Sam Bales
11/14/2019
MATH 624

Regularization Methods for Constructing Optimal Portfolios

Introduction & Summary

Constructing Optimal Sparse Portfolios using Regularization Methods by B. Fastrich, S. Paterlini, and P. Winker (2014) presents a novel approach for using regularization methods to select an optimal portfolio. To preface, the objective of an investor is to maximize risk-adjusted return from a vector of K asset returns $\boldsymbol{\mu}$ and asset return covariance matrix $\boldsymbol{\Sigma}$ by selecting an optimal vector of asset weights $\boldsymbol{w}$. This is a simple optimization problem in theory, but estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ tend to be very unreliable. Thus, implementation of this naïve approach often results in sub-optimal portfolios that are locally optimized to the most recent data. It is well-established in the literature that regularization methods can increase portfolio performance. Fastrich, Paterlini, and Winker (2014) presents a comprehensive comparison of regularization methods for portfolio selection.

The authors start with the objective function:

$$w^* = argmin\left[w'\boldsymbol{\Sigma}w + \lambda\, \Sigma_{\mathrm{i=1}}^{\mathrm{K}} \rho(w_i)\right] \; such\ that\ \mathbf{1}'_K w = 1 \qquad (1)$$

where $w^* \in R^K$ is the optimal vector of asset weights, $\mathbf{1}'_K \in R^{k\,x\,k}$ matrix of 1's, λ is the regularization parameter and $\rho(\ldots)$ is the penalty function. If L_1 and L_2 norm regularization is used, one gets the ordinary LASSO and Ridge estimators. The main focus of the paper is to examine the impact of choosing $\rho(\ldots)$'s that differ from the traditional penalties, and to develop a new penalty function that includes prior information.

One assumption of Ridge and LASSO regularization methods is explanatory variables have equal importance a priori. In the context of portfolio construction, this is a suboptimal property since there are many are well-known asset-price behaviors and the potential for expert knowledge to increase portfolio performance. The authors developed three signal functions that estimate a quantity ω_{i} that incorporates prior information into the objective function, turning $\rho(\ldots)$ in (1) into $\omega_{\mathrm{i}} w_i$. Adding ω_i allows information from the signal functions to influence the contribution of that weight to the aggregate penalty, thereby allowing precise adjustments of asset weights. In addition to this new development, the authors compare and contrast the effects of choosing among pre-established $\rho(\ldots)$'s.

To increase performance and mitigate transactions costs, sparse portfolios are preferred over complex portfolio, ceteris paribus. Hence, the authors focused on four non-convex $\rho(\ldots)$'s, which allow for heavy penalization of small absolute values of w_i. Penalizing small asset weights leads to a more concentrated portfolio with meaningful positions. The authors compared these four existing penalty functions to the new penalty function on a small dataset and a larger, more realistic dataset. They found the non-traditional regularization methods outperformed conventional regularization methods on the small dataset. In addition, the non-convex penalty functions all outperformed Ridge and LASSO penalties in selecting sparse portfolios with minimal portfolio risk (standard deviation) when employed on a large dataset with many potential assets to invest in. However, these solutions require a larger investment in deployment.

The performance of regularized methods is highly dependent upon the value of the tuning parameter. Choosing the correct value of lambda becomes more important when using more "aggressive" regularization functions, such as the non-convex models proposed in the papers. To deploy these optimization methods effectively, the authors cautioned that portfolio managers must be diligent in carefully selecting the tuning parameter to avoid negative consequences. The authors employed cross-validation to tune their hyperparameters and encouraged practitioners to do the same.

Reflection

This implementation of regularization methods differs moderately from that which was learned in 624. First, regularization methods in this paper are applied to a convex-optimization problem of selecting asset weights, instead of optimizing a regression loss function. It is important to keep in mind these tasks are essentially mathematically the same. In the case of regression, Lagrange multipliers are used to minimize the loss function while constraining the magnitude of the coefficients. In the case of portfolio optimization, the loss function—portfolio risk—is minimized while constraining the number of positions in the portfolio. Although this is not explicitly a regression tasks, they are mathematical cousins and still require many estimates and statistical properties.

Second, the paper proposed many $\rho(\ldots)'s$ that were not discussed in class. I think this is a results from the difference int the outcome of the two methods. For statistical learning methods, we just want more accurate predictions by selecting the most important predictors; the predictors ultimately selected don't matter to us, we just want accuracy. Portfolio construction is only concerned with portfolio performance, and thus the assets—analogous to features—selected becomes critically important. I think it is important to consider the overarching objective of the method when considering these methods, and to keep in mind domain-specific regularization functions may exist.

Sam Bales
11/14/2019
MATH 624

Finally, I find this paper to be utterly fascinating, but am skeptical of being able to deploy it in a conventional setting. Tuning hyperparameters for relatively simple $\rho(\ldots)$'s is no small task. When the complexity of $\rho(\ldots)$ increases, the tradeoff between increased performance and computational complexity may no longer be justified. Similarly, the authors only deployed the model on small datasets (~1,200 securities). Given there are around hundreds of thousands of assets available to invest in worldwide, I am skeptical this technique can be employed on a large universe of securities.

Sam Bales
11/14/2019
MATH 624

References

Fastrich, B., Paterlini, S., & Winker, P. (2014). Constructing optimal sparse portfolios using regularization methods. *Computational Management Science*, *12*(3), 417–434. doi: 10.1007/s10287-014-0227-5