten-year period (2003–2012) out-of-sample test, the CLEIR method outperforms the benchmark (*S&P*500) by 2.3% per year while tracking the index closely. More importantly, the CLEIR index had a much lower loss compared to the *S&P*500 index during the financial crisis in 2008. We believe the methodology would also be very useful for enhanced index fund managers engaged in tracking indices having large number of stocks.

VI. Proof

Proof of lemma 1

Proof.

$$
\begin{aligned}
E(X \mid X \geq \zeta) &= P(X \geq \zeta)^{-1} \int_{\{X\geq\zeta\}} x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= P(X \geq \zeta)^{-1} \int_{\{X\geq\zeta\}} (x-\mu+\mu) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= P(X \geq \zeta)^{-1} \mu \int_{\{X\geq\zeta\}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&+ P(X \geq \zeta)^{-1} (-\sigma^2) \int_{\{X\geq\zeta\}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) d\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
&= \mu + P(X \geq \zeta)^{-1} \left(-\sqrt{\frac{\sigma^2}{2\pi}}\right) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \Bigg|_{x=\zeta}^{x=+\infty} \\
&= \mu + \frac{1}{P(X \geq \zeta)} \sqrt{\frac{\sigma^2}{2\pi}} \exp\left(-\frac{(\zeta-\mu)^2}{2\sigma^2}\right).
\end{aligned}
$$

Proof of theorem 1

(6.1) *Proof*

Since the inequality (3.1) in result is valid almost surely in $\mathcal{T}$, we assume $(Y, R_i) \in \mathcal{T}$ in the following proof. We first prove for some $\tilde{w}$ near w^*, the inequality (3.1) holds, and then show $\hat{w}$ is also near w^*. Let

$$
t = \frac{\varepsilon^*}{\varepsilon^* + \lambda_0 \sum_{i=1}^{p} |\hat{w}_i - w_i^*|},
$$

and we define $\tilde{w}_i = t\hat{w}_i + (1-t)w_i^*$, $\tilde{\mathcal{E}} = \mathcal{E}(\tilde{w})$, $\mathcal{E}^* = \mathcal{E}(w^*)$. Then

$$
\begin{aligned}
\lambda_0 \sum_{i=1}^{p} |\tilde{w}_i - w_i^*| &= t\lambda_0 \sum_{i=1}^{p} |\hat{w}_i - w_i^*| \\
&= \varepsilon^* \frac{\lambda_0 \sum_{i=1}^{p} |\hat{w}_i - w_i^*|}{\varepsilon^* + \lambda_0 \sum_{i=1}^{p} |\hat{w}_i - w_i^*|} \leq \varepsilon^*,
\end{aligned}
$$

which means $\tilde{w}$ is around w^*. Then by definition,

$$\begin{aligned}\tilde{\mathcal{E}} + \lambda\sum_{i=1}^{p}|\tilde{w}_i| &= E\rho_{\tilde{w},\alpha} - E\rho_{w^0,\alpha} + \lambda\sum_{i=1}^{p}|\tilde{w}_i|\\ &= \left[\left(E_n\rho_{\tilde{w},\alpha} - E\rho_{\tilde{w},\alpha}\right) - \left(E_n\rho_{w^*,\alpha} - E\rho_{w^*,\alpha}\right)\right]\\ &\quad + \left(E_n\rho_{\tilde{w},\alpha} - E_n\rho_{w^*,\alpha}\right)\\ &\quad + \left(E\rho_{w^*,\alpha} - E\rho_{w^0,\alpha}\right)\end{aligned}$$

$$\begin{aligned}&= -[v_n(\tilde{w}) - v_n(w^*)]\\ &\quad + \left[\left(E_n\rho_{\tilde{w},\alpha} + \lambda\sum_{i=1}^{p}|\tilde{w}_i|\right) - \left(E_n\rho_{w^*,\alpha} + \lambda\sum_{i=1}^{p}\left|w_i^*\right|\right)\right]\\ &\quad + \mathcal{E}^* + \lambda\sum_{i=1}^{p}\left|w_i^*\right|,\\ &= -P + L + E,\end{aligned}$$

Where,

$$P = v_n(\tilde{w}) - v_n(w^*),$$

is the empirical process,

$$L = \left(E_n\rho_{\tilde{w},\alpha} + \lambda\sum_{i=1}^{p}|\tilde{w}_i|\right) - \left(E_n\rho_{w^*,\alpha} + \lambda\sum_{i=1}^{p}\left|w_i^*\right|\right),$$

is the difference of objective function of CVaR-LASSO,

$$E = \mathcal{E}^* + \lambda\sum_{i=1}^{p}\left|w_i^*\right|,$$

is the error term evaluated at w^*. Because we assume $(Y, R_i) \in \mathcal{T}$, and by the definition of the set $\mathcal{T}$,

$$\begin{aligned}-P &= -[v_n(\tilde{w}) - v_n(w^*)]\\ &\leq \sup_{\lambda_0\sum_{i=1}^{p}\left|w_i - w_i^*\right| \leq \varepsilon_0^*} |v_n(w) - v_n(w^*)| \leq \varepsilon^*.\end{aligned}$$

Because we assume ρ is convex, and $\hat{w}$ is defined as the global minimum,

$$\begin{aligned}E_n\rho_{\tilde{w},\alpha} + \lambda\sum_{i=1}^{p}|\tilde{w}_i| &\leq t\left(E_n\rho_{\hat{w}_i,\alpha} + \lambda\sum_{i=1}^{p}|\hat{w}_i|\right)\\ &\quad + (1-t)\left(E_n\rho_{w^*,\alpha} + \lambda\sum_{i=1}^{p}\left|w_i^*\right|\right)\\ &\leq t\left(E_n\rho_{w^*,\alpha} + \lambda\sum_{i=1}^{p}\left|w_i^*\right|\right) + (1-t)\\ &\quad \left(E_n\rho_{w^*,\alpha} + \lambda\sum_{i=1}^{p}\left|w_i^*\right|\right)\\ &= E_n\rho_{w^*,\alpha} + \lambda\sum_{i=1}^{p}\left|w_i^*\right|.\end{aligned}$$

Hence $L \leq 0$. Therefore, $\tilde{\mathcal{E}} + \lambda\sum_{i=1}^{p}|\tilde{w}_i| \leq -P + L + E \leq \varepsilon^* + \mathcal{E}^* + \lambda\sum_{i=1}^{p}\left|w_i^*\right|$. By definition of ε^*, we know $\mathcal{E}^* \leq \varepsilon^*$, then

$$\begin{aligned}\tilde{\mathcal{E}} + \lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^{*C},i}\right| &= \tilde{\mathcal{E}} + \lambda\sum_{i=1}^{p}|\tilde{w}_i| - \lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i}\right|\\ &\leq \varepsilon^* + \mathcal{E}^* + \left(\lambda\sum_{i=1}^{p}\left|w_i^*\right| - \lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i}\right|\right)\\ &\leq 2\varepsilon^* + \lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right|.\end{aligned}$$

Hence

$$\begin{aligned}&\tilde{\mathcal{E}} + \lambda\sum_{i=1}^{p}\left|\tilde{w}_i - w_i^*\right|\\ &= \tilde{\mathcal{E}} + \lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^{*C},i}\right| + \lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right|\\ &\leq 2\varepsilon^* + \lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right| + \lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right|\\ &\leq 2\varepsilon^* + 2\lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right|.\end{aligned}$$

The inequality (3.1) will hold if $2\lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right|$ is also bounded by ε^*. Now we consider two cases: i) $\lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right| < \varepsilon^*$ and ii) $\lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right| \geq \varepsilon^*$.

In case i), $\lambda\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right| < \varepsilon^*$, by (7.1)

$$\tilde{\mathcal{E}} + \lambda\sum_{i=1}^{p}\left|\tilde{w}_i - w_i^*\right| \leq 4\varepsilon^*.$$

This will leads to

$$\lambda_0 \sum_{i=1}^{p} \left|\tilde{w}_i - w_i^*\right| = \frac{\lambda_0}{\lambda}\left(\lambda \sum_{i=1}^{p}\left|\tilde{w}_i - w_i^*\right|\right) \leq \frac{\lambda_0}{\lambda} 4\varepsilon^* \leq \frac{\varepsilon^*}{2},$$

if $\lambda \geq 8\lambda_0$ and $\lambda_0 \sum_{i=1}^{p}\left|\hat{w}_i - w_i^*\right| \leq \varepsilon^*$.

In case ii), note that $w^*_{S^{*C},i} = 0$,

$$\lambda \sum_{i=1}^{p}\left|\tilde{w}_{S^{*C},i} - w^*_{S^{*C},i}\right| = \lambda \sum_{i=1}^{p}\left|\tilde{w}_{S^{*C},i}\right| \leq 2\varepsilon^* + \lambda \sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right| \leq 3\lambda \sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right|.$$

We apply the compatibility condition in Definition 2 for coefficients $\tilde{w} - w^*$ and set S^*,

$$\left(\sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right|\right)^2 \leq \sum_{t=1}^{n}\left[\sum_{i=1}^{p}\left(\tilde{w}_{S^*,i} - w_i^*\right)R_{ij}\right]^2 \frac{|S^*|}{\varphi^2(S^*)}.$$

Hence by (7.1),

$$\tilde{\mathcal{E}} + \lambda \sum_{i=1}^{p}\left|\tilde{w}_{S^{*C},i}\right| + \lambda \sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right| \leq \varepsilon^* + \mathcal{E}^* + \left\{\sum_{t=1}^{n}\left[\sum_{i=1}^{p}\left(\tilde{w}_{S^*,i} - w_i^*\right)R_{ij}\right]^2\right\}\left[\frac{2\lambda|S^*|}{\varphi^2(S^*)}\right]$$

By the definition of $H(v)$, we have

$$uv \leq H(v) + G(u),$$

hence together with marginal condition,

$$\frac{1}{2}\left\{\sum_{t=1}^{n}\left[\sum_{i=1}^{p}\left(\tilde{w}_{S^*,i} - w_i^*\right)R_{ij}\right]^2\right\}\left[\frac{4\lambda|S^*|}{\varphi^2(S^*)}\right]$$
$$\leq \frac{1}{2}\left\{\sum_{t=1}^{n}\left[\sum_{i=1}^{p}\left(\tilde{w}_{S^*,i} - w_i^0\right)R_{ij}\right]^2 + \sum_{t=1}^{n}\left[\sum_{i=1}^{p}\left(\tilde{w}_{S^*,i} - w_i^0\right)R_{ij}\right]^2\right\}\left[\frac{4\lambda|S^*|}{\varphi^2(S^*)}\right]$$
$$\leq H\left(\frac{4\lambda|S^*|}{\varphi^2(S^*)}\right) + G\left(\frac{1}{2}\left\{\sum_{t=1}^{n}\left[\sum_{i=1}^{p}\left(\tilde{w}_{S^*,i} - w_i^0\right)R_{ij}\right]^2 + \sum_{t=1}^{n}\left[\sum_{i=1}^{p}\left(\tilde{w}_{S^*,i} - w_i^0\right)R_{ij}\right]^2\right\}\right)$$
$$\leq H\left(\frac{4\lambda|S^*|}{\varphi^2(S^*)}\right) + \frac{1}{2}G\left(\sum_{t=1}^{n}\left[\sum_{i=1}^{p}\left(\tilde{w}_{S^*,i} - w_i^0\right)R_{ij}\right]^2\right) + \frac{1}{2}G\left(\sum_{t=1}^{n}\left[\sum_{i=1}^{p}\left(\tilde{w}_{S^*,i} - w_i^0\right)R_{ij}\right]^2\right)$$
$$\leq H\left(\frac{4\lambda|S^*|}{\varphi^2(S^*)}\right) + \frac{\tilde{\mathcal{E}}}{2} + \frac{\mathcal{E}^*}{2}.$$

Hence,

$$\tilde{\mathcal{E}} + \lambda \sum_{i=1}^{p}\left|\tilde{w}_i - w_i^*\right| = \tilde{\mathcal{E}} + \lambda \sum_{i=1}^{p}\left|\tilde{w}_{S^{*C},i}\right| + \lambda \sum_{i=1}^{p}\left|\tilde{w}_{S^*,i} - w_i^*\right| \leq \varepsilon^* + \mathcal{E}^* + H\left(\frac{4\lambda|S^*|}{\varphi^2(S^*)}\right) + \frac{\tilde{\mathcal{E}}}{2} + \frac{\mathcal{E}^*}{2} = 2\varepsilon^* + \frac{\tilde{\mathcal{E}}}{2},$$

or

$$\frac{\tilde{\mathcal{E}}}{2} + \lambda \sum_{i=1}^{p}\left|\tilde{w}_i - w_i^*\right| \leq 2\varepsilon^*,$$

$$\tilde{\mathcal{E}} + \lambda \sum_{i=1}^{p}\left|\tilde{w}_i - w_i^*\right| \leq \tilde{\mathcal{E}} + 2\lambda \sum_{i=1}^{p}\left|\tilde{w}_i - w_i^*\right| \leq 4\varepsilon^*.$$

By choosing $\lambda \geq 8\lambda_0$,

$$\lambda_0 \sum_{i=1}^{p} \left|\tilde{w}_i - w_i^*\right| = \frac{\lambda_0}{\lambda}\left(\lambda \sum_{i=1}^{p} \left|\tilde{w}_i - w_i^*\right|\right) \leq \frac{\lambda_0}{\lambda} 2\varepsilon^* \leq \frac{\varepsilon^*}{2}.$$

By the definition of $\tilde{w}$,

$$\frac{\varepsilon^*}{2} \geq \lambda_0 t \sum_{i=1}^{p} \left|\hat{w}_i - w_i^*\right| = \frac{\varepsilon^*\left(\lambda_0 \sum_{i=1}^{p} \left|\hat{w}_i - w_i^*\right|\right)}{\varepsilon^* + \lambda_0 \sum_{i=1}^{p} \left|\hat{w}_i - w_i^*\right|},$$

hence $\lambda_0 \sum_{i=1}^{p} \left|\hat{w}_i - w_i^*\right| \leq \varepsilon^*$

Until now, we prove two facts, a) $\lambda_0 \sum_{i=1}^{p} \left|\hat{w}_i - w_i^*\right| \leq \varepsilon^*$, and b) for any $\tilde{w}$ satisfying $\lambda_0 \sum_{i=1}^{p} \left|\tilde{w}_i - w_i^*\right| \leq \varepsilon^*$,

$$\tilde{\mathcal{E}} + \lambda \sum_{i=1}^{p} \left|\tilde{w}_i - w_i^*\right| \leq 4\varepsilon^*.$$

Hence

$$\hat{\mathcal{E}} + \lambda \sum_{i=1}^{p} \left|\hat{w}_i - w_i^*\right| \leq 4\varepsilon^*.$$

Theorem 1 holds.

Acknowledgments

We thank Mark Taylor (Editor), the anonymous reviewer, for helpful discussions and useful suggestions. Yong Jin also acknowledges generous financial support of the Research Grant Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 25508217) and the Learning and Teaching Enhancement Grant (Project No.: 1.21.xx.8ADP). All errors are our own.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Hong Kong Polytechnic University [The Learning and Teaching Enhancement Grant (Project No.: 1.21.xx.8ADP)];Research Grants Council, University Grants Committee [Project No. PolyU 25508217].

ORCID

Yong Jin http://orcid.org/0000-0002-1544-1082

References

Amenc, N., F. Goltz, L. Martellini, and P. Retkowsky. 2011. "Efficient Indexation: An Alternative to Cap-Weighted Indices." *Journal Of Investment Management* 9 (4): 1–23.

Arnott, R. D., J. Hsu, and P. Moore. 2005. "Fundamental Indexation." *Financial Analysts Journal* 61 (2): 83–99.

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath. 1999. "Coherent Measures of Risk." *Mathematical Finance* 9 (3): 203–228.

Bamberg, G., and N. Wagner. 2000. "Equity Index Replication with Standard and Robust Regression Estimators." *OR-Spectrum* 22 (4): 525–543.

Basak, S., and A. Shapiro. 2001. "Value-At-Risk-Based Risk Management: Optimal Policies and Asset Prices." *Review of Financial Studies* 14 (2): 371–405.

Basak, S., A. Shapiro, and L. Tepla. 2006. "Risk Management with Benchmarking." *Management Science* 52 (4): 542–557.

Blume, M. E., and R. M. Edelen 2002. "On Replicating the *S&P*500 Index." Rodney L White Center for Financial Research - Working Papers.

Bonami, P., and M. A. Lejeune. 2009. "An Exact Solution Approach for Portfolio Optimization Problems under Stochastic and Integer Constraints." *Operations Research* 57 (3): 650–670.

Bousquet, O. 2002. "A Bennett Concentration Inequality and Its Application to Suprema of Empirical Processes." *Comptes Rendus Mathematique* 334 (6): 495–500.

Cai, L., Y. Jin, Q. Qi, and X. Xu. 2018. "A Comprehensive Study on Smart Beta Strategies in the A-Share Market." *Applied Economics* 50 (55): 6024–6033.

Chavez-Bedoya, L., and J. R. Birge. 2014. "Index Tracking and Enhanced Indexation Using a Parametric Approach." *Journal of Economics Finance and Administrative Science* 19 (36): 19–44.

Chen, H., H. Desai, and S. Krishnamurthy. 2013. "A First Look at Mutual Funds that Use Short Sales." *Journal of Financial and Quantitative Analysis* 48 (3): 761–787.

Chopra, V. K., and W. T. Ziemba. 1993. "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice." *The Journal of Portfolio Management* 19 (2): 6–11.

Choueifaty, Y., and Y. Coignard. 2008. "Toward Maximum Diversification." *Journal of Portfolio Management* 35 (1): 40.

Chow, T. M., J. Hsu, V. Kalesnik, and B. Little. 2011. "A Survey of Alternative Equity Index Strategies." *Financial Analysts Journal* 67 (5): 37–57.

Clarke, R. G., H. De Silva, and S. Thorley. 2006. "Minimum-Variance Portfolios in the US Equity Market." *The Journal of Portfolio Management* 33 (1): 10–24.

DeMiguel, V., L. Garlappi, and R. Uppal. 2009. "Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?" *Review of Financial Studies* 22 (5): 1915–1953.

Fernholz, R. 1995. "Portfolio Generating Functions." Working paper, INTECH (December).

Goel, A., A. Sharma, and A. Mehra. 2018. "Index Tracking and Enhanced Indexing Using Mixed Conditional Value-At-Risk." *Journal of Computational and Applied Mathematics* 335: 361–380.

Goel, A., A. Sharma, and A. Mehra. 2019. "Robust Optimization of Mixed CVaR STARR Ratio Using Copulas." *Journal of Computational and Applied Mathematics* 347: 62–83.

Goto, S., and Y. Xu. 2015. "Improving Mean Variance Optimization through Sparse Hedging Restrictions." *Journal of Financial and Quantitative Analysis* 50 (6): 1415–1441.

Haugen, R. A., and N. L. Baker. 1991. "The Efficient Market Inefficiency of Capitalization-Weighted Stock Portfolios." *The Journal of Portfolio Management* 17 (3): 35–40.

Investment Company Institute (US). 2017. *Investment Company Fact Book*. Washington, DC: Investment Company Institute.

Jin, Y., and L. Wang. 2016. "Beat Equal Weighting: A Strategy for Portfolio Optimisation." *Risk* December: 87–91.

Kan, R., and G. Zhou. 2007. "Optimal Portfolio Choice with Parameter Uncertainty." *Journal of Financial and Quantitative Analysis* 42 (3): 621–656.

Lejeune, M. A., and G. Samatli-Pa. 2013. "Construction of Risk-Averse Enhanced Index Funds." *INFORMS Journal on Computing* 25 (4): 701–719.

Michaud, R. O. 1989. "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal* 45 (1): 31–42.

Rockafellar, R. T., and S. Uryasev. 2000. "Optimization of Conditional Value-At-Risk." *Journal of Risk* 2: 21–42.

Rockafellar, R. T., and S. Uryasev. 2002. "Conditional Value-At-Risk for General Loss Distributions." *Journal of Banking & Finance* 26 (7): 1443–1471.

Sarykalin, S., G. Serraino, and S. Uryasev. 2008. "VaR Vs CVaR in Risk Management and Optimization." *Tutorials in Operations Research, INFORMS* 270–294.

Tibshirani, R. 1996. "Regression Shrinkage and Selection via the LASSO." *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1): 267–288.