Diversity-weighted Portfolio in Stochastic Portfolio Optimization: A closer look at its application and efficacy

Chenghao Zhu, Haoyang Shang, Yuhan Wang, Yuxuan Zhou

Boston University, Questrom School of Business

December 11, 2023

Abstract

In this paper, we provided a comprehensive analysis of diversity-weighted portfolios within Stochastic Portfolio Theory (SPT). We examined the application and effectiveness of Bayesian non-parametric methods, particularly in addressing challenges inherent in SPT, like market imperfections and model limitations. The study compared Frequentist and Bayesian parametric inference methods in portfolio optimization and explored the impact of negative parameter values on portfolio returns. Empirical analysis included the evaluation of portfolio performance using data from the S&P 500 index, assessing the merits of diversity-weighted portfolios with both positive and negative parameters.

1 Introduction

Stochastic Portfolio Theory, an alternative approach to portfolio selection developed by Robert Fernholz in 2002, aims to outperform the market index with probability one. Fernholz uses a 'master equation', a pathwise decomposition of relative performance, to avoid challenges encountered in classical portfolio optimization methods, such as explicit model postulation and calibration, as well as the (normative) no-arbitrage assumption.

However, Stochastics Portfolio Theory remains imperfect for wider adoption due to several remaining problems and limitations. Three of the most notable issues are: the difficulty of finding relative arbitrages under reasonable assumptions (inverse problem), ignorance of several market imperfections such as the possibility of default and transaction cost, and lack of adoption by factors other than market capitalization when exploiting market inefficiency.

The study "Stochastic Portfolio Theory: A Machine Learning Perspective" by Yves-Laurent Kom Samo and Alexander Vervuurt, published in 2016, introduces a Bayesian non-parametric method to overcome challenges in Stochastic Portfolio Theory. The researchers explore a variety of investment strategies influenced by a function acting on a diverse set of trading characteristics, like market capitalization. This function is governed by a Gaussian process (GP) prior. They evaluate the likelihood of a strategy being 'exceptional' based on a performance metric chosen by the user (such as excess return compared to the market index or Sharpe ratio), and benchmarks that are considered 'exceptional'. The authors utilize Monte Carlo Markov Chain (MCMC) techniques to generate samples from the posterior of the GP that defines the 'exceptional' strategy.

In our project, we explored how authors used Bayesian Inference in Diversity-weighted Portfolios, to tackle the challenges of Stochastic Portfolio Theory. Moreover, we compared the performance of Frequentist parametric inference with that of the Bayesian parametric inference method when sampling posterior distribution. In the last part of the project, we also explored how a negative parameter p in a Diversity-weighted portfolio could drastically increase its return compared to a positive parameter p.

2 Diversity-weighted Portfolio

We provide a brief overview of SPT, outlining the broad category of market models that its findings apply to, the nature of the portfolio selection criterion, and the methods used to develop strategies that meet this requirement.

2.1 The Model of Stock Market

In SPT, the stock capitalizations are modelled as Itô processes. Namely, the dynamics of the n positive stock capitalization processes $X_i(\cdot)$, $i = 1, \ldots, n$

are described by the following system of SDEs

$$dX_i(t) = X_i(t)\left(b_i(t)dt + \sum_{\nu=1}^{d}\sigma_{i\nu}(t)dW_\nu(t)\right), \tag{1}$$

for $t \geq 0$ and $i = 1, \ldots, n$. Here, $W_1(\cdot), \ldots, W_d(\cdot)$ are independent standard Brownian motions with $d \geq n$, and $X_i(0) > 0$, $i = 1, \ldots, n$ are the initial capitalizations.

We assume all processes to be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and adapted to a filtration $\mathcal{F} = \{\mathcal{F}(t)\}_{0 \leq t < \infty}$.

The rates of return $b_i(\cdot)$, $i = 1, \ldots, n$ and volatilities $\sigma(\cdot) = (\sigma_{i\nu}(\cdot))_{1 \leq i \leq n, 1 \leq \nu \leq d}$, are some unspecified $\mathcal{F}$ progressively measurable processes and are assumed to satisfy the integrability condition

$$\sum_{i=1}^{n}\int_0^T \left(|b_i(t)| + \sum_{\nu=1}^{d}(\sigma_{i\nu}(t))^2\right) dt < \infty \tag{2}$$

for all $T \in (0, \infty)$, and the non-degeneracy condition

$$\exists \epsilon > 0 \text{ s.t. } \xi^\top \sigma(t)\sigma^\top(t)\xi \geq \epsilon\|\xi\|^2, \forall \xi \in \mathbb{R}^n, t \geq 0 \tag{3}$$

2.2 Relative Arbitrage

It is now necessary for us to understand how to use portfolios to build and assess a portfolio's performance within this model framework. These are $\mathbb{R}^n$-valued and $\mathcal{F}$-progressively measurable processes $\pi(\cdot) = (\pi_1(\cdot), \ldots, \pi_n(\cdot))^T$, where $\pi_i(t)$ stands for the proportion of wealth invested in stock i at time t.

Long-only portfolios is what we used and there is no money market. These invest solely in the stocks; they take values in the closure $\overline{\Delta_+^n}$ of the set

$$\Delta_n^+ = \{x \in \mathbb{R}^n : x_1 + \ldots + x_n = 1, 0 < x_i < 1, i = 1, \ldots, n\} \quad (4)$$

Without losing generality, let us assume that each firm has one outstanding share. The related wealth process $V^\pi(\cdot)$ of an investor executing $\pi(\cdot)$ can be observed to evolve in the following way.

$$\frac{dV_\pi(t)}{V_\pi(t)} = \sum_{i=1}^{n} \pi_i(t) \frac{dX_i(t)}{X_i(t)}, \quad V^\pi(0) = 1 \quad (5)$$

Performance is primarily measured in SPT in relation to the market index. This is the wealth process $V^\mu(\cdot)$ that results from a buy-and-hold portfolio, given by the vector process $\mu(\cdot) = (\mu_1(\cdot), \ldots, \mu_n(\cdot))^T$

$$\mu_i(t) \frac{X_i(t)}{\sum_{j=1}^{n} X_j(t)} \quad (6)$$

Let $T > 0$. A strong relative arbitrage with respect to the market over the time-horizon $[0, T]$ is a portfolio $\pi(\cdot)$ such that

$$P(V^\pi(T) > V^\mu(T)) = 1 \quad (7)$$

Any portfolio that beats the market in the context of (7) is considered a relative arbitrage in SPT, and the degree of the beat is theoretically meaningless.

2.3 Functionally-generated Portfolios

Now we will introduce a particular type of portfolios named functionally-generated portfolios studied by Fernholz [1999].

Consider a function $G \in C^2(U, \mathbb{R}_+)$, where U is an open neighborhood of Δ_n^+ and such that $x \mapsto x_i D_i \log G(x)$ is bounded on Δ_n^+ for $i = 1, \ldots, n$. Then G is said to be the generating function of the functionally-generated portfolio $\pi(\cdot)$, given, for $i = 1, \ldots, n$, by

$$\frac{\pi_i(t)}{\mu_i(t)} = \frac{D_i G(\mu(t))}{G(\mu(t))} + 1 - \sum_{j=1}^{n} \mu_j(t) \frac{D_j G(\mu(t))}{G(\mu(t))} \quad (7)$$

In this case, the partial derivative with regard to the variable i is written as D_i, and the second partial derivative with respect to the variables i and j is written as D_{ij}^2.

Theorem 3.1 of Fernholz [1999] asserts that the performance of the wealth process corresponding to $\pi(\cdot)$, when measured relative to the market, satisfies the $\mathbb{P}$-almost sure decomposition

$$\log\left(\frac{V_\pi(T)}{V_\mu(T)}\right) = \log\left(\frac{G(\mu(T))}{G(\mu(0))}\right) + \int_0^T g(t)\, dt \quad (8)$$

where the quantity

$$g(t) := -\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{D_{ij}^2 G(\mu(t))}{2G(\mu(t))} \mu_i(t) \mu_j(t) \tau_{ij}^{\mu}(t) \quad (9)$$

is called the drift process of the portfolio $\pi(\cdot)$. Here, we have written $\tau_{ij}^{\mu}(\cdot)$ for the relative covariances; denoting by e_i the i-th unit vector in $\mathbb{R}^n$, these are defined for $1 \leq i, j \leq n$ as

$$\tau_{\mu_{ij}}(t) := (\mu(t) - e_i)^T \sigma(t) \sigma(t)^T (\mu(t) - e_j) \quad (10)$$

If $T > 0$, the left side of master equation (8) may be constrained away from zero under appropriate conditions on the market model (1), demonstrating that $\pi(\cdot)$ represents an arbitrage with respect to the market over $[0, T]$.

2.4 Diversity-weighted Portfolio

One of the most-studied FGPs is the diversity-weighted portfolio (DWP) with parameter $p \in \mathbb{R}$ (defined in (4.4) of Fernholz et al. [2005]).

$$\pi_i^{(p)}(t)\frac{(\mu_i(t))^p}{\sum_{j=1}^n(\mu_j(t))^p}, \quad i = 1, \ldots, n., \quad (11)$$

This portfolio is a relative arbitrage with respect to $\mu(\cdot)$ over $[0, T]$ for any p and $T > \frac{2\log n}{\varepsilon\delta p}$, under the condition (ND^ε), and that of diversity (D^δ), introduced below.

$$\exists\delta \in (0,1)\colon P\left(\max_{\{1\leq i\leq n, t\in[0,T]\}} \mu_i(t) < 1-\delta\right) = 1 \quad (D^\delta)$$

3 Positive and Negative Parameter P

In Vervuurt and Karatzas [2015], authors found that Diversity-weighted Portfolios with negative parameter p was shown to outperform the market over sufficiently long time horizons and under suitable market assumptions.

First, we will illustrate that under ND^ε and no-failure (NF) condition, the diversity-weighted portfolio $\pi_i^{(p)}(\cdot)$ with parameter

$$p \in \left(\frac{\log n}{\log(n\varphi)}, 0\right)$$

is a strong arbitrage relative to the market $\mu(\cdot)$ for $[0, T]$, for any number

$$T > \frac{-2n\log(n\varphi)}{\varepsilon(1-p)(n-(n\varphi)^p)}$$

$$\exists\varphi \in (0, \frac{1}{n}) \text{ s.t. } P(\mu_{(n)}(t) > \varphi, \forall t \in [0,T]) = 1 \quad (NF)$$

To prove it, given the fact that the portfolio (11) is generated by the function

$$G_p : x \mapsto \left(\sum_{i=1}^{n} x_i^p \right)^{1/p}, \quad (12)$$

Under the condition that $\sum_{i=1}^{n} x_i$, we apply Lagrange multipliers method to maximize G_p for $p < 0$, which comes up with $x_i = \frac{1}{n}$, we can then derive the upper and lower bounds for G_p,

$$n^{(1-p)} = \sum_{i=1}^{n} \left(\frac{1}{n} \right)^p \leq \sum_{i=1}^{n} (\mu_i(t))^p = (G_p(\mu(t)))^p < \sum_{i=1}^{n} \varphi^p = n\varphi^p, \quad (13)$$

Therefore, with negative p, we can say that

$$\log \left(\frac{G_p(\mu(T))}{G_p(\mu(0))} \right) \geq \log(n\varphi), \quad (14)$$

And for $\pi_{(1)}^{(p)}(t) = \max(\pi^{(p)}(t))$ and $\mu_{(n)}(t) = \min(\mu(t))$,

$$\pi_{(1)}^{(p)}(t) = \frac{(\mu_{(n)}(t))^p}{\sum_{i=1}^{n} (\mu_i(t))^p} < \frac{\varphi^p}{n^{(1-p)}} = \frac{(n\varphi)^p}{p} < 1, \quad (15)$$

Then for the excess growth rate $\gamma_\pi^*(t) = \frac{1}{2} \left(\sum_{i=1}^{n} \pi_i(t) a_{ii}(t) - \sum_{i,j=1}^{n} \pi_i(t) a_{ij}(t) \pi_j(t) \right)$, where $a(\cdot) = \sigma(\cdot)\sigma^T(\cdot)$, combined with ND^ε, we can get that for long-only portfolios,

$$\gamma_\pi^*(t) \geq \frac{\varepsilon}{2}(1 - \pi_{(1)}(t)), \forall t \geq 0, \quad (16)$$

In conjunction with (15), equation (16) can be turned into

$$\int_0^T \gamma_{\pi^{(p)}}^*(t)\, dt \geq \frac{\varepsilon}{2} \int_0^T (1 - \pi_{(1)}^{(p)}(t))\, dt \geq \frac{\varepsilon}{2} T \left(1 - \frac{(n\varphi)^p}{n} \right), \quad (17)$$

Finally, combining all these together, we will obtain the relative performance as

$$\begin{aligned} \log \left(\frac{V^{\pi^{(p)}}(T)}{V^\mu(T)} \right) &= \log \left(\frac{G_p(\mu(T))}{G_p(\mu(0))} \right) \\ + (1-p) \int_0^T \gamma_{\pi^{(p)}}^*(t)\, dt &> \log(n\varphi) + (1-p)\frac{\varepsilon}{2} T \left(1 - \frac{(n\varphi)^p}{n} \right) > 0, \quad (18) \end{aligned} \quad (4)$$

Then we move on to the comparison between negative and positive parameters. Specifically, the diversity-weighted portfolio with $\pi^{(p^-)}(\cdot)$ with negative parameter

$$p^- \in \left(\frac{\log n}{\log(n\varphi)}, 0\right)$$

is then a strong arbitrage relative to the diversity-weighted portfolio $\pi^{(p^+)}(\cdot)$ with positive parameter

$$p^+ \in \left(\max\left\{0, 1 - \frac{\varepsilon(n - (n\varphi)^{p^-})(1-p^-)}{4K(n-1)}\right\}, 1\right)$$

Over any horizon $[0, T]$ of length

$$T > \frac{-2\log(n\varphi)}{C}$$

Here the positive constant C is defined as

$$C := \frac{\varepsilon}{2}\left(1 - \frac{(n\varphi)^{p^-}}{n}\right)(1-p^-) - \frac{2K}{n(n-1)}(1-p^+)$$

The proof is as follows. We first introduce a similar condition as ND^ε, which is called the bounded variance assumption:

$$\exists K > 0 \text{ s.t. } \xi^T\sigma(t)\sigma^T(t)\xi \le K\|\xi\|^2, \forall \xi \in \mathbb{R}^n, t \ge 0, \quad (BV)$$

$$\gamma_{\pi^*}(t) \le 2K(1 - \pi_{(1)}(t)), \forall t \ge 0, \quad (19)$$

To make this easier to understand, we write $\pi^{\pm}(\cdot)$ and $G_{\pm}$ for $\pi^{(p^{\pm})}(\cdot)$ and $G_{p^{\pm}}$, respectively. Note that for $p^+ > 0$ there is

$$n\varphi^{p^+} < (G_+(\mu(t)))^{p^+} \le n^{(1-p^+)} \quad (20)$$

Which gives the lower bound for $p = p^+$. Use observation $\pi^+_{(1)}(t) \ge \frac{1}{n}$, we get that

$$\int_0^T \gamma^*_{\pi^+}(t)\, dt \le 2K \int_0^T (1 - \pi^+_{(1)}(t))\, dt \le 2KT\left(1 - \frac{1}{n}\right) \quad (21)$$

Hence, we see that by the virtue of the two equations the upper bound is then

$$\log\left(\frac{V^{\pi^+}(T)}{V^{\mu}(T)}\right) < -\log(n\varphi) + 2(1-p^+)KT\left(1-\frac{1}{n}\right) \quad (22)$$

Combining this with equation (18) above, we find that

$$\log\left(\frac{V^{\pi^-}(T)}{V^{\pi^+}(T)}\right) = \log\left(\frac{V^{\pi^-}(T)}{V^{\mu}(T)}\right) - \log\left(\frac{V^{\pi^+}(T)}{V^{\mu}(T)}\right)$$

$$> 2\log(n\varphi) + CT > 0 \quad (23)$$

given that

$$CT > -2\log(n\varphi) > 0 \quad (24)$$

An easy calculation shows that $C > 0$, whereas the last inequality in (24) comes from $\varphi < \frac{1}{n}$. Therefore in this case $\pi^-(\cdot)$ outperforms $\pi^+(\cdot)$ strongly over the time-horizon $[0, T]$.

4 Model Specification

4.1 Definition

This part is focused on defining a model for long-only portfolios based on a set of trading characteristics $X \subset \mathbb{R}^d$ for some $d \geq 1$. The portfolio weights are determined by a continuous function $f : X \to \mathbb{R}^+$:

$$\pi_i^f(t) = \frac{f(x_i(t))}{\sum_{j=1}^n f(x_j(t))}, \quad i = 1, \ldots, n, \quad (25)$$

Numerous elements, including market-to-book value, credit ratings, industry momentum, firm size, and balance sheet variables, might influence the choice of trading characteristics. Functionally-generated portfolios, such as equally-weighted, entropy-, and diversity-generated portfolios, are included in the model as special examples.

The key to this model is the investment map f, which determines how trading opportunities, indicated by evolving trading characteristics $x_i(t)$, are sized. Whence, learning an investment strategy in our framework is equivalent to learning an investment map f. To do so, we consider two families of functions:

Parametric Approach: Where f takes a simple parametric form:

$f : \mu \mapsto \mu^p, \quad$ for $p \in \mathbb{R}$ which corresponds to the diversity-weighted portfolio

Non-Parametric Approach: Where $\log f$ is a path of a mean-zero Gaussian process with a covariance function k:

$$\log f \sim \mathrm{GP}(0, k(\cdot, \cdot)).$$

We must provide an optimality criterion that encodes the user's investment purpose in order to learn "good" investment maps. To do so, we consider a performance functional P_D that maps the logarithm of a candidate investment map to the historical performance $P_D(\log f)$ of the portfolio $\pi^{f(\cdot)}$ as in Eq. (25) over some finite time horizon, given historical data D. An example of a performance functional is the excess return relative to a benchmark portfolio π^*.

$$ER(\pi^f|\pi^*) = \prod_{t=1}^{T}(1 + \sum_{i=1}^{n} r_i(t)\pi_i^f(t)) - \prod_{t=1}^{T}(1 + \sum_{i=1}^{n} r_i(t)\pi_i^*(t)), \quad (26)$$

Another example performance functional is the annualized Sharpe Ratio, defined as

$$"SR"(\pi) = \left(\sqrt{B}\right)\left(\frac{\hat{E}(\{r(1),\ldots,r(T)\})}{\hat{S}(\{r(1),\ldots,r(T)\})}\right), \quad (27)$$

4.2 Inference

For the remainder of this article, the total excess return, adjusted for transaction costs, will be used as the performance functional in relation to the equally weighted portfolio (EWP), which has constant weights.

$$\pi_i(t) = \frac{1}{n}, \quad i = 1,\ldots,n, \quad \forall t \geq 0$$

over the whole training period

$$P_D(\log f) = \text{"ER"}(\pi_f \,|\, \text{"EWP"}), \quad (28)$$

When needed, we use as likelihood model

$$L(\text{"PD"}(\log f)) = \gamma(\text{"PD"}(\log f); a, b), \quad (29)$$

The probability density function of the Gamma distribution with mean a and standard deviation b is represented by the symbol $\gamma(\cdot; a, b)$. As previously discussed, a and b need not be learned as they reflect the investment manager's risk appetite. In the tests that follow, $a = 7.0$ and $b = 0.5$ are used. Stated differently, we conjecture that the ideal investment plan ought to be planned such that, given an initial unit of wealth, the terminal wealth at the conclusion of the training period will, on average, exceed by 7.0 units the terminal wealth that the equally weighted portfolio will attain at the end of the same trading tenure.

Frequentist parametric: Maximizing $P_D(\log f)$ for $p \in [-10, 10]$ yields the optimal parameter of the DWP, which is the first inference method we explore. In contrast, the SPT literature often considers p to be in the range of $[-1, 1]$. On the uniform grid with mesh size 0.05^2, we use brute force maximizing to prevent any issues with local maxima.

Bayesian parametric: The second inference strategy we examine involves sampling from the posterior distribution over the exponent p in the DWP scenario using the Metropolis-Hastings algorithm (Hastings [1970]).

$$p(p|D) \propto L(P_D(p)) \times \mathbf{1}_{\{p \in [-10,10]\}}, \quad (30)$$

where we have rewritten $L(P_D(\log f))$ as $L(P_D(p))$ to make the dependency on p explicit. We sample a proposal update p^* from a Gaussian centered at the current exponent p and with standard deviation 0.5. The acceptance probability is easily found to be

$$r = \min\left(1, \frac{L(P_D(p^*))}{L(P_D(p))} \times \mathbf{1}_{\{p^* \in [-10,10]\}}\right), \quad (31)$$

Specifically, we note that as long as p is started within $[-10, 10]$, the indicator function in Eq. (30) won't interfere with the Markov chain. Once 10,000 MH repeats have been completed, the first 5,000 are often eliminated as "burn-in." After the proper DWP, we use the posterior mean exponent that we learned from training data to swap in our testing horizon.

$$\hat{f}(\mu) = \mu^{\mathbb{E}(p|D)}, \quad (32)$$

5 Empirical Findings

The stocks we analyze in our experiments consist of the constituents of the S&P 500 index, accounting for changes in index components. We adjust our portfolios on a monthly basis. Our data is sourced from Yahoo Finance, covering the period from January 2008 to December 2022.

We explore optimal investment strategies, as outlined in the prior section, utilizing five consecutive years of data for training purposes and five subsequent years for testing. Our initial data set starts in January 2008 for

Portfolio	IS Return(%)	OOS Return(%)	IS SR	OOS SR
Market	13.1686	13.279	0.3169	0.3043
EWP	20.1405	16.9966	1.4615	1.236
Positive DWP	17.3811	14.8478	1.0947	0.8915
Negative DWP	53.7388	53.6636	1.1967	0.9483
Positive MCMC DWP	19.5541	15.567	1.3059	1.1809
NEgative MCMC DWP	46.7826	56.8979	1.2454	1.0223

Figure 1: Average performance for the portfolios

training, followed by a sequential shift of both training and testing datasets by one year. This process results in six pairs of training and testing subsets. We assess and compare the performance of these portfolios, including the equally-weighted portfolio, the market portfolio, and four diversity-weighted portfolios. In one diversity-weighted portfolio, the exponent p is determined by maximizing the evaluation functional, while in the other, p is learned using MCMC, with both positive and negative one.

Figure 1 summarizes the average in-sample and out-of-sample return and Sharpe ratio for our 6 portfolios during 6 different time periods. We can see that diversity-weighted portfolios with negative p indeed outperform those with positive one in annual return no matter whether MCMC is applied, while the DWP with positive p cannot even beat the equally weighted portfolio. However, when it comes to Sharpe ratio, the equally weighted portfolio surprisingly becomes the best one, which may be due to the fact that diversity-weighted portfolios are poor at controlling volatility. Figure 1 shows the exact wealth process for these 6 portfolios with an in-sample-period of 2009-2013 and out-of-sample-period of 2014-2018, which can help better illustrate our findings above.

Besides, based on our research, we can also see that the diversity-weighted-portfolio coming up by MCMC can slightly outperform that using simple maximization, with a better effort at risk management.

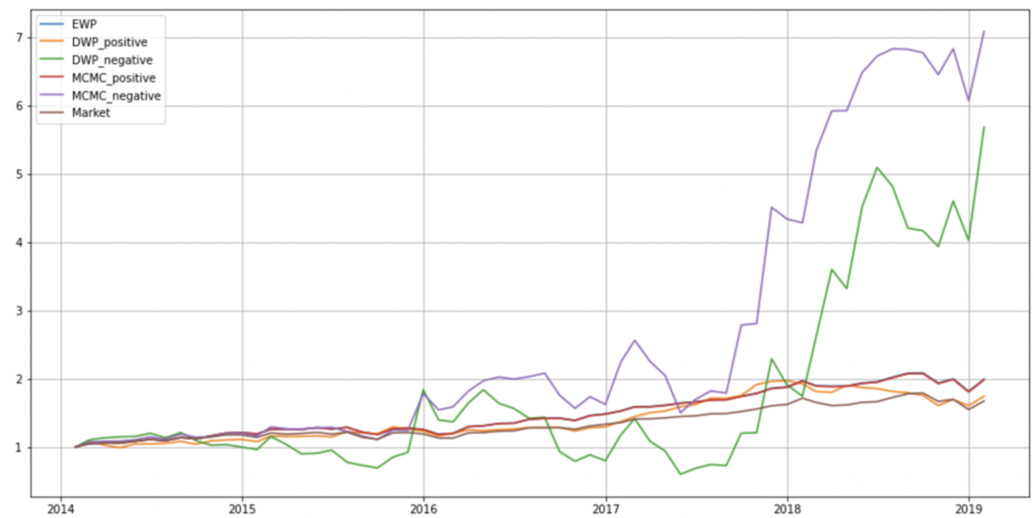

Figure 2: Exact wealth process for 2009-2018

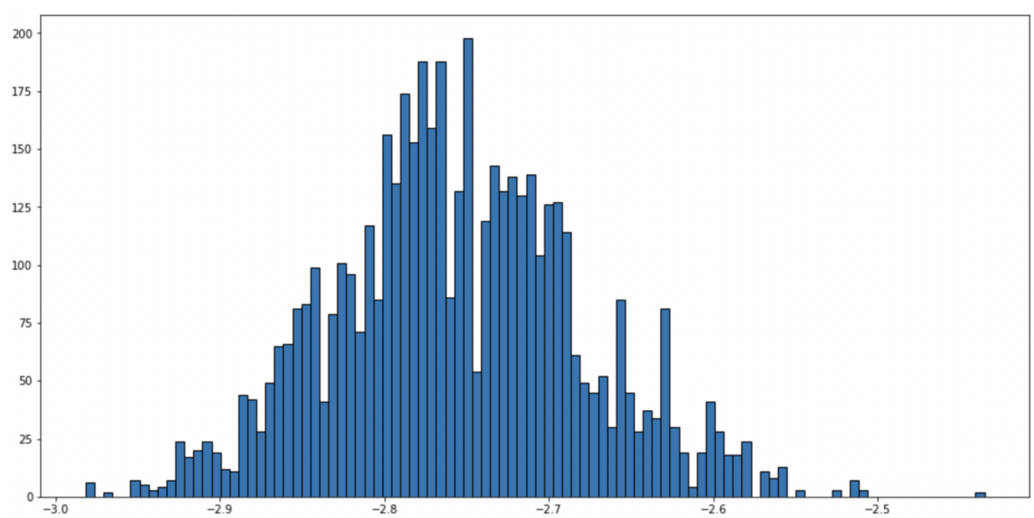

Figure 3: Posterior distribution over the exponent p

An example of the posterior distribution over the exponent p in the Bayesian parametric method is also illustrated in Figure 3, which suggests that p can be highly concentrated after 5000 tries.

6 Conclusion

The empirical study shows that, Diversity-weighted portfolios with negative p indeed outperforms those with positive parameters, since having superior returns. However, Diversity-weighted-portfolio may be poor at risk-management, as the extra returns did not result in a better Sharpe ratio, suggesting increased volatility. Overall, Diversity-weighted-portfolio coming up by MCMC can slightly outperform that using simple maximization, although by a small margin.

References

[1] Samo, Yves-Laurent Kom, and Alexander Vervuurt. "Stochastic Portfolio Theory: A Machine Learning Perspective." *arXiv.Org*, 9 May 2016, arxiv.org/abs/1605.02654.

[2] Vervuurt, Alexander, and Ioannis Karatzas. "Diversity-Weighted Portfolios with Negative Parameter." *arXiv.Org*, 1 July 2015, arxiv.org/abs/1504.01026.