Portfolio Optimization in the Presence of Extreme Risks: A Pareto-Dirichlet Approach

Olivier Le Courtois *

Xia Xu †

January 11, 2019

Abstract

Investor's aversion to extreme risk can affect portfolio choice. We use kurtosis as extreme risk indicator, and construct a mean variance kurtosis portfolio optimization. Our approach combines Pareto improvement method and Dirichlet simulations. We show that previous high-order moment portfolio optimization methods can mis-classify inefficient portfolios as efficient, and we propose the Pareto improvement method to detect efficient set. For its implementation, we use Dirichlet simulations to approximate the feasible portfolio set. Sharpe ratio is generalized by synthesizing variance and kurtosis into a single risk indicator. We show that mean variance optimization is a sub-case of mean variance kurtosis optimization, and the mean variance kurtosis efficient portfolios are always better than the mean variance efficient portfolios in terms of generalized Sharpe ratio.

EFM Code: 370, 380

Keywords: Portfolio Choice, Efficient Frontier, Extreme Risk, Dirichlet Distribution

*O. Le Courtois is at emlyon business school. lecourtois@em-lyon.com. Address: 23, Avenue Guy de Collongue, 69134 Ecully Cedex, France. Phone: 33-(0)4-78-33-77-49. Fax: 33-(0)4-78-33-79-28.

†Xia Xu is at emlyon business school. xu@em-lyon.com. Address: 23, Avenue Guy de Collongue, 69134 Ecully Cedex, France.

The last decades have witnessed several economic disasters, among which the global financial crisis and the European sovereign debt crisis are the two most dramatic ones in recent memory. This event risk concern can be consequential to investor behavior and portfolio development. Some researchers like Acharya and Pedersen (2005) try to account for this concern by incorporating liquidity risk as an additional systematic risk factor, so economic crises are interpreted as liquidity crises. The key drawback for this additional systematic risk factor approach is that it has no suggestion on portfolio construction. Other researchers, like de Athayde and Flôres (2004) and Briec, Kerstens, and Jokung (2007) develop higher order moment portfolio optimization primarily in the mean variance skewness space. However, the existing higher order optimization methods can misclassify inefficient portfolios as efficient. Besides, neither of the additional risk factor approach nor the higher order moment portfolio optimization approach is specifically dedicated to extreme risk in the analysis.

In this paper, we capture the extreme risk by use of kurtosis, and consider a novel portfolio selection problem based on the mean, variance, and kurtosis of the asset return distribution.[1] Efficient portfolios are specified by Pareto improvement method in the spirit of the Pareto efficiency of neoclassical economics. Its intuition is to recognize a portfolio with some return moments worse off and non moments better off as inefficient. It is a generalized and simple algorithm for inefficient portfolio detection and is free from the aforementioned misclassification error by the existing higher order portfolio optimization methods. This intuitive Pareto improvement method is fueled by Dirichlet simulations for feasible portfolio set approximation, which is another innovation of this paper. The Dirichlet distribution generates multivariate vectors with non negative values that sum to 1, which is consistent with the non short selling portfolio formation as highlighted by Jagannathan and Ma (2003). The mean variance efficient portfolios are all mean variance kurtosis efficient, and the mean variance inefficient portfolios can be mean variance kurtosis efficient. The inferences on the mean variance and mean variance kurtosis efficient sets are helpful to parameterize the Dirichlet simulations for better efficacy. Given the mean variance kurtosis efficient set, we propose a generalized Sharpe ratio to synthesize variance and kurtosis into a single risk indicator for subsequent optimal portfolio choice. When variance is more weighted in the generalized Sharpe ratio, the optimal portfolios are always mean variance efficient. When kurtosis is more weighted, the optimal portfolios are often mean variance inefficient. Therefore, the mean variance kurtosis optimization is more comprehensive than the mean variance optimization, as kurtosis can give different risk information from variance.

[1] When the terms of skewness and kurtosis go with variance in the context, usually they refer to the third moment and the fourth moment for convenience. This notation is commonly used in papers of portfolio optimization with higher order moments. For example, in Briec, Kerstens, and Jokung (2007, p. 138) skewness is defined as the third moment explicitly. In Jurczenko, Maillet, and Merlin (2006) and other papers, the same notation is used.

In modern finance, risk is primarily defined as variance. Remarkably, Markowitz (1952) interprets risk as return variance and investors goal is to minimize portfolio variance at a given level of expected portfolio return. However, the effect of higher order moments on asset pricing and return determination gets more attention, especially after the 2008 global financial crisis. Barro (2006) explores the effect of three rare economic disasters, namely two World Wars and the Great Depression, on asset markets, and calibrates disaster probability to account for asset pricing puzzles like high equity premium. Cvitanić, Polimenis, and Zapatero (2008) warn that ignoring the effect of higher moments leads to wealth loss by overinvesting in risky assets. Bates (2012) inspects US stock market crash risk since 1926 to 2010, and features the importance of fat tail property in the explanation of market crash. Santa-Clara and Yan (2010), Benzoni, Collin-Dufresne, and Goldstein (2011), Bollerslev and Todorov (2011), Gabaix (2012), Drechsler (2013), Wachter (2013) and others study general equilibrium models subject to tail events, or rare jump shocks, in attempt to explain market anomalies like the equity premium puzzle. Extreme risk has become an important concern for asset pricing models.

Given the importance of extreme risk, some studies try to augment the mean variance portfolio optimization by incorporating the effect of higher order moments. Constrained variance minimization method and shortage function method are the two most representative higher order portfolio optimization methods. The constrained variance minimization method, as used in de Athayde and Flôres (2004) and Mencía and Sentana (2009), is to find the portfolio with the minimized variance given the constraint that the mean return and skewness are no less than a certain level. The shortage function method looks for the portfolio with the best return profile improvement available in the feasible set, relative to a certain evaluated portfolio. Briec, Kerstens, and Jokung (2007), Briec and Kerstens (2010), and Kerstens, Mounir, and Van de Woestyne (2011) give instructive descriptions of how to detect mean variance skewness efficient portfolios using the shortage function method. If no moment profile improvement exists, the evaluated portfolio is efficient; otherwise, the portfolio achieving the largest return profile improvement is label as efficient relative to the evaluated.[2]However, for marginal return improvement where not all of the return moments are improved, the two methods can misclassify inefficient portfolios as efficient. This mis-identification impacts the efficient set composition.

Kurtosis is appropriate to measure the tail thickness of return distribution. It is good at reflecting extreme risk, but a paucity of literature focuses on kurtosis in portfolio selection

[2]The Bayesian approach for higher order portfolio optimization by Harvey, Liechty, Liechty, and Müller (2010) discuss more about resampling technique to reduce estimation error rather than efficient frontier construction for higher order optimization. Their approach is based on a grid of portfolio weights between 0 and 1, which can be cumbersome and unimplementable when the underlying asset size is large.

without divergent consideration of skewness. To our est knowledge, this paper proposes a first portfolio selection method with an explicit emphasis on kurtosis, taking advantage of the Pareto improvement method to minimize the misclassification error.

Note that our focus on kurtosis does not implicate that skewness is out of investment consideration. Rather, the first extension of mean variance approach to higher order moment is mean variance skewness portfolio optimization, highlighting investors' preference for positive skewness. Compared to skewness, kurtosis can better spotlight the effect of extreme events, not only negative but also positive. It is more consistent with the general perception of risk, which is the lower the better. For this sake, we study mean variance kurtosis optimization to provide a complementary perspective for high order portfolio choice.

We contribute to the portfolio simulation literature as well. Most discussions on sampling methods are about alleviating the problem of estimation errors and associated high optimization sensitivity, like the "resampled efficient frontier" by Michaud and Michaud (2008) in the mean variance framework. The sampling method can also be used for random portfolio formation. There are many papers on random portfolio construction from a certain security pool. For example, Jagannathan and Ma (2003) use random samples of 500 stocks to examine weight constraint on the mean variance optimization. Note that random portfolios come with equal weights for portfolio formation. For example, Porter and Bey (1974) generate random equal-weighted portfolios at various sizes to form second order stochastic dominance efficient sets, and then compare them with the mean variance efficient sets. Levy and Kaplanski (2015) similarly use random samples to examine stochastic dominance and mean variance efficient sets. In fact, equal-weighted scheme is not representative enough for general portfolio development by less diversification benefits, we innovatively use Dirichlet distribution to generate portfolio weights for its innate fitness. Chotikapanich and Griffiths (2002) initially use Dirichlet distribution to sample income shares for the Lorenz curve estimation. This is an inspiration for the use of Dirichlet distribution for portfolio selection, given the similarity of income shares and portfolio weights. More than that, Dirichlet simulations make the Pareto improvement method implementable, and the combination of the two programs offers a simple and straightforward approach for the higher moments efficient frontier construction. The inferences about the relationship between mean variance and mean variance kurtosis efficient sets facilitate the efficient portfolio identification and increases the efficacy of Dirichlet efficient frontier.

The rest of this paper is arranged as follows. Section 1 examines the importance of kurtosis to capture extreme risk for investor's portfolio choice. Section 2 presents the Pareto improvement method for efficient portfolio detection, as well as its advantages to the existing higher order portfolio optimization methods. Section 3 introduces Dirichlet distribution for portfolio simulation, and Section 4 further discusses how to enhance the efficacy of Dirichlet efficient frontier. Section 5 proposes an empirical study of the ten economic sectors of the S&P 500 index, and the Sharpe ratio is generalized to take optimal choice from the Dirichlet efficient portfolio set. Section 6 concludes the paper.

1 Extreme risk and kurtosis

Suppose there are N risky assets in the financial market, and an investor constructs portfolio by deciding on the weight vector $w = [w_1\ w_2\ \ldots\ w_N]'$ where w_i is the portfolio weight allocated to asset i with $i \in [1, 2, \ldots, N]$. For a regular portfolio, it is guaranteed that $\Sigma_{i=1}^N w_i = 1$, plus $w_i \geqslant 0$ with no short selling. And, $r_i = [r_{i1}\ r_{i2}\ \ldots\ r_{iT}]'$ is the return series of asset i, r_{it} its return observation at time t, $E[r_i]$ its expected return. Then the portfolio return series and its expected return are $R = [r_1\ r_2\ \ldots\ r_N]w$ and $E[R] = \big[E[r_1]\ E[r_2]\ \ldots\ E[r_N]\big]w$.

We calculate the central moment at order k as

$$M_k = E\big[R - E[R]\big]^k.$$

Ranging k from 2 to 4, we get variance (σ^2), skewness (s^3), and kurtosis (κ^4). Apparently, extreme observations of R have more significant effect on higher order moments, because the compounding index is bigger. This indicates the advantage of higher order moment in capturing the impact of extreme returns.

Kurtosis is frequently used to capture extreme risks, which have come into the spotlight after recent series of financial crises and economic disasters. For example, Barro (2006) uses kurtosis to reflect disaster risk and shows that the economic disasters of last century led to a 29% average decline in real per capita GDP. He finds that disaster risk has considerable effects on asset returns and on the equity premium. Dittmar (2002), Eraker, Johannes, and Polson (2003), Liu, Longstaff, and Pan (2003), Poon, Rockinger, and Tawn (2003), Bakshi and Madan (2006), Bates (2008), Todorov (2009), Benzoni, Collin-Dufresne, and Goldstein (2011), Bollerslev and Todorov (2011), Bates (2012), Gabaix (2012), Drechsler (2013), Wachter (2013), Cremers, Halling, and Weinbaum (2015), among many, notice the significant role of extreme risks in the determination of returns. More interestingly, Malmendier and Nagel (2011) argue that low stock returns contribute to shaping individual preferences towards greater risk aversion, and this effect is stronger for young people. Thus, the effect of extreme risks on risk preference and return determination is not only contemporaneous but also prospective.

Kurtosis is a complementary risk indicator to variance for three reasons. Firstly, risk encompasses not only downside risk but also upside risk, as Damodaran (2003, Chapter 2) advocates. In this sense, kurtosis is more appropriate than skewness at depicting event risks, because it assigns the same penalties to extreme events, no matter positive or negative. Secondly, extreme negative events are more significant than extreme positive events, contributing to a deflated skewness benefit and an escalated dispersion apprehension. Cont (2001) obtains that the large upward movements are not as equally as the large drawdowns. This loss asymmetry implies that any potential positive extreme values are likely offset by negative extreme values in similar or greater size. When comparing two distributions with negative skewness, the skewness difference is not as informative as the kurtosis difference to infer the general dispersion level of the distributions. Thirdly, it is also well discussed about investor's preference for odd moments and aversion to even moments, like by Jondeau and Rockinger (2006). Kurtosis better depicts the investor's psychology of risk as dis-utility, and it is also consistent with the traditional treatment of variance as a risk indicator. Therefore, we extend the risk definition beyond variance, taking kurtosis as the complementary risk measure for its better characterization of extreme risks.

Table 1: The variation of skewness and kurtosis across global markets

	US	Europe	Japan	Asia Pacific	Globe
CV(Skewness)	2.11	1.43	2.49	2.22	0.97
CV(Kurtosis)	0.44	0.32	0.21	0.29	0.32

Note: Market returns are monthly data from Fama French data library with their best availability up to Nov 2018. Skewness and kurtosis are calculated with a 3-year moving window. The coefficient of variation (CV) is the absolute ratio of the standard deviation to the mean. The Asia Pacific market does not include Japan.

Besides, kurtosis estimates involve less variability than skewness estimates. Assembling the skewness and kurtosis estimates in a 3-year moving window of monthly returns on the Fama French market factor, we see that the coefficient of variation for skewness is 2.11 in US market, much higher than that of kurtosis, which is 0.44. It is similar in the European market, Japanese market, Asia Pacific ex Japan market, as well as the global market. To sum up, kurtosis estimates are more stable than skenwess, which is helpful to ameliorate the input sensitivity problem of portfolio optimization.

We highlight the necessity of including kurtosis into portfolio selection to investors who are averse to extreme risk with a simple example in Table 2.

Suppose we have 3 portfolios, A, B and C, and each of them has return realizations at corresponding state detailed the the table. For simplicity the return distributions are symmetric, so the skewness are uniformly 0. Note that Portfolios A and B have the same mean and variance, so mean variance optimization can not tell which one is better. Kurtosis gives critical additional distribution information about extreme observations. Portfolio A

Table 2: Tail return and moments: A portfolio choice example

	Portfolio A		Portfolio B		Portfolio C	
State	Return	Prob	Return	Prob	Return	Prob
1	0	0.02	5-$\sqrt{3/2}$	1/3	3	0.5
2	5	0.96	5	1/3	7	0.5
3	10	0.02	5+$\sqrt{3/2}$	1/3		
Mean	5		5		5	
Variance	1		1		4	
Kurtosis	25		1.5		16	
$U(R) = 1 - e^{-R}$	0.97353		0.98945		0.97465	

has much bigger kurtosis, so investors who are averse to risk including extreme risk prefer Portfolio B. Suppose their utility function is $U(R) = 1 - e^{-R}$, then even the mean return for Portfolio B decreases to 4.5, its expected utility is still bigger than that of Portfolio A. In this way, the investors trade off mean and kurtosis, and they may well accept a slight mean return reduction for better kurtosis exposure. This kind of choices is also consistent with Tversky and Kahneman (1986) insights, where loss aversion is an important property of value function and extreme negative losses are overweighed.

Brockett and Kahane (1992) illustrate that variance can not fully reflect risk aversion with an example similar to our Portfolios A and C.[3] Specifically, the two portfolios have the same mean return, and Portfolio A has smaller variance. So Portfolio C is out of the mean variance efficient portfolio set, but in fact this portfolio is more preferable to those investors with utility functions like $U(R) = 1 - e^{-R}$. The Brockett and Kahane (1992) criticism can be ameliorated by including kurtosis as complementary risk indicator to variance. Portfolio A has lower variance and higher kurtosis, while Portfolio C has higher variance and lower kurtosis, so no mean-risk dominance is between them. Therefore, the mean variance kurtosis optimization is superior to the mean variance optimization, and the inclusion of taking kurtosis into portfolio selection improves optimal choice.

There are several reasons why an investor is repugnant to extreme returns, especially extreme negative returns. Note that extreme positive returns are negative events to those investors who invest in short strategies. The following reasons are some explanations to the extreme risk aversion.

First, the financing burden transmitted by mark-to-market system. Individual investors especially those using futures and other derivatives to trade the underlying assets have to meet the margin call, which is often triggered by an extreme return. If the return is not so extreme, the investors probably do not have to make adjustments since the margin surplus

[3] We set ϵ of Brockett and Kahane (1992, Example 1) to be 1 to get our Portfolio C.

may well offset the loss. Also, institutional investors especially mutual funds and exchange-traded funds are more sensitive to such big price changes, since notable fund outflows and share redemptions may force them to realize losses and restrain their maneuver for subsequent trading. Even for hedge funds, extreme returns also have great impact to managers due to the requirements of hurdle rate and high water mark. The failure of LTCM is an exhibition that how wild asset price movements can be destructive.

Second, the danger of evolving into fire sales. When an asset is experiencing extreme negative return, it has prospective downward price pressure as investors holding this asset can go into panic and compete for sales. Therefore, the divergence of market price and intrinsic value can be wider, and finally the competition leads to fire sales for this asset. Not only individual investors can get involved in this vicious circle, but also intuitional investors like mutual funds and hedge funds for their distressed selling, as explained by Coval and Stafford (2007) and Stein (2009).

Third, financial contagion across assets. The effect of extreme returns is not restricted to one specific asset, and it can spillover to other assets, asset classes, and financial markets. Canaries in the coal mine, the sharp price decline excites the investor psychology, which can be contagious due to the rising cross-sectional correlation and financial globalization. This systematic correlation is enhanced in bear markets for extreme events, see Longin and Solnik (2001) and Bae, Karolyi, and Stulz (2003).

In short, we maintain that variance does not capture extreme risk well and kurtosis is its companion. Kurtosis can provide different risk information, and the combination of variance and kurtosis better characterizes risk. The inclusion of kurtosis into portfolio optimization can substantially improve the portfolio choice at the presence of extreme risk, as illustrated by the example in Table 2.

2 Pareto improvement method

We have highlighted the advantage of including kurtosis into portfolio optimization in previous section. To exted the mean variance portfolio optimization to higher order moments, there are two representative methods in literature: the constrained variance minimization approach and the shortage function approach. We will show the incompetency of the two approaches in detecting marginal profile improvements for higher moment efficient portfolio identification. We then introduce a more general and straightforward approach – the Pareto improvement method. Illustrating with an example, we affirm the merit of Pareto improvement method as well as the potential misclassification of inefficient portfolio as efficient by the other two methods. This example also reveals the key relationship of efficient sets in mean variance space and in higher order moment space, which gives a clue about the composition of the higher order moment efficient set.

2.1 Existing high order optimization methods

For each portfolio weight vector w, we compute its corresponding mean, variance and kurtosis, then stack them as the moment profile

$$\mathcal{P} = [\mu\ \sigma^2\ \kappa^4]'.$$

The classic optimization problem for the mean variance analysis is

$$\begin{aligned} &\text{minimize} && \sigma^2 \\ &\text{subject to} && \mu \geqslant \hat{\mu} \end{aligned}\ , \tag{P1}$$

where μ and σ^2 are elements in the portfolio's moment profile, and $\hat{\mu}$ a specified expected return level. The optimal portfolio by P1 has the minimum variance under the condition that its expected return is no less than the prespecified $\hat{\mu}$. This is a constrained optimization problem in the two dimensional space of mean and variance, and its natural extension to the three dimensional space of mean, variance, and kurtosis is

$$\begin{aligned} &\text{minimize} && \kappa^4 \\ &\text{subject to} && \mu \geqslant \hat{\mu} \\ & && \sigma^2 \leqslant \hat{\sigma}^2 \end{aligned}\ , \tag{P2}$$

where the newly added $\hat{\sigma}^2$ is a specified variance level. The optimal solution of P2 has the minimum kurtosis with the constraints that the mean is no less than $\hat{\mu}$ and the variance is no bigger than $\hat{\sigma}^2$. For high order portfolio optimization, it is very difficult to get analytical solutions like mean variance optimization case for involving tensor product for higher order moments calculation. Besides, the classic optimization can fail to detect marginal return profile improvement, thus misclassify inefficient portfolio as efficient. We will demonstrate this point in next subsection.

The constrained variance minimization method pioneered by de Athayde and Flôres (2004) for higher order moment is essentially P2. Since de Athayde and Flôres (2004) discuss efficient portfolios in the mean variance skewness space, the minimization objective is variance with constraints that the mean is no less than $\hat{\mu}$ and the skewness is no less than a specified level $\hat{s}^3$. This method goes like P1 with additional skewness constraint. de Athayde and Flôres (2004) state the optimization problem by Lagrange multipliers as follows[4]

$$\min_{w} \mathscr{L} = \sigma^2 + \lambda_1(\hat{\mu} - \mu) + \lambda_2(\hat{s}^3 - s^3)$$

where λ_1 and λ_2 are two associated Lagrange multipliers. Note that de Athayde and Flôres (2004) include riskfree asset into the variance minimization program, so this method can not produce a pure efficient portfolio of risky assets. In other words, the optimal portfolios by this method can not be characterized as a common risky investment part plus a riskfree investment part, where the mixture depends on the investor-specific risk preference. The optimal portfolio by the constrained variance minimization method has the same mixture of risky investment part and riskfree investment part across investors' risk preferences. This seems counter-intuitive since we expect more risk averse people to increase their riskfree investment part.

The other higher order moment portfolio optimization method is the shortage function method, introduced by Briec, Kerstens, and Jokung (2007). Its intuition is clear: for an evaluated portfolio, if we can find another portfolio with better moment profile, then this evaluated is inefficient and the portfolio with the largest moment profile improvement is efficient. Assume we evaluate portfolio i with moment profile $\mathcal{P}^i = [\mu_i \ \sigma_i^2 \ \kappa_i^4]'$. If we can find another portfolio j for which $\mu_j > \mu_i$, $\sigma_j^2 < \sigma_i^2$, and $\kappa_j^4 < \kappa_i^4$, then we say portfolio i is inefficient since $\mathcal{P}^j \succ \mathcal{P}^i$. If there is no such portfolio achieving dominating moment profile over it, then portfolio i is efficient. Efficient portfolios are those without any dominating moment profiles.

More formally, define a directional improvement as

$$g = [g_1 \ \ -g_2 \ \ -g_4]',$$

where g_1, g_2, and g_4 are all positive elements. For an evaluated portfolio i with moment profile $\mathcal{P}^i$, its shortage function is

$$\begin{aligned} \text{maximize} \quad & \delta \\ \text{subject to} \quad & \mu_i + \delta g_1 \leqslant \mu \\ & \sigma_i^2 - \delta g_2 \geqslant \sigma^2 \\ & \kappa_i^4 - \delta g_4 \geqslant \kappa^4 \end{aligned}, \tag{P3}$$

[4] This method involves the riskfree asset in the minimum variance portfolio determination, explicated in de Athayde and Flôres (2004, p. 1339) as

$$\min_{w} \mathcal{L} = w'\Sigma w + \lambda_1\big(\mu_P - w'\tilde{\mu} - (1 - w'\mathbf{1})r_f\big) + \lambda_2\big(s_P^3 - w'S(w \otimes w)\big),$$

where Σ and S are co-variance and co-skewness matrices respectively, μ_P and s_P^3 are prespecified mean return and skewness level, $\tilde{\mu}$ is the expected return vector of risky assets, $\otimes$ is the tensor product.

where δ, μ, σ^2, and κ^4 are associated with the portfolios in the feasible set.[5] A positive δ means that the evaluated portfolio i is inefficient, and the portfolio with δ unit of the directional improvement g is efficient relative to it. If δ is 0, then portfolio i is efficient along the direction of g. The case of negative δ makes no sense, because it is equivalent to positive δ accompanied by a negative directional improvement $-g$, which is in fact a moment profile deterioration. It implies that the evaluated portfolio is not attainable within the feasible set.

We can rewrite P3 in a concise way:

$$\begin{array}{ll} \underset{w}{\text{maximize}} & \delta \\ \text{subject to} & \mathcal{P}^i + \delta g \preceq \mathcal{P}(w) \end{array} . \tag{P4}$$

P4 indicates the essence of the shortage function method: in the feasible set, find the portfolio with the best moment profile along the directional improvement. δ is the improvement distance, and also the efficiency indicator.

The shortage function method has an evident limitation: the choice of directional improvement is arbitrary and this vector can not coincide with its projections in the space. That is to say, g can not include 0, otherwise the computational complexity for the optimization rockets. Consequentially, the choice of g can impact portfolio efficiency determination, as one can easily detect the different results by $g = [1,000\ -1\ -1]$ and by $g = [1\ -1\ -1,000]$.

2.2 Classic optimization may fail

In this subsection, we will illustrate that the classic optimization method is not adequate for high order portfolio optimization.

Since a portfolio is a linear combination of the underlying assets, its return range is set by the asset with the minimum mean return μ_{min} and the asset with the maximum mean return μ_{max}. For any return level μ in the range, there is always a portfolio achieving the lowest variance, exhibited by the left panel of Figure 1. Obviously, Portfolios A, B and C are efficient and Portfolios a and b are inefficient. This risk minimization is program P1.

However, in the mean variance kurtosis space, this constrained minimization is not adequate. Now suppose Portfolios a', b' and c' have the same kurtosis. Then, by program P2 with mean return constraint of μ_1 and variance constraint of σ_1^2, all of the three portfolios are efficient. Nonetheless, Portfolio C' has higher mean return and lower variance than a' and b'. Given the same kurtosis level, it is apparent that a' and b' are inefficient relative to C', so the constrained kurtosis minimization can give wrong answer about efficient portfolios.

Another interesting and more probable case is that Portfolios A' and a' have the same kurtosis. Now with the mean return constraint of μ_1 and variance constraint of σ_1^2, program

[5] We rephrase the problem in the context of mean variance kurtosis space. For details refere to the third section of Briec, Kerstens, and Jokung (2007).

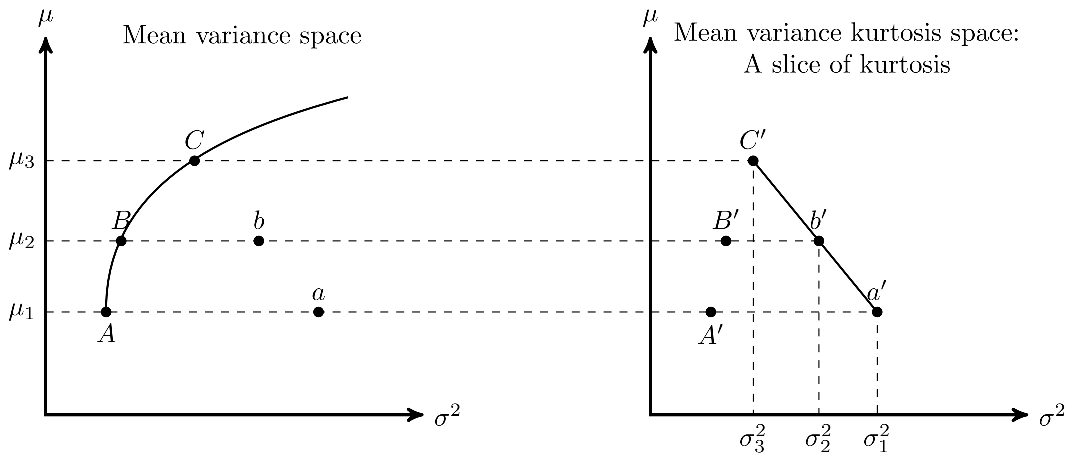

Figure 1: Optimization objective and constraint in mean variance space and mean variance kurtosis space

P2 detects that they are all efficient. Intuitively, Portfolio A' is better than a' due to its lower variance. So the classic optimization can miss the marginal return improvement like between A' and a', and between C' and a'. Their common moment happens to be the optimization objective, and the optimization constraint ignores the portfolio quality information by their different moments. The shortage function method also has this drawback of marginal return improvement detection due to the strict positivity of directional improvement vector.

It is worthwhile to notice that the classic optimization methods of P2 and P4 are more like a tool to evaluate whether a certain portfolio, especially like the market portfolio or a market index, is efficient or not. That is to say, efficient frontier is a by-product. To assemble a decent number of efficient portfolios, the optimization has to be executed for corresponding times, which usually indicates a low efficacy. The more important thing is the feasibility of the evaluated portfolio. In program P2, the constraints of $\hat{\mu}$ and $\hat{\sigma}^2$ have to be specified, and the feasibility of this combination remains a question. Similarly for the input of the portfolio return profile $[\mu_i\ \sigma_i^2\ \kappa_i^4]$ in P3.

Briec, Kerstens, and Jokung (2007) explicitly admit that the optimization has to search over "all possible combinations of returns, risk, and skewness of the portfolios" for efficient portfolios. Note that the set of all possible portfolio return profiles is equivalent to the set of all possible portfolio weights. Rather than examine the feasibility of portfolio return profile one by one, we can directly approximate the feasible portfolio set. Then, the portfolio optimization problem reduces to simple pairwise comparisons. This is the central idea for our mean variance kurtosis portfolio optimization method.

2.3 Pareto improvement method

The potential failure of classic optimization methods prompts us for a more general method, which takes into consideration of marginal profile improvement and the feasibility of portfolio profiles. This general algorithm is the Pareto improvement method.

The motivation of the Pareto improvement method is the same with that of the shortage function method: only the portfolios without any moment profile improvement can be labeled as efficient. However, the two methods are different at the efficient portfolio set specification. Given a well-defined directional improvement g and a set of evaluated portfolios, P4 is implemented for each evaluated portfolio to spot a corresponding efficient portfolio. Then the efficient portfolio set is assembled. On the contrary, the Pareto improvement method screens out inefficient portfolios from the feasible set, and this deletion does not depend on the directional improvement g specification. Portfolios whose moment profiles are found to be dominated by any other portfolio in the pairwise comparison are removed immediately from the candidate pool for portfolio efficiency. The surviving portfolios are efficient and they constitute the efficient set.

We make formal exhibition for this method. Assume we have a feasible portfolio set

$$\Omega = \{w \in \mathbb{R}^N; \Sigma_{i=1}^N w_i = 1, w \geqslant \mathbf{0}\}.$$

Corresponding moment calculations are defined in Section 1. The moment profile $\mathcal{P}$ is a function of w as

$$\mathcal{P}(w) = [\mu(w)\ \sigma^2(w)\ \kappa^4(w)]'.$$

The set for portfolios' moment profiles is defined as

$$\mathscr{P} = \{\mathcal{P}(w); w \in \Omega\}.$$

A Pareto improvement is an available dominance surplus over the moment profile.

Definition 2.1. *Portfolio j is a Pareto improvement relative to Portfolio i, denoted as*

$$\langle \mathcal{P}^j, \mathcal{P}^i \rangle \gneqq \mathbf{0},$$

where $\langle \mathcal{P}^j, \mathcal{P}^i \rangle = [\mu_j - \mu_i\ \sigma_i^2 - \sigma_j^2\ \kappa_i^4 - \kappa_j^4]'$ and $\gneqq$ means "no less than but not equal to." For an arbitrary vector ε and the zero vector $\mathbf{0}$ in same size, $\varepsilon \gneqq \mathbf{0}$ indicates that each element of ε is no less than 0 but at least one is not 0.

Thus, the set of efficient portfolios' moment profiles, $\mathscr{E}\mathscr{P}$, can be restated as a subset of $\mathscr{P}$ as

$$\mathscr{E}\mathscr{P} = \{\mathcal{P}(w); \langle \mathcal{P}(w'), \mathcal{P}(w) \rangle \gneqq \mathbf{0} \Rightarrow w' \notin \Omega\}.$$

Similarly, the efficient portfolio set Φ is a subset of the feasible portfolio set Ω,

$$\Phi = \{w \in \Omega; \mathcal{P}(w) \in \mathscr{E}\mathscr{P}\}.$$

Compared with the shortage function method, the Pareto improvement method has two important advantages: generality and simplicity. It is more general, because the definition of the Pareto improvement is more relaxed than that of the directional improvement. The Pareto improvement is not pre-specified, so it is free from potential optimization bias due to the choice of directional improvement, which is embedded in P4 to perform on all evaluated portfolios for corresponding efficient portfolios. And the Pareto improvement can be any linear combination of the standard bases with non-negative coefficients, while the pre-specified directional improvement in the shortage function method has to be a linear combination of the standard bases with all positive coefficients. In other words, the shortage function method demands improvement to be along each moment, while the Pareto improvement method only requires improvement to be along at least one moment, *ceteris paribus*. This method epitomizes the very essence of Pareto efficiency analysis in neoclassical economics: improve some without deteriorating any, and we apply such spirit in the context of portfolio selection.

Moreover, this method is simpler than the shortage function method in terms of computational complexity. By the shortage function method, a large set of evaluated portfolios is necessary for a commensurate set of efficient portfolios, and the optimization has to perform for each evaluated portfolio. For example, if we want 1,000 efficient portfolios, we have to prepare at least 1,000 evaluated portfolios and compute P4 every time. However, by the Pareto improvement method only pairwise comparisons among portfolio moment profiles are needed, thus it is much less computationally demanding. It is effective since the confirmation of only one Pareto improvement relative to the evaluated portfolio is enough to classify the latter as inefficient.

Some may argue that the efficacy of Pareto improvement method relies on the defining of feasible portfolio set Ω, and its related set of moment profiles $\mathscr{P}$. Actually, the shortage function method also depends on $\mathscr{P}$, which is shown by the right hand side of the constraint in P4. As Briec, Kerstens, and Jokung (2007) point out, the optimization has to go over all possible portfolio return profiles, which clearly refers to $\mathscr{P}$. The claim that the Pareto improvement method is not as theoretically sound as the shortage function method with respect to $\mathscr{P}$ is untenable. The key is to approach $\mathscr{P}$ as close as possible, which will be discussed in Section 3.

2.4 An illustration

We use a simple example to illustrate the advantage of Pareto improvement method for higher order moment efficient set identification, as well as the reason why the constrained variance minimization method and the shortage function method may misclassify portfolios.

Table 3: Marginal improvements and portfolio efficiency

	Mean	Variance	Kurtosis
Portfolio 1	8%	12%	2.1%/1.9%
Portfolio 2	8%	10%	2.1%
Portfolio 3	6%	11%	2%
Portfolio 4	6%	9%	2%

Table 3 gives 4 portfolios with their corresponding moment profiles. The mean variance optimization P1 easily spots Portfolio 2 as efficient if $\hat{\mu} = 8\%$ since it gives variance of 10% lower than Portfolio 1, and spots Portfolio 4 as efficient if $\hat{\mu} = 6\%$ since its variance of 9% is lower than any other portfolio. By the Pareto improvement method, Portfolio 1 has a Pareto improvement towards Portfolio 2 along variance, and the situation is similar for the comparison of Portfolios 3 and 4. Based on the Pareto improvements identified, this method classifies Portfolios 1 and 3 as inefficient. The mean variance analysis and the Pareto improvement method yield the same solution in the mean variance space, since the duality connection of mean and variance is direct. On the other hand, the shortage function method can give different answer. For a directional improvement of $g = [2 \ \ -1]'$, Portfolio 3 is inefficient since it has $\delta = 1$ towards Portfolio 2. However, $\delta = 0$ for Portfolio 1 and the shortage function method misclassifies it as efficient.

In the mean variance kurtosis space, we first use the constrained kurtosis minimization P2 to detect efficient portfolios.[6] Now set $\hat{\mu} = 6\%$ and $\hat{\sigma}^2 = 11\%$, the solutions are Portfolios 3 and 4, both giving the same kurtosis of 2%. The shortage function method gives similar answer with a directional improvement of $g = [2 \ -1 \ -1]'$. However, we can clearly identify the dominance of Portfolio 4 over Portfolio 3 by its lower variance, which is suggested by the Pareto improvement method. The pair of Portfolios 1 and 2 is a similar case, where P2 spots them all as efficient at $\hat{\mu} = 8\%$ and $\hat{\sigma}^2 = 12\%$ for the same kurtosis of 2.1%, while the Pareto improvement method gives a different answer which makes more sense. Thus, in the mean variance kurtosis space the constrained kurtosis minimization method and the shortage function method may misclassify inefficient portfolio as efficient.

This misclassification error results from the fact that the three dimensional connection of mean, variance, and kurtosis is intertwined. The constrained kurtosis minimization has to search all the data grid of mean and variance, taking them as pure constraints and ignoring

[6]The constrained kurtosis minimization and the constrained variance minimization are essentially two variants of the constrained risk minimization in the mean variance kurtosis space. We exhibit the constrained kurtosis minimization here, because it is convenient to explicate the relationship between the mean variance and mean variance kurtosis efficient sets. In similar fashion we can also construct another example to illustrate the same misclassification concern with the constrained variance minimization.

the innate portfolio quality revealed by the mean and variance information. Yet, the Pareto improvement method circumvents such an error by equalizing the constraints and objective. Therefore, this method is better than the constrained risk minimization method and the shortage function method in high dimensional portfolio optimization. The Pareto improvement method also obviates the technical difficulty of reaching an explicit solution beyond the mean variance space, as indicated by the attempt of de Athayde and Flôres (2004).

Another key point demonstrated by the example is that Portfolios 2 and 4, which are efficient in the mean variance space, are also mean variance kurtosis efficient. Due to their advantage in the $[\mu\ \sigma^2]$ part, they can not be inferior in each moment no matter if their κ^4 are superior or not. Only portfolios with all moments inferior will be classified as inefficient by the Pareto improvement method. So, we can infer a key relationship between the set of mean variance efficient portfolios and the set of mean variance kurtosis efficient portfolios: the former set is a subset of the latter one. This inference plays an important role at enhancing approximation efficacy in the later step.

We have another key relationship inference about the sets of efficient portfolios. If Portfolio 1 has a kurtosis of 1.9%, then it is also efficient in the mean variance kurtosis space since it has the overall lowest kurtosis. This suggests that mean variance inefficient portfolio can be mean variance kurtosis efficient. These two inferences give us a hint on the composition of the set of mean variance kurtosis efficient portfolios: the mean variance efficient portfolios are all mean variance kurtosis efficient, and the mean variance inefficient portfolios with kurtosis edge can be mean variance kurtosis efficient.

3 Dirichlet simulations

To take full advantage of the Pareto improvement method, it is necessary to approach the feasible portfolio set Ω as close as possible. The representative sampling for portfolios is both practically and theoretically feasible. In practice, market frictions deter minimal portfolio adjustments. Because of trading costs like brokerage commissions, fees, and taxes, a small portfolio change, like a 0.1% increase in one asset's weight at the cost of another asset's, is not profitable or even subject to losses. Several papers have documented the large economic losses by investors' overtrading costs, like Barber, Lee, Liu, and Odean (2008).

Chotikapanich and Griffiths (2002) initiate the use of Dirichlet distribution for sampling income shares to estimate Lorenz curve. This offers an inspiration that the Dirichlet distribution can also be used to sample portfolio weights, whose characteristics are similar to income shares. Next we introduce the Dirichlet distribution and its application in the context of portfolio selection.

The Dirichlet distribution is the multivariate generalization of the beta distribution. It has the following probability density function:

$$f(w_1, w_2, ..., w_N; \alpha_1, \alpha_2, ..., \alpha_N) = \frac{\Gamma(\sum_{i=1}^{N} \alpha_i)}{\prod_{i=1}^{N} \Gamma(\alpha_i)} \prod_{i=1}^{N} w_i^{\alpha_i - 1}$$

where $\alpha_1, \alpha_2, ..., \alpha_N$ are concentration parameters and are all positive. $\Gamma(\cdot)$ is the Gamma function such that $\Gamma(\alpha_i) = \int_0^\infty x^{\alpha_i - 1} e^{-x} \mathrm{d}x$.

For the Dirichlet distribution, $\sum_{i=1}^{N} w_i = 1$ and $w_i \geqslant 0$. This regular sum property makes Dirichlet distribution a naturally appropriate distribution scheme to simulate portfolio weights. Its non-negativity is consistent with the constraint of non-short selling, and it also includes extreme cases where individual assets are fully invested.

If we define $\alpha_0 = \sum_{i=1}^{N} \alpha_i$, then the expected value and variance of the marginal distribution are

$$\mu_i = \frac{\alpha_i}{\alpha_0},$$

$$\sigma_i^2 = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)},$$

showing that the concentration parameters affect the marginal distribution.

For a comprehensive understanding of the Dirichlet distribution, first consider the case of equal concentration parameters where we simulate the weights of three-asset portfolios. Let w_1 be the weight for Asset 1, w_2 the weight for Asset 2, and w_3 the weight for Asset 3, and $\alpha = \alpha_1 = \alpha_2 = \alpha_3$. The weight schemes generated by Drichlet distribution here form a standard 2 simplex in the 3 dimensional space, as shown by Figure 2. The impact of concentration parameters on simulated outcomes is straightforward. We see that the simulated points are evenly scattered when α is 1, and in this case, the Dirichlet distribution is equal to the uniform distribution. This special setting is called the flat Dirichlet distribution. As α gets smaller, points are more condensed towards the boundaries, and corner simulations including 0 values are more frequent. If α is bigger than 1, points are more condensed around the simplex center point, and simulations including similar values are more frequent. For the symmetric Dirichlet distribution, we can also get a clue about the impact of concentration parameter from the expected value and variance equations. In this three-asset case, the expected values are equally to be 1/3, and the variance is $\frac{2}{9(3\alpha+1)}$. So, the four sub-cases in Figure 2 with varying α share the same expected value, but the variance is negatively related with α.

Then we come to see how unequal parameters lead to asymmetric concentration shown in Figure 3. It is obvious that the higher concentration parameter relative to the others, the more condensed high values for this asset. In the first sub-figure, α_3 is way bigger than the

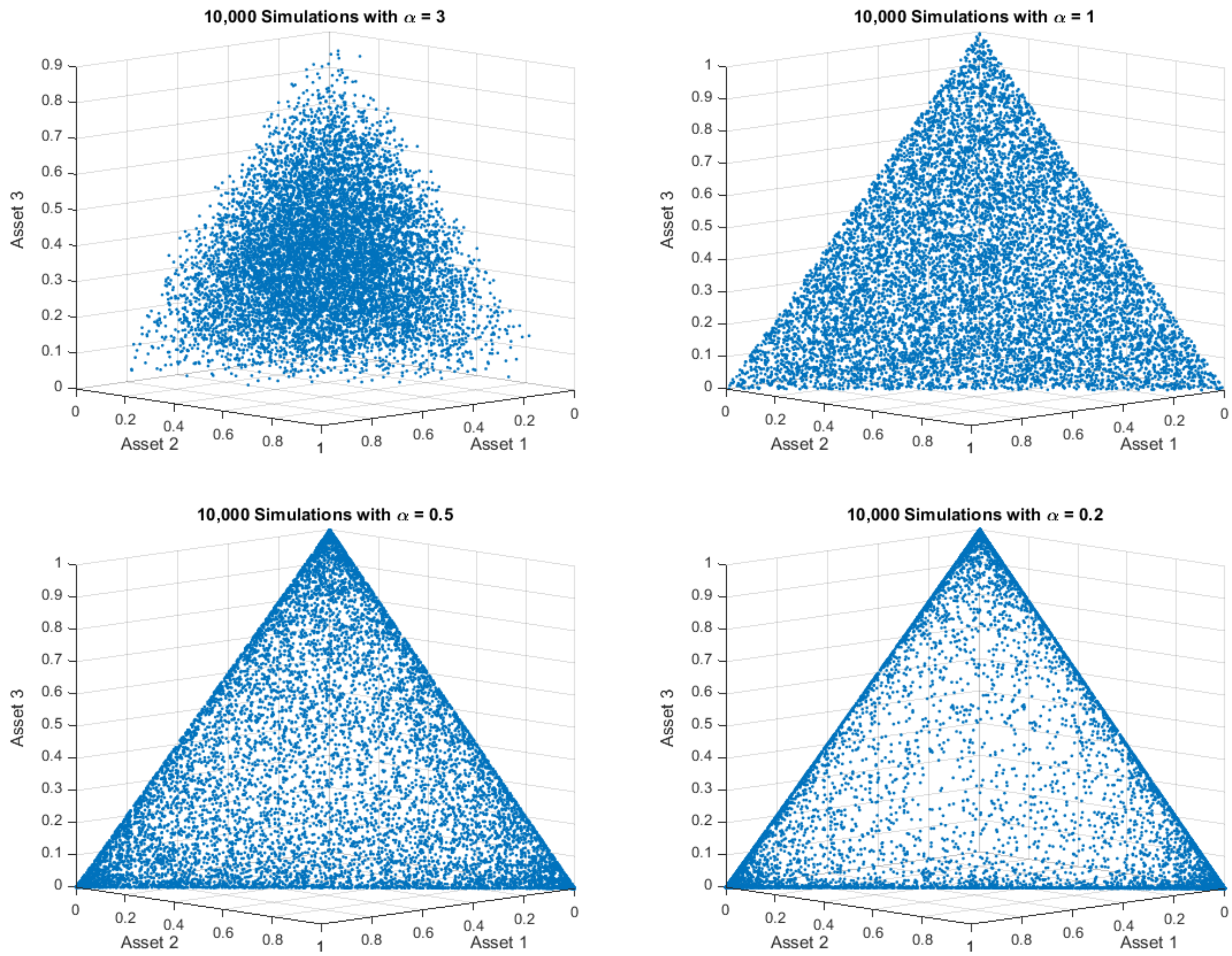

Figure 2: Dirichlet simulations with equal α

others, thus more portfolios are close to the full investment in Asset 3. On the contrary, if α_3 is only half of the others as in the second sub-figure, more portfolios are simulated as major investment in Assets 1 and 2 plus minimal investment in Asset 3. The asymmetric feature can also be inferred from the expected value and variance equations. In the first subcase, $\mu_1 = \mu_2 = 0.2$ and $\mu_3 = 0.6$, so the Drichlet portfolios have more weights in Asset 3, which is about 60%. In the second subcase, $\mu_1 = \mu_2 = 0.4$ and $\mu_3 = 0.2$, so the Drichlet portfolios have less weights in Asset 3, which is about 20%.

Now we have an intuitive illustration of the Dirichlet distribution and an understanding of how the concentration parameters impact the Dirichlet portfolios. Then we use Dirichlet portfolios as an input to the Pareto improvement method to construct mean variance kurtosis efficient frontier.

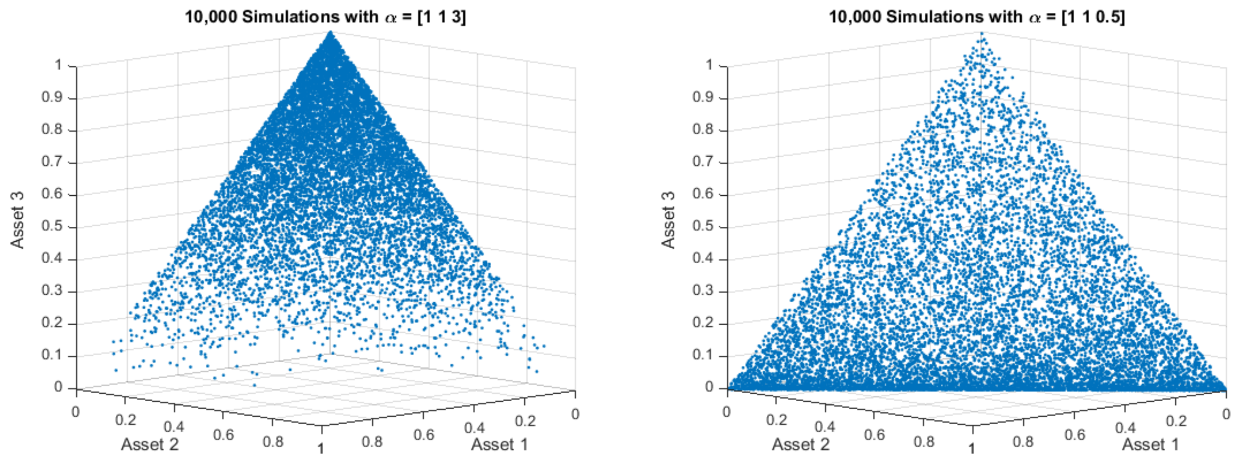

Figure 3: Dirichlet simulations with unequal α

4 The efficacy of the Dirichlet efficient frontier

The core of our innovative way for mean variance kurtosis efficient frontier construction is the combination of the Pareto improvement method and the Dirichlet simulations. Specifically, we approximate the feasible portfolio set with Dirichlet portfolios. These Dirichlet portfolios are the input to the Pareto improvement method, and the Pareto pairwise comparisons between moment profiles directly eliminate the inefficient portfolios and their associated moment profiles from the efficiency pool. The leftover ones are efficient and they constitute the Dirichlet efficient frontier.

Formally, we approximate the feasible portfolio set Ω by sampling Dirichlet portfolios. We get the Dirichlet feasible portfolio set Ω^* as

$$\Omega^* = \{w \mid w \sim \mathcal{D}ir(\boldsymbol{\alpha}), \boldsymbol{\alpha} > \mathbf{0}\},$$

where $\boldsymbol{\alpha}$ is the vector of concentration parameters defined in Section 3, and it is in the same size as of the asset universe N. Then the corresponding set for the Dirichlet portfolios' moment profiles is

$$\mathscr{P}^* = \{\mathcal{P}(w); w \in \Omega^*\}.$$

After the Pareto improvement comparisons within $\mathscr{P}^*$ and the identification of inefficient portfolios, the set for the Dirichlet efficient portfolios' moment profiles is

$$\mathscr{E}\mathscr{P}^* = \{\mathcal{P}(w); \langle \mathcal{P}(w^{'}), \mathcal{P}(w)\rangle \gneqq \mathbf{0} \Rightarrow w^{'} \notin \Omega^*\},$$

and the set of Dirichlet efficient portfolios is

$$\Phi^* = \{w \in \Omega^*; \mathcal{P}(w) \in \mathscr{E}\mathscr{P}^*\}.$$

Note that efficient frontier is a direct exposition of the set of efficient portfolios' moment profiles, other than the set of efficient portfolios. The latter set is just a collection of efficient portfolio weights. Therefore, the Dirichlet efficient frontier in the mean variance kurtosis space is a display of $\mathscr{E}\mathscr{P}^*$.

Regarding the efficacy of the Dirichlet efficient frontier $\mathscr{E}\mathscr{P}^*$, as well as its relationship with the population efficient frontier $\mathscr{E}\mathscr{P}$, we have 3 propositions as following.

Proposition 4.1. *The Dirichlet efficient frontier is the lower bound for the population efficient frontier.*

Proof. Note that $\Phi^* \subseteq \Omega^* \subseteq \Omega; \forall\ w \in \Phi^*, \exists\ w^{'} \in \Omega$ *such that* $\mathcal{P}(w^{'}) \succeq \mathcal{P}(w)$*, thus* $\mathscr{E}\mathscr{P} \succeq \mathscr{E}\mathscr{P}^*$*.*

Proposition 4.2. *At certain number of Dirichlet portfolios, the efficacy of Dirichlet efficient frontier is negatively associated with the size of asset universe* N*.*

This proposition maintains that the Dirichlet feasible portfolio set Ω^* is less representative as the size of asset universe increases, thus the Dirichlet efficient frontier $\mathscr{E}\mathscr{P}^*$ is more prone to be biased, *ceteris paribus*. Theoretically, the feasible portfolio set should span over the standard $N-1$ simplex. However, the coverage ratio for a fixed number of simulations decreases as the dimension increases. For example, a set of 1,000 Dirichlet portfolios may well occupy a line (standard 1 simplex) reasonably, but not so for a plane (standard 2 simplex). In fact, this proposition is a direct reflection of "curse of dimensionality" (Bellman, 1961; Donoho, 2000) in simulation.

Proposition 4.3. *At an appropriate size of asset universe* N*, the Dirichlet efficient frontier is asymptotically consistent with the population efficient frontier. That is to say,*

$$\lim_{Z \to \infty} d\big(\mathcal{P}(w), \mathcal{P}(w^{'})\big) = 0,$$

where Z *is the number of Dirichlet portfolios,* $\mathcal{P}(w)$ *is any moment profile from* $\mathscr{E}\mathscr{P}^*$ *and* $\mathcal{P}(w^{'})$ *its counterpart from* $\mathscr{E}\mathscr{P}$*, and* $d(\cdot,\cdot)$ *is the Euclidean distance.*

In the mean variance space, suppose that at an expected return level of R, σ is the corresponding volatility along the population efficient frontier, $\tilde{\sigma}$ the corresponding volatility along the Dirichlet efficient frontier. This proposition states that as the number of Dirichlet portfolios Z goes to infinity, the difference of $\tilde{\sigma}$ over σ goes to 0. That is to say,

$$\lim_{Z \to \infty} \tilde{\sigma}(R) = \sigma(R)$$

for all $R_{\text{minimum variance portfolio}} \leqslant R \leqslant R_{\text{maximum return portfolio}}$.

This proposition is an application of the asymptotic consistency of nonparametric frontier estimation (Cazals, Florens, and Simar, 2002; Simar and Wilson, 2000, for example) in production theory. There, production possibility frontier can be approached by parametric and nonparametric ways. Briec, Kerstens, and Jokung (2007) translate Luenberger shortage function in the portfolio context, so they follow a parametric way. The Pareto improvement and Dirichlet way is nonparametric. We then rephrase this proposition in terms of nonparametric production frontier estimation.

Suppose production activity is characterized as using a set of inputs x to produce a set of outputs y. $\varphi(y)$ denotes the minimum input for an output level no less than y, so it also depicts the cost frontier. For m independent identically distributed random inputs $(X^1, X^2, \ldots, X^m)$ given $Y \geq y$, the expected minimum input function is defined as

$$\varphi_m(y) = E\big(\min(X^1, X^2, \ldots, X^m)|Y \geq y\big).$$

Cazals, Florens, and Simar (2002, Theorem 2.3) prove that for any fixed y, $\lim_{m\to\infty} \varphi_m(y) = \varphi(y)$. They then show that when m goes to large, the free disposal hull (FDH) estimator of the expected minimum input function of order m converges to the input efficient frontier. Cazals, Florens, and Simar (2002, Theorem 5.2) generalize this convergence in the setting with p inputs and q outputs. The rate of convergence can also be measured. For a production level (x_0, y_0), the associated efficiency measure $\theta(x_0, y_0)$ is the minimum scalar such that θx_0 can still produce y_0. Intuitively, $\theta = 1$ for efficient production levels. Simar and Wilson (2000, equation 2.16) provide that

$$\hat{\theta}_{FDH}(x_0, y_0) - \theta(x_0, y_0) = \mathcal{O}(n^{-\frac{1}{p+q}}),$$

where n is the production sample size.

In our Dirichlet context, we take risk as the input and return as the output. That is too say, $p = 2$ to account for variance and kurtosis, and $q = 1$ to account for mean return. This assignment is more intuitive to perceive the balance of risk and return (input and output respectively) than the balance of portfolio weight and portfolio return profile (input and output respectively). Note that Dirichlet portfolios are independent and identically distributed, and return moments as the function of Dirichlet portfolios are also independent. So the discussion of efficient portfolios transforms into the discussion of efficient production units, and the expected minimum input function is equivalent to the expected minimum risk function. The rate of convergence of the FDH efficiency measure, which is analogues to the rate of convergence of the Dirichlet efficient frontier to the population efficient frontier, is $n^{\frac{1}{3}}$.

There are three noticeable points for this nonparametric efficient frontier translation and corroboration. First, Cazals, Florens, and Simar (2002, page 20) recognize that no simple explicit expression of $\hat{\theta}(x, y)$ is possible, and the easiest way is to use Monte Carlo simulations.

This is consistent to the spirit of our Dirichlet portfolios for feasible portfolio set approximation. Second, our Pareto improvement method facilitates the convergence in practice, as the default efficiency measure θ in the production theory relies on an efficiency determination very similar to the shortage function method. Third, the size effect has a hint by the rate of convergence. For the mean variance approach, the rate of convergence $n^{\frac{1}{2}}$ is larger than that of any higher order moment portfolio optimization approach. So the curse of high dimensionality on portfolio optimization has two effects: on the approximation of feasible portfolio set and on the rate of convergence for efficiency measure.

We experiment on these propositions for the efficacy of Dirichlet efficient frontier. We fetch the Dow Jones Industrial Average (DJIA) components' monthly data based on their adjusted close prices from September, 2009 to August, 2015.[7] Setting concentration parameters uniformly to 1, we have 10,000 Dirichlet portfolio for the first 3 assets of DJIA[8] and for the total 30 assets respectively. We plot these Dirichlet portfolio profiles as red points, and also plot the corresponding population efficient frontiers by blue curve in Figure 4 for better comparison. Apparently, the Dirichlet efficient portfolios of the 3 assets perfectly approach the population efficient frontier, as we can see that the blue curve is occupied by the red points in the first sub-figure. However, in the 30 assets case, the Dirichlet portfolios cluster below the population efficient frontier in the second sub-figure, suggesting that the Dirichlet efficient frontier has a material downward bias. This observation lends direct evidence for Proposition 4.2 on the size effect. As the number of Dirichlet portfolios increases to 1,000,000 in the third sub-figure, the portfolio cluster expands and the Dirichlet efficient frontier improves substantially compared to the second sub-figure. This observation gives support evidence for Proposition 4.3 on the asymptotic consistency.

Note that the concentration parameter α also plays a critical role in the Dirichlet distribution and we a different α in Figure 4. For the same number of Dirichlet portfolios, a change of α can significantly enhances Dirichlet efficient frontier efficacy. Compared to the case of $\alpha = 1$, the 1,000,000 Dirichlet portfolio cluster of $\alpha = 0.1$ march towards northwest, indicating a substantial enhancement for the Dirichlet efficient frontier.[9]

From the experiments, we show that a careful treatment on the asset universe, Dirichlet portfolio number, and concentration parameter can substantially improve the efficacy of Dirichlet efficient frontier. Thus, it is feasible to make the Dirichlet efficient frontier more

[7]This experiment is mainly illustrative and we omit the impact of 5 DJIA component changes during this period. Data are from Yahoo Finance. The companies and their tickers can be found at the following website `https://www.cnbc.com/quotes/?symbol=AAPL,AXP,BA,CAT,CSCO,CVX,DD,DIS,GE,GS,HD,IBM,INTC,JNJ,JPM,KO,MCD,MMM,MRK,MSFT,NKE,PFE,PG,TRV,UNH,UTX,V,VZ,WMT,XOM`

[8]The three companies are Apple (AAPL), American Express (AXP), and Boeing (BA).

[9]More exhibitions for Dirichlet efficient frontier efficacy with respect to asset universe, simulation numbers, and concentration parameters are available upon request from the authors.

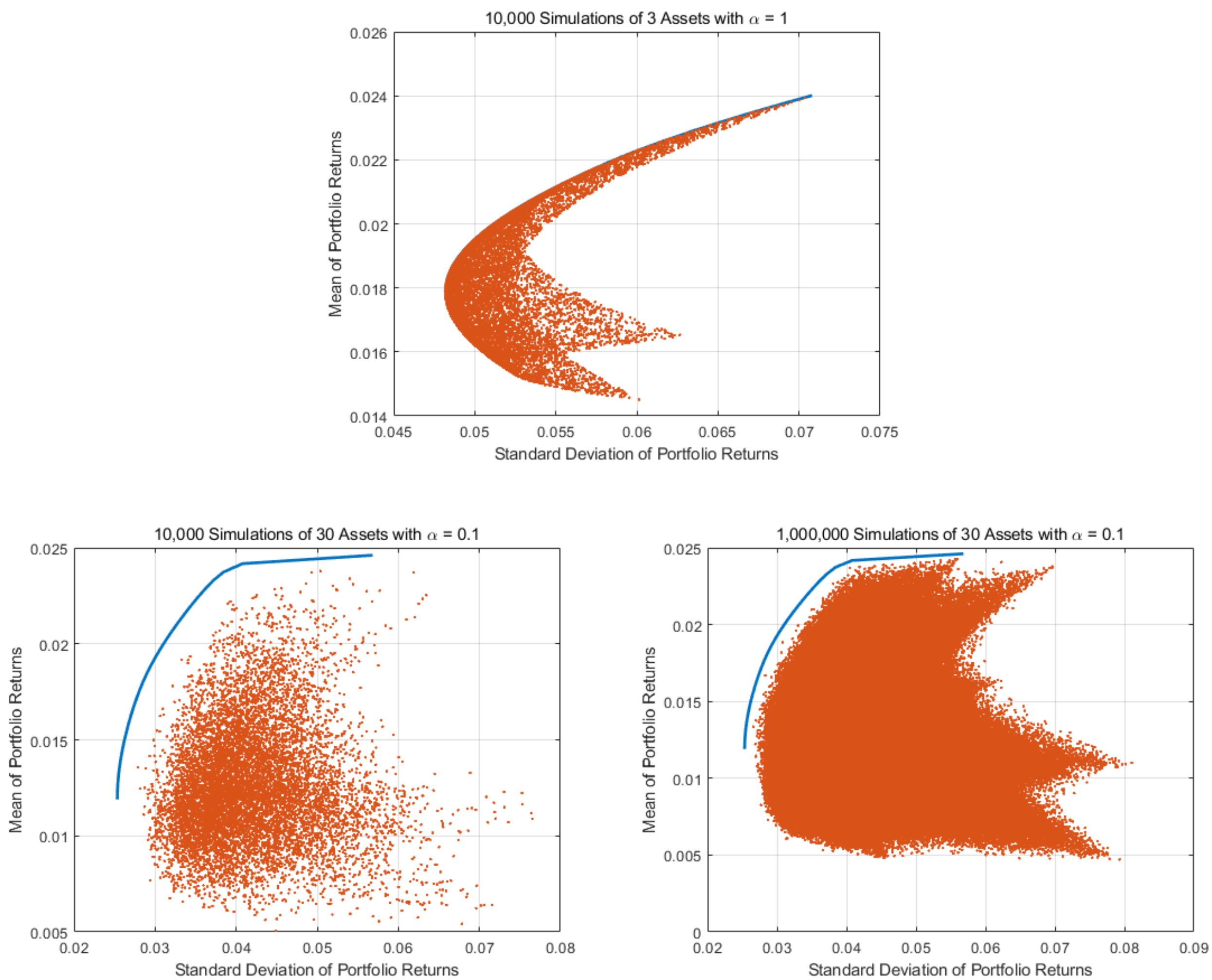

Figure 4: Dirichlet efficient frontier efficacy: Asset universe and simulation number

consistent with the population efficient frontier.

Recall that we have two key inferences on the relationship of the mean variance efficient portfolio set and the mean variance kurtosis efficient portfolio set. These inferences are useful in the construction of mean variance kurtosis efficient frontier. Because the mean variance efficient portfolio set is a necessary subset of the mean variance kurtosis efficient portfolio set, this fact calls for a specification of the Dirichlet concentration parameters which produce the most consistent mean variance efficient frontier first. We identify the concentration parameters so that the Dirichlet mean variance efficient frontier best mimic the population mean variance efficient frontier. This set of parameters is included to construct the efficient subset by mapping mean variance efficiency to mean variance kurtosis efficiency. The other inference that the mean variance inefficient portfolios can be mean variance kurtosis efficient requires a broad coverage for the Dirichlet portfolios. Therefore we vary the set of parameters in a wide range to generate portfolios as diverse as possible. These two inferences, combined with a careful treatment of the asset universe size and the Dirichlet portfolio size by the preceding propositions, facilitate the combination of the Pareto improvement method and the Dirichlet simulations in developing the mean variance kurtosis efficient frontier with favorable efficacy.

5 Empirical implementation on the S&P 500 index

5.1 Reverse optimization for equilibrium returns

In this section, we choose the S&P 500 index constituents as our asset universe for empirical implementation. As the most widely followed stock market index, the S&P 500 index covers almost all of the influential blue-chips over diverse industries, and accounts for about 80% of the total capitalization of US stock markets. This index is also commonly used as a proxy for the market portfolio, so it is reasonable to assume that its components span the practical asset universe.

Rather than taking the individual stocks as risky assets for investment, we use a sector approach due to the aforementioned asset universe concern on the Dirichlet efficient frontier efficacy. If we invest directly in the constituents, the size of asset universe N is 500.[10] As explained, the weight vectors are in a standard 499 simplex, and the Dirichlet efficient frontier in this case can be inconsistent to the population efficient frontier. Moreover, the adjustments of index membership can affect the asset universe, as additions and deletions make some assets newly tradable or untradeable anymore. To keep up with the index dynamics, investment

[10]In fact, there are 505 stocks in the index, since some component companies have several listed share classes, like Discovery Communications, Google, Comcast, Twenty-First Century Fox, and News Corporation in April 2018.

universe has to change accordingly. This problem is serious when a long sample period is needed for better return estimates.

Correspondingly, the sector investment circumvents these drawbacks. Despite minor differences across various industry classification methods, the number of sectors are about 10, a great reduction from the number of constituents 500. Therefore, this sector investment helps to enhance the Dirichlet efficient frontier efficacy. Besides, since the sectors are synthesized from the constituents, the sector composition has already adjusted for the constituent changes. All in all, the sector investment has obvious advantages compared with the direct investment in components, thus we use sector investment in the empirical study.

Datastream offers data for 10 economic sectors of the S&P 500 index. These economic sectors are classified according to the Global Industry Classification Standard (GICS). The 10 sectors are Consumer Discretionary (CDis), Consumer Staples (CSta), Energy (Ene), Financials (Fin), Health Care (HCare), Industrials (Ind), Information Technology (Info), Materials (Mat), Telecommunication Services (Tel), and Utilities (Utl), respectively. We take these 10 economic sectors as our underlying assets.

We construct a reverse optimization as in the Black Litterman model for equilibrium returns, see Black and Litterman (1992) and Idzorek (2011). We take the equilibrium returns as a neutral input for the mean variance kurtosis portfolio optimization. Here the implied excess equilibrium return Π is calculated as

$$\Pi = \lambda \Sigma w_{mkt},$$

where λ is risk aversion coefficient, Σ the covariance matrix of excess returns, and w_{mkt} the vector of market capitalization weights. The market implied risk aversion coefficient is defined as

$$\lambda = \frac{E[r_{mkt}] - r_f}{\sigma^2} = \frac{E[r_{mkt}] - r_f}{w_{mkt}' \Sigma w_{mkt}},$$

where r_{mkt} is market portfolio return, σ^2 its variance, and r_f the risk free rate. See Idzorek (2011) for details.

Therefore, we have to collect data for the market portfolio return (the S&P 500 index), the risk free rate (the 3-month T-bill), the covariance matrix (10 sector returns), and the market weight (10 sector capitalizations). We use total return (RI) data for all of these return series. Note that in Datastream, RI data for the 10 sectors uniformly start from September 11th, 1989 and the market capitalization (MV) data mostly from January 23rd, 1995. The S&P 500 index data and the 3M T-bill data have better period coverage. Therefore, we set sample period to be from 1995 to 2015, which is just before the GICS sector restructuring. For each year, we get daily data for both RI and MV. Covariance matrix is calculated with daily excess return for these 10 sectors, and market cap weight is characterized as the capitalization fraction of each sector to the S&P 500 index at the last day of that year. Excess return is set as the daily return difference between the S&P 500 index or economic sectors returns over the 3M T-bill return.

Table 4: Market implied risk aversion and equilibrium returns

	λ	Ene	Mat	Ind	CDis	CSta	HCare	Fin	Info	Tel	Utl
1995	44	0.18	0.25	0.31	0.28	0.25	0.22	0.4	0.63	0.33	0.2
1996	11.34	0.13	0.11	0.17	0.17	0.16	0.18	0.22	0.25	0.17	0.11
1997	7.72	0.26	0.22	0.27	0.22	0.29	0.34	0.34	0.39	0.23	0.13
1998	5.3	0.16	0.19	0.24	0.27	0.19	0.25	0.33	0.34	0.18	0.06
1999	4.37	0.07	0.06	0.12	0.16	0.09	0.15	0.19	0.27	0.16	0.05
2000	-2.98	-0.04	-0.08	-0.12	-0.12	-0.06	-0.08	-0.15	-0.19	-0.11	-0.05
2001	-3.03	-0.05	-0.11	-0.15	-0.15	-0.03	-0.05	-0.13	-0.25	-0.11	-0.05
2002	-3.41	-0.18	-0.19	-0.21	-0.21	-0.1	-0.17	-0.23	-0.29	-0.23	-0.16
2003	8.75	0.16	0.28	0.28	0.33	0.18	0.24	0.32	0.43	0.33	0.18
2004	7.74	0.08	0.12	0.11	0.11	0.07	0.09	0.1	0.14	0.09	0.06
2005	2.02	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
2006	10.27	0.13	0.15	0.11	0.11	0.07	0.08	0.11	0.15	0.09	0.07
2007	0.87	0.03	0.03	0.02	0.02	0.01	0.02	0.03	0.02	0.02	0.02
2008	-2.37	-0.38	-0.36	-0.31	-0.34	-0.21	-0.25	-0.46	-0.31	-0.31	-0.27
2009	3.25	0.31	0.33	0.33	0.32	0.13	0.15	0.77	0.27	0.19	0.16
2010	4.68	0.19	0.21	0.2	0.18	0.09	0.11	0.23	0.17	0.1	0.11
2011	0.88	0.06	0.06	0.06	0.05	0.03	0.04	0.07	0.05	0.03	0.03
2012	9.56	0.21	0.21	0.19	0.17	0.09	0.12	0.22	0.19	0.1	0.06
2013	23.18	0.35	0.38	0.35	0.35	0.26	0.31	0.41	0.3	0.23	0.23
2014	10.37	0.16	0.15	0.16	0.15	0.09	0.16	0.15	0.16	0.09	0.07
2015	1.05	0.03	0.03	0.02	0.03	0.02	0.03	0.03	0.03	0.02	0.02
Mean	6.84	0.09	0.10	0.10	0.10	0.08	0.09	0.14	0.13	0.08	0.05
StD	10.59	0.16	0.18	0.18	0.18	0.12	0.15	0.26	0.24	0.16	0.12
Min	-3.41	-0.38	-0.36	-0.31	-0.34	-0.21	-0.25	-0.46	-0.31	-0.31	-0.27
Median	4.68	0.13	0.12	0.12	0.15	0.09	0.11	0.15	0.16	0.09	0.06
Max	44	0.35	0.38	0.35	0.35	0.29	0.34	0.77	0.63	0.33	0.23
Skew	2.22	-1.05	-0.79	-0.66	-0.73	-0.33	-0.49	-0.02	-0.24	-0.64	-0.98
Kurt	8.44	4.54	3.42	2.67	2.88	3.02	2.83	3.75	2.62	3.13	4.25

With this reverse optimization technique, we compute annualized equilibrium excess returns for these economic sectors and market implied risk aversion coefficient for each year, summarized in Table 4. We see that the commonly believed high beta sectors, Financials and Information Technology, for example, have the highest average annual excess return as well as the highest standard deviation. The commonly believed low beta sectors, Utilities and Consumer Staples, for example, have the lowest average annual excess return as well as the lowest standard deviation.

More importantly, we see the risk aversion varies a lot from -3.41 to 44, and it is positively skewed. This indicates that the investors are likely to be highly risk averse. Note that variance and kurtosis are correlated, so in periods when the investors are highly risk averse, they also tend to be highly extreme risk averse. This is an evidence of extreme risk in affecting investors' portfolio decisions.

Another key observation is that, the sorting for these economic sectors based on volatility is quite different from the sorting based on kurtosis. For example, the Utilities sector has the lowest volatility of 0.12 but has one of the highest kurtosis of 4.25. And also, the Information Technology sector has one of the highest volatility of 0.24 but at the same time has the lowest kurtosis of 2.62. This suggests that kurtosis can give a different risk aspect than volatility, and help to depict a complete risk profile.

5.2 Dirichlet parameterization

Our goal is to take advantage of Dirichlet distribution to simulate abundant portfolios for the construction of efficient frontier in the mean variance kurtosis space. Remind the first key inference about the mean variance efficient set and mean variance kurtosis efficient set: mean variance efficient portfolios must be mean variance kurtosis efficient, and the mean variance efficient set is a distinctive subset of the mean variance kurtosis efficient set. With this reasoning, we can primarily focus on those Dirichlet distributions best matching the mean variance efficient frontier. We experiment 100,000 Dirichlet portfolios with different concentration parameters in Figure 5. The blue curve is the population efficient frontier constructed by the mean and covariance matrix of the implied equilibrium excess returns from the reverse optimization. Red points are coordinates of mean and volatility for each simulated portfolio.

We see that a vary in concentration parameter can significantly enhances the Dirichlet portfolio efficiency, consistent to Section 4 about the efficacy of Dirichlet efficient frontier. For the uniform case of $\alpha = 1$, the 100,000 Dirichlet portfolios have substantial gap to the population efficient frontier, especially when the expected return are at the two ends. However, when the concentration parameter goes to small as $\alpha = 0.1$, the 100,000 Dirichlet portfolios realize significant performance improvement. The Dirichlet efficient frontier nearly coincide with the population efficient frontier, because the blue curve is occupied by red points along all expected return levels. Therefore, we have convincing justification in including $\alpha = 0.1$ for the mean variance kurtosis efficient frontier construction, reflecting the inference that the mean variance efficient set is a subset of the mean variance kurtosis efficient set.

The second key inference is that the mean variance inefficient portfolios can be mean variance kurtosis efficient. With regard to this inference, we take more simulation scenarios by varying the Dirichlet concentration parameters in a wide range additional to the previous

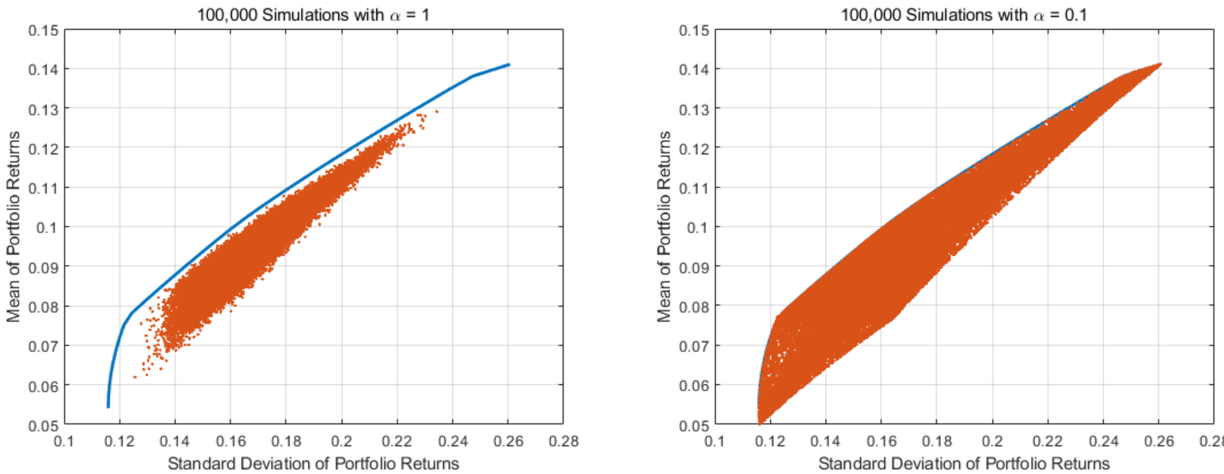

Figure 5: Efficient frontier relative to α

specification of $\alpha = 0.1$, which serves the best of the mean variance efficient frontier. To enlarge the Dirichlet portfolio variability, we simulate 100,000 portfolios for each concentration parameter ranging from 0 (since the concentration parameter has to be positive, we use 0.01 to approach 0) to 5 with a step of 0.1, so we parameterize α to be [0.01 0.1 0.2 ... 5]. Therefore, we have 5,100,000 Dirichlet portfolios in total.

Figure 6 shows the 5.1 million Dirichlet portfolios in the mean variance kurtosis space. Instead of being a widespread plane, the Dirichlet portfolios constitute a part of cone, with the vertex towards the origin. We see that the mean variance kurtosis combinations are not valid at all of the mean variance kurtosis grids. For the middle part of expected return levels, the combinations of variance and kurtosis stretch.

5.3 Mean variance kurtosis portfolios

For more concrete impression on the Dirichlet portfolios, Table 5 reports the summary statistics in terms of mean variance efficiency and mean variance kurtosis efficiency. Out of the 5.1 million Dirichlet portfolios, 3,499 are mean variance efficient and 4,873 are mean variance kurtosis efficient. As expected, all of the mean variance efficient portfolios are mean variance kurtosis efficient, and 1,374 portfolios are mean variance inefficient but mean variance kurtosis efficient. Interestingly, the mean variance kurtosis efficient portfolios have higher mean return (0.0964 vs 0.0962) and Sharpe ratio (0.6055 vs 0.6020), lower variance (0.0285 vs 0.0295) and kurosis (0.0034 vs 0.0040) than the mean variance efficient portfolios, on average.[11] Even in terms of standard deviation, the mean variance kurtosis efficient portfolios are superior to the mean variance efficient portfolios along all of the indicators.

[11] Note that we use excess equilibrium return, so the Sharpe ratio can be intuitively calculated by dividing the mean return by the volatility.

For example, the standard deviation of the mean variance efficient portfolios' mean return is 0.0265, higher than that of the mean variance kurtosis efficient portfolios' 0.0237. Panel C reports the value difference, and the mean improvement and standard deviation reduction by the mean variance kurtosis efficient portfolios are statistically significant. This corroborated the advantage of mean variance kurtosis portfolio optimization to mean variance portfolio optimization.

The general superiority of mean variance kurtosis efficient portfolios implies that the mean variance inefficient but mean variance kurtosis efficient portfolios perform relatively well to the mean variance efficient portfolios, which is confirms by Panel D. These portfolios outperform the mean variance efficient portfolios at almost all of the indicators: higher mean value, lower standard deviation, higher minimum, median, and maximum values, and higher skewness.

Figure 7 complements the summary statistics by showing the efficient frontiers in different space. In the first sub-figure, we plot the efficient frontiers in the mean variance space: the population mean variance efficient frontier by blue curve, the Dirichlet mean variance frontier portfolios by red circles, and the Dirichlet mean variance kurtosis frontier portfolios by yellow points. We see that the three frontiers overlap to much extent, indicating the efficacy of Dirichlet efficient frontiers. At the two ends of low mean return and high mean return parts, the Dirichlet mean variance efficient frontier is identical to the Dirichlet mean variance kurtosis efficient frontier. At the upper-middle part, the mean variance kurtosis efficient portfolios also stroll around the mean variance efficient portfolios.

To reaffirm these observations, we exhibit the mean variance kurtosis efficient frontier in the second sub-figure. We can easily see the trade-off between return and risk: the higher expected return, the higher variance; also, the higher expected return, the higher kurtosis. This mean variance kurtosis efficient frontier is similar to the mean variance efficient frontier exhibition especially at the two ends, while in the middle section several points stroll around the trend at the same expected return level, which is about the trade-off between variance and kurtosis.

In fact, the relationship between variance and kurtosis for the efficient portfolios is non-monotonic, as revealed by Figure 8. Consistent to previous observation, the variance and kurtosis are generally positively correlated. In the middle part, the same level of variance corresponds to several portfolios with different kurtosis. More than that, the non-monotonicity is more obvious if we zoom out at the first 500 efficient portfolios, shown by the blue points in the upper-left part. When the mean return is at the very low end, variance and kurtosis is negatively correlated as the portfolio return increases. Then the correlation gets to be positive. This figure once again lends evidence that kurtosis can offer risk information different

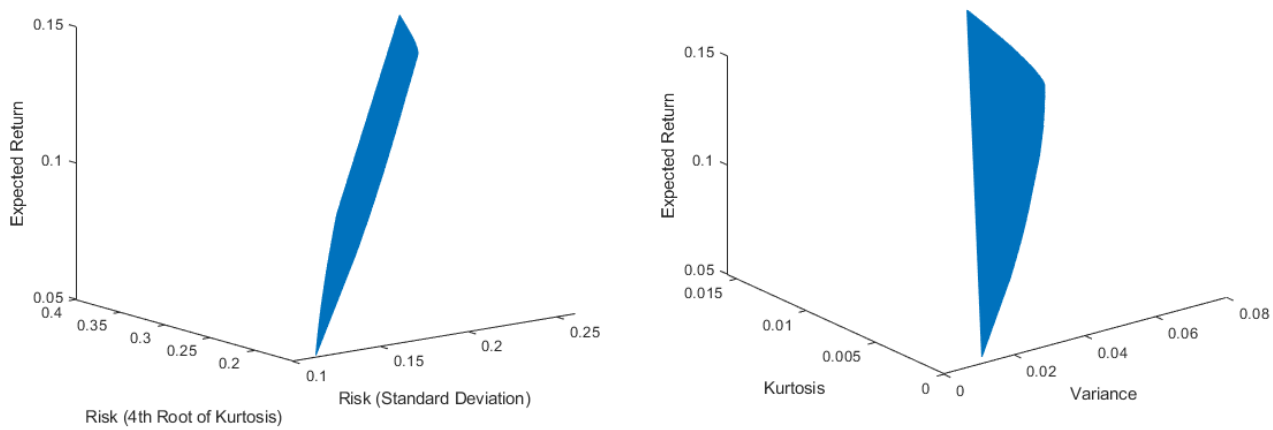

Figure 6: Dirichlet portfolios

and complementary to variance, legitimating its involvement into the portfolio optimization process. After all, if variance and kurtosis are strictly monotonically related, there is no need to include kurtosis since it is redundant given the presence of variance. Although it is less likely that the investors choose efficient portfolios with minimal return as optimal, the non-monotonicity around the minimum return still invites attention since some extremely risk averse investors can choose the minimum variance/risk portfolio as their optimal investment.

The inference between efficient portfolios' variance and kurtosis is: generally variance and kurtosis give similar risk information, but they also give different risk information especially when the expected return is medium high, or extremely low.

5.4 Efficient portfolio maximizing Sharpe ratio

It is still a question that how the investors choose their optimal portfolios among the mean variance kurtosis efficient portfolios. In the mean variance space, the optimal choice is the efficient portfolio with the highest Sharpe ratio. The tradeoff between return and risk is straightforward by balancing the expected return and volatility. In the mean variance kurtosis space, we need a generalized Sharpe ratio to characterize such a tradeoff, as illustrated in Figure 10. As variance and kurtosis are two risk indicators, we can make a synthesized measure to balance these risk aspects. In the risk plane, i.e., the variance and kurtosis plane, the vector addition is convenient for risk synthesis. Now we define the generalized Sharpe ratio SR in the mean variance kurtosis space as

$$SR = \frac{\mu}{\sqrt{\alpha_1^2\sigma^2 + \alpha_2^2\kappa^2}}.$$

If we consider a simple case of $\alpha_1 = 1$ and $\alpha_2 = 0$, the generalized Sharpe ratio is the simple Sharpe ratio. If $\alpha_1 = \alpha_2 = 1$, the generalized Sharpe ratio is $\frac{\mu}{\sqrt{\sigma^2+\kappa^2}}$. Here we

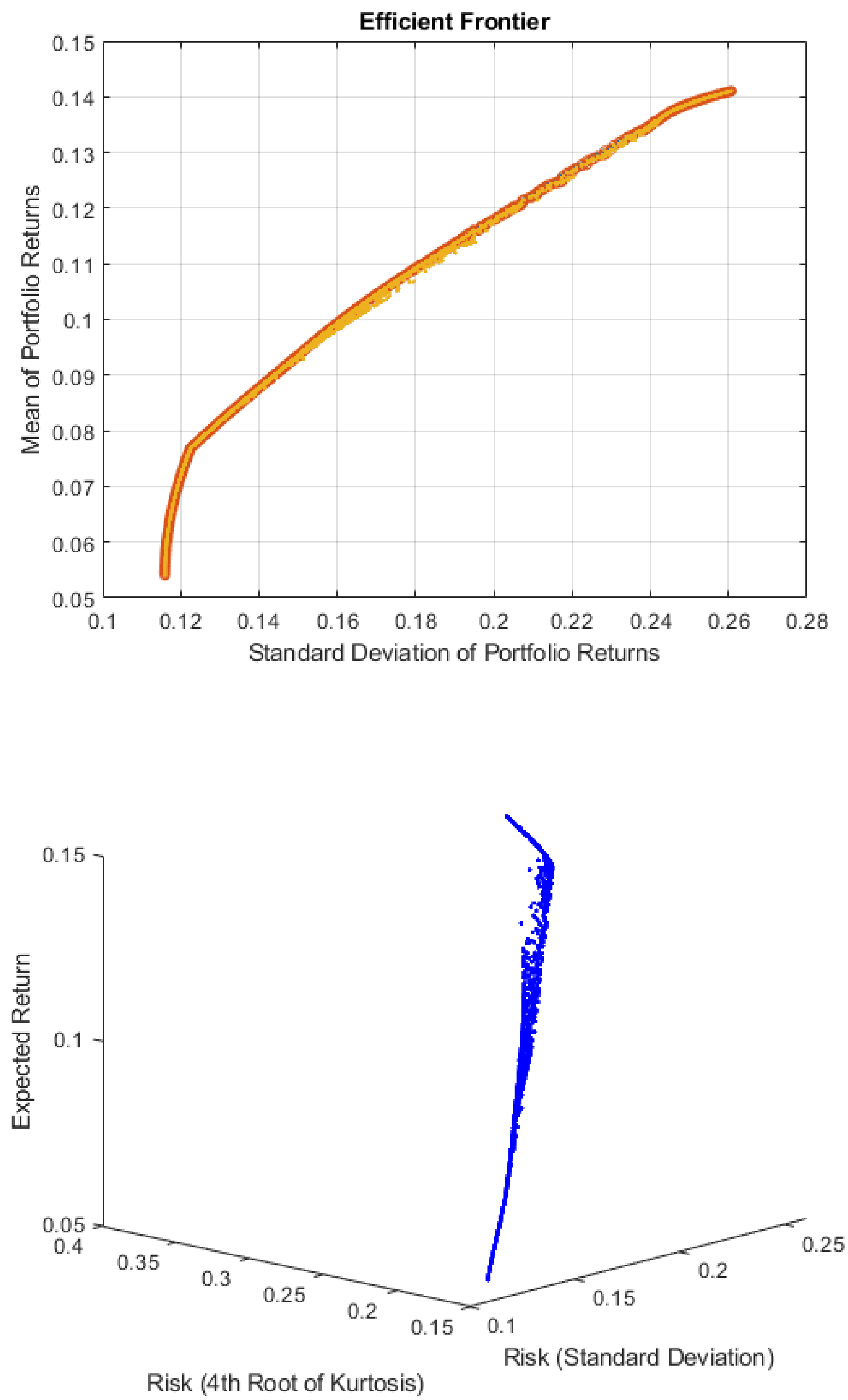

Figure 7: Mean variance kurtosis efficient portfolios

Table 5: Mean variance and mean variance kurtosis efficient portfolio

This table reports the summary statistics for the 3,499 mean variance efficient portfolios and the 4,873 mean variance kurtosis efficient portfolios. For each efficient portfolio, we calculate its mean return, variance, kurtosis, and Sharpe ratio. For each indicator, we examine the efficient portfolios by their mean value, standard deviation, minimum value, median, maximum value, skewness, and kurtosis. Panel A reports these indicators for the mean variance efficient portfolios, and Panel B reports these indicators for the mean variance kurtosis efficient portfolios. Panel C gives the indicator value difference of the mean variance kurtosis efficient portfolios over the mean variance efficient portfolios. We use t test for the mean value and F test for the variance as to the statistical significance. *** denotes significance at the 1% level, and ** denotes significance at the 5% level. Panel D reports statistics for the 1,374 MV inefficient but MVK efficient portfolios.

	Mean	StD	Min	Median	Max	Skewness	Kurtosis
Panel A: Mean variance efficient portfolios							
Mean	0.0962	0.0265	0.0541	0.0794	0.1410	0.8188	2.0446
Variance	0.0295	0.0210	0.0134	0.0160	0.0679	1.0822	2.3632
Kurtosis	0.0040	0.0057	0.0006	0.0007	0.0157	1.3538	3.0208
Sharpe ratio	0.6020	0.0372	0.4666	0.6251	0.6284	-1.1279	2.8461
Panel B: Mean variance kurtosis efficient portfolios							
Mean	0.0964	0.0237	0.0541	0.0925	0.1410	0.8841	2.4528
Variance	0.0285	0.0187	0.0134	0.0220	0.0679	1.3067	3.0834
Kurtosis	0.0034	0.0050	0.0006	0.0012	0.0157	1.7569	4.4065
Sharpe ratio	0.6055	0.0335	0.4666	0.6232	0.6284	-1.4179	3.7523
Panel C: Value difference of MV and MVK efficient portfolios							
Mean	0.0001	−0.0028***	0.0000	0.0131	0.0000	0.0654	0.4082
Variance	−0.0010**	−0.0023***	0.0000	0.0060	0.0000	0.2245	0.7202
Kurtosis	−0.0006***	−0.0007***	0.0000	0.0006	0.0000	0.4031	1.3857
Sharpe ratio	0.0035***	−0.0038***	0.0000	-0.0019	0.0000	-0.2900	0.9062
Panel D: MVK efficient portfolios beyond MV efficiency							
Mean	0.0967	0.0142	0.0769	0.0932	0.1410	1.1844	4.5881
Variance	0.0259	0.0106	0.0150	0.0224	0.0679	2.0081	6.7356
Kurtosis	0.0019	0.0019	0.0006	0.0013	0.0157	3.4756	18.3923
Sharpe ratio	0.6145	0.0182	0.5410	0.6229	0.6280	-2.0892	6.4940

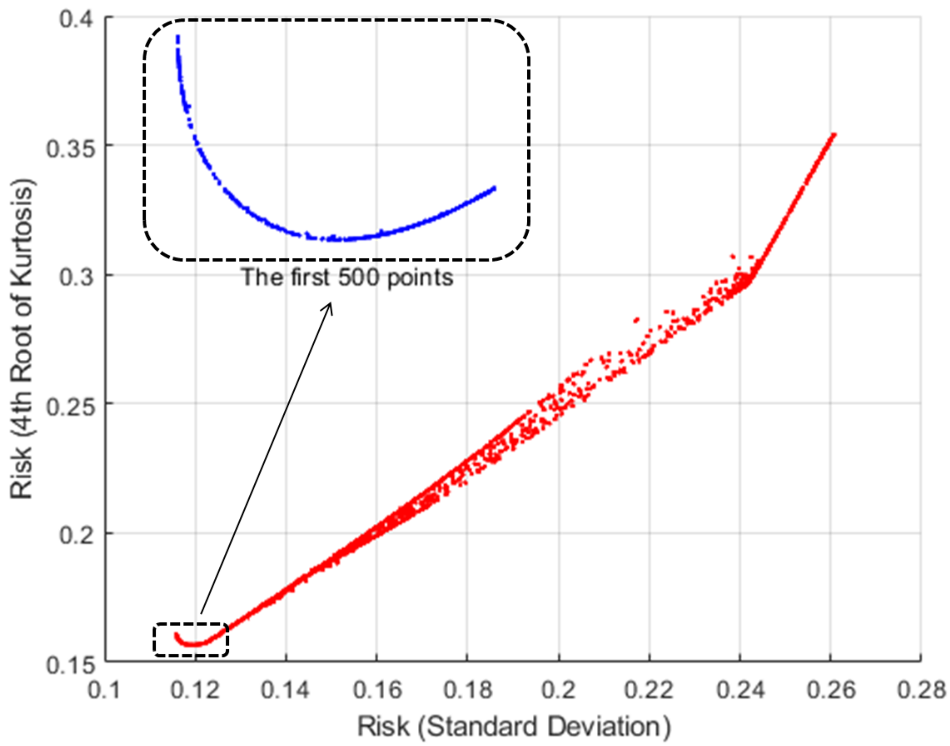

Figure 8: Non-monotonic risk relationship for the efficient portfolios

give variance and kurtosis the same weight to synthesize risk. The diagonal of variance and kurtosis square is the synthesized risk indicator, and the return indicator is vertical to it. Thus, the generalized Sharpe ratio is the tangent ratio of return over risk, consistent to the exhibition of Sharpe ratio in the mean variance space.

Note that we can easily manipulate the mixture of variance and kurtosis in the risk synthesis. We give more weight to kurtosis than variance by setting $\alpha_1 = 1$ and $\alpha_2 > 1$. The return indicator is still vertical to it, and the tangent explanation for the generalized Sharpe ratio does not change.

The mean variance kurtosis efficient set is more comprehensive than the mean variance efficient set for optimal portfolio choice. Table 6 reports top 10 efficient portfolios that maximize the Sharpe ratio in the mean variance efficient set and mean variance kurtosis efficient set respectively. We see that the two sets provide the same efficient portfolios for Sharpe ratio maximization, and these portfolios are both mean variance efficient and mean variance kurtosis efficient. This is reasonable because mean variance efficient set is a subset of mean variance kurtosis efficient set, and the Sharpe ratio prefers those portfolios with higher mean and lower variance, which are exactly the mean variance efficient portfolios. Therefore, we have consistent optimal portfolio from the mean variance efficient set and mean variance kurtosis efficient set.

The advantage of mean variance kurtosis portfolio optimization is that, when the Sharpe ratio involves kurtosis, the investors can choose mean variance inefficient but mean variance kurtosis efficient portfolio as the optimal choice. Table 7 shows that, when the investors care more about kurtosis in the Sharpe ratio evaluation, they are more likely to choose the mean variance inefficient portfolios as their optimal choice. When we give the same weight to variance and kurtosis, the Sharpe ratio maximizing portfolios are still mean variance efficient, illustrated by Panel A. However, when the weight for kurtosis increases, the mean variance inefficient portfolios tend to maximize the generalized Sharpe ratio. Panel B suggests than when the weight of kurtosis doubles that of variance, the mean variance inefficient but mean variance kurtosis efficient portfolio gives the highest generalized Sharpe ratio. Therefore, among all the mean variance kurtosis efficient portfolios the investors choose this one as the optimal choice, although it is mean variance inefficient. More strikingly, if the weight of kurtosis triples that of variance, the top 10 mean variance kurtosis efficient portfolios that maximize the generalized Sharpe ratio are mainly mean variance inefficient. This is not necessary that the optimal portfolio must have smaller kurtosis, despite we give more weight to penalize it. As shown in Panel C, the optimal choice in the mean variance kurtosis efficient set gives higher variance (0.020841 vs 0.020429) and kurtosis (0.001091 vs 0.001081) than the optimal choice in the mean variance efficient set. It is because the former one has better risk return tradeoff, with a higher mean return (0.089762 vs 0.089420). In fact, for the 10 mean variance kurtosis efficient portfolios which maximize the generalized Sharpe ratio, only 2 out of them are mean variance efficient. This once again confirms the necessity of mean variance kurtosis portfolio optimization: this optimization can give the same answer as the mean variance portfolio optimization does, like what Table 6 and Table 7 Panel A show; this optimization can also give an answer beyond the mean variance portfolio optimization, like what Table 7 Panel B and C show.

The advantage of mean variance kurtosis optimization is ostensibly mediocre at first glance, since the Sharpe ratio improvement in Table 7 seems minimal. However, by compounding over years, the difference is significant. We consider the case of Panel C in Table 7 when α_1=1 and α_2=3, and see how the mean variance kurtosis optimal portfolio outperforms its corresponding mean variance optimal portfolio. Figure 9 plots the cumulative return surplus of the MVK efficient portfolio with generalized Sharpe ratio of 0.259187 over the MV efficient portfolio with generalized Sharpe ratio of 0.259125. Suppose the investor has an initial investment of \$1 on the two portfolios at the very beginning of 1995, and we first look at how they perform with excess equilibrium returns across the whole period from 1995 to 2015. The first subfigure shows that, the MVK optimal portfolio is always better than

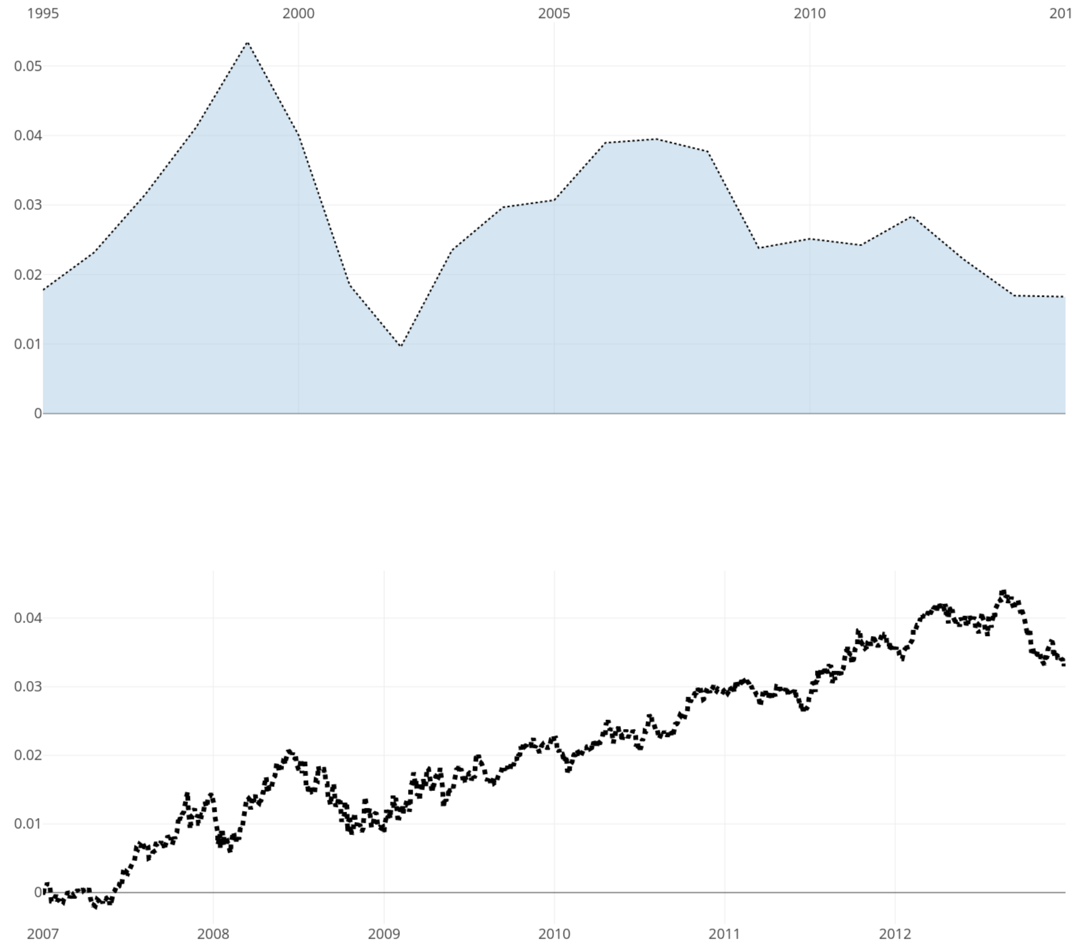

Figure 9: Cumulative return surplus of mean variance kurtosis optimal portfolio: α_1=1, α_2=3

the MV optimal portfolio across the sample period. Even during the Internet bubble bust around 2001 and the global financial crisis around 2008, the MVK optimal portfolio still have an apparent higher return, and the return surplus is 1.7% at the end. We then see how the two portfolio perform with actual excess returns during the global financial crisis from 2007 to 2012. The second subfigure shows that, the MVK optimal portfolio return surplus grows over time, and ends up with 3.4% at 2012. This is evidence that the MVK optimal portfolio is more robust than the MV optimal portfolio especially when extreme risk prevails.

Till now we consider the origin (0,0,0), as in Figure 10, as the reference point. This is due to the fact that we take the implied excess equilibrium return as the input for our portfolio optimization. However, it is possible that the interest rate for risk free asset experience a sudden change, so that the reference point can deviate from the origin at the time following portfolio construction. Figure 11 depicts the market yields on US Treasury securities at 3-month, 1-year, and 10-year. Here we primarily look at the 3-month T-bill, which is one of the most referred riskfree investment. We see that the market yield is quite volatile especially before the global financial crisis. The yield is 0.84% on Sep 16, 2008 then decreases to 0.03% on Sep 17, 2008, and then increases by 0.76% just one day later. Similarly, the yield in the first trading day of 2007 is 5.07%, of 2008 is 3.26%, of 2009 is 0.08%.

We can also examine the impact of sudden riskfree rate change on portfolio optimization by reallocate the reference point. A Sharpe ratio variant in the mean variance kurtosis space SR^* is

$$SR^* = \frac{\mu - \delta_1}{\sqrt{(\sigma - \delta_2)^2 + (\kappa - \delta_3)^2}}.$$

In this case, we give variance and kurtosis same weight to synthesize risk. Table 8 shows how the reference point change may influence the investors' optimial portfolio choice within the mean variance kurtosis efficient set. A mild riskfree rate increase can tilt the optimal choice towards the mean variance kurtosis efficient but mean variance inefficient portfolios. Panel A suggests that, when the reference point is $(0.007, 0, 0)$, the investors prefer an mean variance inefficient portfolio achieving the best adjusted Sharpe ratio. Given that the yields in the first trading days of 2017 and 2018 are 0.53% and 1.44% respectively, an increase of 0.7% in riskfree return is quite likely. Similarly, an increase in variance can make an mean variance inefficient portfolio optimal, as shown by Panel B. The combination effect is also significant, as indicated by Panel C. Here, the reference point is an moderate unexpected riskfree return change with $(0.01, 0.02, 0.02)$. Once again, the best mean variance inefficient portfolio is superior to the best mean variance efficient portfolio. In the first three panels, we talk about non-decrease cases for the riskfree return. In the case of riskfree asset return decrease, we have similar conclusion. Strikingly, for a reference point of $(-0.04, 0.2, 0.2)$, none of the top 5 mean variance kurtosis efficient portfolios with the highest adjusted Sharpe ratios are mean variance efficient. Remind that such an decrease is realistic: the yields on Mar 19 in 2007 and 2008 are 5.06% and 0.61% respectively, and the decrease is even greater than 4%.

In this part, we show that the mean variance portfolio optimization is more comprehensive than the mean variance portfolio optimization. The mean variance efficient set is a subset of mean variance kurtosis efficient set, and the two portfolio optimization methods gives the same optimal choice in terms of Sharpe ratio maximization. Besides, the mean variance kurtosis portfolio optimization can give different optimal choice from the mean variance portfolio optimization, when kurtosis gains greater weight in the risk synthesis than variance. Kurtosis is not a monotonic function of variance, so it can provide additional risk aspect to variance. Therefore, the inclusion of kurtosis can change the optimal choice. The optimal

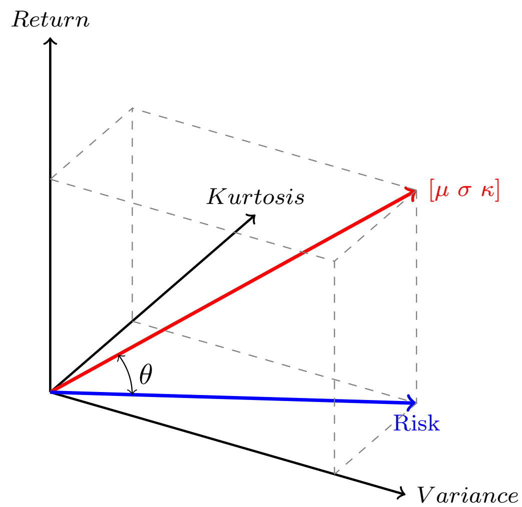

Figure 10: Generalized Sharpe ratio

reallocation does not necessarily exhibit as kurtosis minimization. The mean variance kurtosis portfolio optimization looks for the best combination of mean return, variance, and kurtosis, so it is possible that an efficient portfolio with higher variance and/or kurtosis maximizes the generalized Sharpe ratio.

5.5 Robustness check

The value of mean variance kurtosis portfolio optimization is based on the fact that kurtosis is not a monotonic function of variance. We implement the mean variance kurtosis portfolio optimization by use of Pareto improvement method and Dirichlet simulations, and we use the ten economic sectors of the S&P 500 index as the underlying assets.

Some may argue that the Dirichlet efficient frontier can be biased from the population efficient frontier, so the non-monotonic relationship between variance and kurtosis exhibited in Figure 8 can be misleading. The actual risk relationship should be strictly monotonic across the plane, and kurtosis can barely provide any new risk information not covered by variance. Therefore, the mean variance kurtosis portfolio optimization is invalid, or unnecessary.

To reaffirm the critical risk relationship between variance and kurtosis, we make a robustness check. Now we consider a simple case of two assets, where the feasible portfolio set can be nearly perfectly examined. Suppose we have two economic sectors, Energy and Materials, as our underlying assets. We use two methods to approach the feasible portfolio set. First, we still adopt the Dirichlet simulation method, and for each concentration parameter in [0.01 0.1 0.2 ... 5], we produce 100 Dirichlet portfolios. Therefore, we have 5,100 Dirichlet portfolios in total. Second, we use the grid method and consider an investment weight for Energy w_1 from 0 to 1, with a step of 0.001. Correspondingly, the weight for Materials w_2 is

Data are from the Federal Reserve "Selected Interest Rates - H.15." The break is due to data absence by the Federal Reserve.

Figure 11: Interest rates by US Treasury securities

Table 6: MV and MVK efficient portfolios maximizing Sharpe ratio

This table details the top 10 efficient portfolios that maximize the simple Sharpe ratio in the mean variance efficient set and the mean variance kurtosis efficient set, respectively. The simple Sharpe ratio is defined as mean return over volatility. The last column of "$\in$ MV" is a dummy variable of the corresponding mean variance efficient portfolio and the mean variance kurtosis efficient portfolio. If the mean variance kurtosis efficient portfolio is mean variance efficient at the same time, it takes the value of 1, and 0 otherwise.

Rank	MV efficient portfolios			MVK efficient portfolios			
	Mean	Variance	Sharpe ratio	Mean	Variance	Sharpe ratio	$\in$ MV
1	0.080668	0.016480	0.628370	0.080668	0.016480	0.628370	1
2	0.080690	0.016489	0.628370	0.080690	0.016489	0.628370	1
3	0.080750	0.016514	0.628369	0.080750	0.016514	0.628369	1
4	0.080804	0.016536	0.628369	0.080804	0.016536	0.628369	1
5	0.080571	0.016441	0.628369	0.080571	0.016441	0.628369	1
6	0.080813	0.016540	0.628369	0.080813	0.016540	0.628369	1
7	0.080460	0.016396	0.628368	0.080460	0.016396	0.628368	1
8	0.080419	0.016379	0.628367	0.080419	0.016379	0.628367	1
9	0.080384	0.016365	0.628366	0.080384	0.016365	0.628366	1
10	0.080993	0.016614	0.628365	0.080993	0.016614	0.628365	1

Table 7: MV and MVK efficient portfolios maximizing generalized Sharpe ratio

This table details the top 10 efficient portfolios that maximize the generalized Sharpe ratio in the mean variance efficient set and the mean variance kurtosis efficient set, respectively. The generalized Sharpe ratio is defined as $SR = \frac{\mu}{\sqrt{\alpha_1^2\sigma^2+\alpha_2^2\kappa^2}}$, and the weight for kurtosis increases in the three panels. The last column of "$\in$ MV" is a dummy variable of the corresponding mean variance efficient portfolio and the mean variance kurtosis efficient portfolio. If the mean variance kurtosis efficient portfolio is mean variance efficient at the same time, it takes the value of 1, and 0 otherwise.

Panel A: α_1=1, α_2=1

Rank	MV efficient portfolios				MVK efficient portfolios				
	Mean	Variance	Kurtosis	Sharpe ratio	Mean	Variance	Kurtosis	Sharpe ratio	$\in$ MV
1	0.085425	0.018537	0.000904	0.387484	0.085425	0.018537	0.000904	0.387484	1
2	0.085565	0.018601	0.000910	0.387484	0.085565	0.018601	0.000910	0.387484	1
3	0.085353	0.018504	0.000901	0.387483	0.085353	0.018504	0.000901	0.387483	1
4	0.085271	0.018467	0.000898	0.387483	0.085271	0.018467	0.000898	0.387483	1
5	0.085691	0.018658	0.000915	0.387483	0.085691	0.018658	0.000915	0.387483	1
6	0.085586	0.018610	0.000911	0.387483	0.085586	0.018610	0.000911	0.387483	1
7	0.085208	0.018438	0.000895	0.387482	0.085208	0.018438	0.000895	0.387482	1
8	0.085761	0.018690	0.000918	0.387482	0.085761	0.018690	0.000918	0.387482	1
9	0.085771	0.018695	0.000918	0.387482	0.085771	0.018695	0.000918	0.387482	1
10	0.085779	0.018699	0.000919	0.387482	0.085779	0.018699	0.000919	0.387482	1

Panel B: α_1=1, α_2=2.1

Rank	MV efficient portfolios				MVK efficient portfolios				
	Mean	Variance	Kurtosis	Sharpe ratio	Mean	Variance	Kurtosis	Sharpe ratio	$\in$ MV
1	0.085860	0.018756	0.000921	0.298938	0.087200	0.019503	0.000975	0.298941	0
2	0.086647	0.019100	0.000955	0.298935	0.085860	0.018756	0.000921	0.298938	1
3	0.086507	0.019036	0.000949	0.298935	0.086647	0.019100	0.000955	0.298935	1
4	0.086501	0.019033	0.000949	0.298934	0.086507	0.019036	0.000949	0.298935	1
5	0.086336	0.018956	0.000942	0.298934	0.086501	0.019033	0.000949	0.298934	1
6	0.086326	0.018952	0.000942	0.298934	0.086336	0.018956	0.000942	0.298934	1
7	0.086325	0.018951	0.000942	0.298934	0.086326	0.018952	0.000942	0.298934	1
8	0.086345	0.018961	0.000943	0.298934	0.086325	0.018951	0.000942	0.298934	1
9	0.086397	0.018984	0.000945	0.298934	0.086345	0.018961	0.000943	0.298934	1
10	0.086874	0.019208	0.000965	0.298933	0.086397	0.018984	0.000945	0.298934	1

Panel C: α_1=1, α_2=3

Rank	MV efficient portfolios				MVK efficient portfolios				
	Mean	Variance	Kurtosis	Sharpe ratio	Mean	Variance	Kurtosis	Sharpe ratio	$\in$ MV
1	0.089420	0.020429	0.001081	0.259125	0.089762	0.020841	0.001091	0.259187	0
2	0.085860	0.018756	0.000921	0.259102	0.087200	0.019503	0.000975	0.259182	0
3	0.086874	0.019208	0.000965	0.259101	0.089699	0.020816	0.001088	0.259168	0
4	0.086931	0.019235	0.000968	0.259101	0.086771	0.019336	0.000957	0.259138	0
5	0.086865	0.019204	0.000965	0.259100	0.089420	0.020429	0.001081	0.259125	1
6	0.087017	0.019275	0.000972	0.259100	0.086325	0.019065	0.000939	0.259117	0
7	0.087126	0.019327	0.000977	0.259100	0.089299	0.020416	0.001075	0.259117	0
8	0.087130	0.019328	0.000977	0.259100	0.087872	0.019806	0.001007	0.259115	0
9	0.087154	0.019340	0.000978	0.259100	0.086373	0.019001	0.000943	0.259108	0
10	0.087128	0.019327	0.000977	0.259099	0.085860	0.018756	0.000921	0.259102	1

Table 8: MV and MVK efficient portfolios maximizing adjusted generalized Sharpe ratio

This table details the top 5 efficient portfolios that maximize the adjusted generalized Sharpe ratio in the mean variance efficient set and the mean variance kurtosis efficient set, respectively. The adjusted generalized Sharpe ratio is defined as $SR^* = \frac{\mu-\delta_1}{\sqrt{(\sigma-\delta_2)^2+(\kappa-\delta_3)^2}}$, and the reference points are explicated in the three panels. The last column of "$\in$ MV" is a dummy variable of the corresponding mean variance efficient portfolio and the mean variance kurtosis efficient portfolio. If the mean variance kurtosis efficient portfolio is mean variance efficient at the same time, it takes the value of 1, and 0 otherwise.

Panel A: δ_1=0.007, δ_2=0, δ_3=0

Rank	MV efficient portfolios				MVK efficient portfolios				
	Mean	Variance	Kurtosis	Sharpe ratio	Mean	Variance	Kurtosis	Sharpe ratio	$\in$ MV
1	0.097371	0.024466	0.001537	0.358156	0.097325	0.024564	0.001523	0.358188	0
2	0.097380	0.024471	0.001537	0.358156	0.097031	0.024289	0.001513	0.358159	0
3	0.097429	0.024498	0.001541	0.358156	0.097371	0.024466	0.001537	0.358156	1
4	0.097231	0.024389	0.001527	0.358156	0.097380	0.024471	0.001537	0.358156	1
5	0.097460	0.024514	0.001543	0.358156	0.097429	0.024498	0.001541	0.358156	1

Panel B: δ_1=0, δ_2=0.12, δ_3=0

Rank	MV efficient portfolios				MVK efficient portfolios				
	Mean	Variance	Kurtosis	Sharpe ratio	Mean	Variance	Kurtosis	Sharpe ratio	$\in$ MV
1	0.083578	0.017712	0.000831	0.490727	0.084185	0.018032	0.000854	0.490791	0
2	0.083647	0.017742	0.000834	0.490727	0.083578	0.017712	0.000831	0.490727	1
3	0.083503	0.017679	0.000829	0.490727	0.083647	0.017742	0.000834	0.490727	1
4	0.083471	0.017665	0.000827	0.490726	0.083503	0.017679	0.000829	0.490727	1
5	0.083461	0.017660	0.000827	0.490726	0.083471	0.017665	0.000827	0.490726	1

Panel C: δ_1=0.01, δ_2=0.02, δ_3=0.02

Rank	MV efficient portfolios				MVK efficient portfolios				
	Mean	Variance	Kurtosis	Sharpe ratio	Mean	Variance	Kurtosis	Sharpe ratio	$\in$ MV
1	0.096878	0.024198	0.001504	0.343397	0.097325	0.024564	0.001523	0.343471	0
2	0.096969	0.024247	0.001510	0.343396	0.097031	0.024289	0.001513	0.343407	0
3	0.096700	0.024102	0.001492	0.343396	0.096878	0.024198	0.001504	0.343397	1
4	0.097037	0.024284	0.001514	0.343395	0.096969	0.024247	0.001510	0.343396	1
5	0.096584	0.024041	0.001485	0.343394	0.096700	0.024102	0.001492	0.343396	1

Panel D: δ_1=-0.04, δ_2=0.2, δ_3=0.2

Rank	MV efficient portfolios				MVK efficient portfolios				
	Mean	Variance	Kurtosis	Sharpe ratio	Mean	Variance	Kurtosis	Sharpe ratio	$\in$ MV
1	0.089353	0.020426	0.001079	0.480802	0.092423	0.022370	0.001226	0.483278	0
2	0.089420	0.020429	0.001081	0.480765	0.092859	0.022886	0.001258	0.483027	0
3	0.090051	0.020752	0.001115	0.480503	0.093047	0.022892	0.001266	0.483005	0
4	0.090019	0.020736	0.001113	0.480503	0.092523	0.022308	0.001232	0.482820	0
5	0.089971	0.020712	0.001111	0.480503	0.089762	0.020841	0.001091	0.482695	0

$1 - w_1$. Therefore, we have 1,001 grid portfolios in total.

For the two feasible portfolio sets, we conduct Pareto improvement method to ascertain mean variance kurtosis efficient portfolios. Figure 12 plots the risk relationship between variance and kurtosis for the efficient portfolios, similar to Figure 8. We see that the Dirichlet method and grid method give the same risk relationship: kurtosis is not a strictly monotonic function of variance, thus kurtosis is not a redundant risk indicator to variance. We see the shape for the two-asset case is similar to the first 500 observations in Figure 8. The risk relationship is negatively correlated at first, meaning that it is possible to trade between variance and kurtosis; then, the relationship is positively correlated, meaning that the tradeoff is more between return and risk.

This robustness check once again manifests the efficiency of Dirichlet simulation method. For small size of the underlying assets, Dirichlet simulation method give the same result as with the grid method to provide proper feasible portfolio set approximation. For large size of the underlying assets, Dirichlet method provides an efficient way to approach the feasible portfolio set, while the grid method is either impossible or cumbersome. We can enhance the efficacy of the Dirichlet feasible set with respect to the two key inferences about efficient sets, as well as the three propositions detailed in Section 4.

It is straightforward that mean variance portfolio optimization can produce an optimal portfolio that is mean variance inefficient, given that kurtosis provides different risk information and is not a monotonic function of variance. For brevity, we omit the examination of the generalized Sharpe ratio maximization. The result is the same: the mean variance kurtosis portfolio optimization is more comprehensive; it can give the same answer as well as different answer to the mean variance portfolio optimization. The higher weight for kurtosis in the risk sythesis, the more likely that the mean variance kurtosis optimal portfolio goes beyond the mean variance efficient set.

6 Conclusion

In this paper, we examine the portfolio optimization problem with the presence of extreme risk, and extend the mean variance optimization to higher order moments. Kurtosis captures the impact of extreme returns, thus we use it as a measure of extreme risks. It complements variance for a comprehensive depiction of investment risk, as investment with relatively low variance may well have relatively large kurtosis. We then consider the portfolio construction problem for those investors who are averse to extreme risks. The extreme risk aversion is instructive and useful given the presence of extreme financial events, especially during the global financial crisis and the European sovereign debt crisis recently. Therefore, we consider the portfolio development in the mean variance kurtosis space, to highlight the impact of

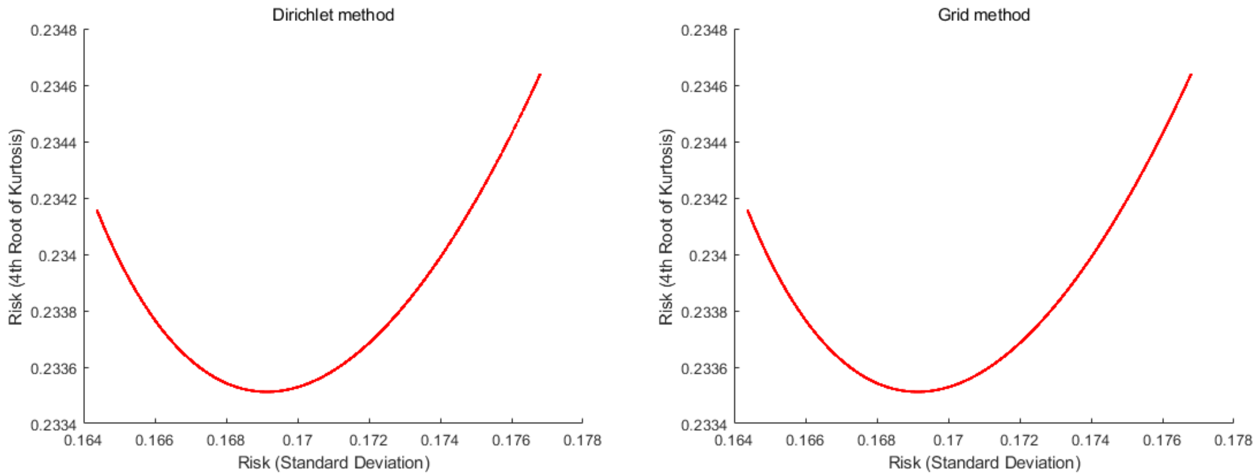

Figure 12: Non-monotonic risk relationship: 2 assets and 2 methods

extreme risks on portfolio selection.

We propose the Pareto improvement method for mean variance kurtosis efficient portfolio specification. Its intuition is to find a portfolio with some return moments better off and non moments worse off as a Pareto improvement relative to the evaluated portfolio. Compared to current methods for higher order moment portfolio optimization, i.e., the constrained variance minimization program by de Athayde and Flôres (2004) and the shortage function method by Briec, Kerstens, and Jokung (2007), the Pareto improvement method is advantageous due to its ability to detect marginal return profile improvements. This advantage is important because the failure of the two existing methods in specifying marginal improvements leads to a potential mis-classification of inefficient portfolios as efficient. Moreover, the Pareto improvement method is simple and efficient. By pairwise comparison among portfolio profiles, this method classifies the portfolios with inferior profiles as inefficient. Only those portfolios without any profile improvement are labeled as efficient.

To implement the Pareto improvement method, a proper approximation of feasible portfolio set is necessary. We approximate this set by using the Dirichlet distribution, which produces vectors of non-negative elements in sum of 1. This feature makes it a suitable program for portfolio weight simulations. We illustrate the use of the Pareto improvement method and Dirichlet simulations for mean variance kurtosis portfolio optimization with an empirical implementation of the ten economic sectors of the S&P 500 index. The market implied risk aversion parameters and equilibrium sector returns are obtained from the Black Litterman model. The equilibrium sector returns are combined with Dirichlet portfolios for Pareto improvement comparisons. The Dirichlet parameterization utilizes two properties about the mean variance and mean variance kurtosis efficient sets: the mean variance efficient portfolios are all mean variance kurtosis efficient, and the mean variance inefficient portfolios can be mean variance kurtosis efficient. With regard to these two properties, we have 5.1 million Dirichlet portfolios. We then produce the mean variance kurtosis efficient frontier, which generally behaves like line segment at the two ends, and a band spread in the middle.

To choose the optimal portfolio among the mean variance kurtosis efficient set, we propose the generalized Sharpe ratio which synthesizes variance and kurtosis into a single risk indicator. The optimal choices should be those portfolios maximizing the generalized Sharpe ratio. We show that when variance is more weighted, the mean variance kurtosis portfolio optimization produces the same optimal choice as with the mean variance portfolio optimization. However, when kurtosis is more weighted, the optimal portfolio maximizing the generalized Sharpe ratio are usually mean variance inefficient. Besides, a sudden change for the riskfree asset can also lead to a portfolio reallocation for optimality. When the riskfree return increase or decrease, the optimal choices are often the mean variance inefficient portfolios. These observations reaffirm that the mean variance kurtosis portfolio optimization is more comprehensive to the mean variance portfolio optimization.

More importantly, the application of Pareto improvement method and the Dirichlet simulations can be generalized and adapted in other settings. The Pareto improvement method provides a compassion framework for efficiency determination. Similarly, the Dirichlet approach provides a practicable way for any portfolio selection problems where an analytical solution is impossible. The combination of Pareto improvement method and the Dirichlet simulations can be easily used for mean variance skewness portfolio optimization, mean variance skewness kurtosis portfolio optimization, mean upside/downside risk portfolio optimization, and higher order stochastic dominance efficiency.

Further work will be on the mean variance kurtosis utility specification, and the tangency of utility surface and efficient frontier in the mean variance kurtosis space. The correlation between the bond market extreme risk and the stock market extreme risk is also an issue to investigate.

References

Acharya, V. V. and L. H. Pedersen (2005). Asset pricing with liquidity risk. *Journal of Financial Economics 77*(2), 375–410.

Bae, K.-H., G. A. Karolyi, and R. M. Stulz (2003). A new approach to measuring financial contagion. *The Review of Financial Studies 16*(3), 717–763.

Bakshi, G. and D. Madan (2006). A theory of volatility spreads. *Management Science 52*(12), 1945–1956.

Barber, B. M., Y.-T. Lee, Y.-J. Liu, and T. Odean (2008). Just how much do individual investors lose by trading? *The Review of Financial Studies 22*(2), 609–632.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics 121*(3), 823–866.

Bates, D. S. (2008). The market for crash risk. *Journal of Economic Dynamics and Control 32*(7), 2291–2321.

Bates, D. S. (2012). US stock market crash risk, 1926–2010. *Journal of Financial Economics 105*(2), 229–259.

Bellman, R. E. (1961). *Adaptive Control Processes: A Guided Tour*. MIT Press.

Benzoni, L., P. Collin-Dufresne, and R. S. Goldstein (2011). Explaining asset pricing puzzles associated with the 1987 market crash. *Journal of Financial Economics 101*(3), 552–573.

Black, F. and R. Litterman (1992). Global portfolio optimization. *Financial Analysts Journal 48*(5), 28–43.

Bollerslev, T. and V. Todorov (2011). Tails, fears, and risk premia. *The Journal of Finance 66*(6), 2165–2211.

Briec, W. and K. Kerstens (2010). Portfolio selection in multidimensional general and partial moment space. *Journal of Economic Dynamics and Control 34*(4), 636–656.

Briec, W., K. Kerstens, and O. Jokung (2007). Mean-variance-skewness portfolio performance gauging: A general shortage function and dual approach. *Management Science 53*(1), 135–149.

Brockett, P. L. and Y. Kahane (1992). Risk, return, skewness and preference. *Management Science 38*(6), 851–866.

Cazals, C., J.-P. Florens, and L. Simar (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics 106*(1), 1–25.

Chotikapanich, D. and W. E. Griffiths (2002). Estimating lorenz curves using a dirichlet distribution. *Journal of Business & Economic Statistics 20*(2), 290–295.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance 1*(2), 223–236.

Coval, J. and E. Stafford (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics 86*(2), 479–512.

Cremers, M., M. Halling, and D. Weinbaum (2015). Aggregate jump and volatility risk in the cross-section of stock returns. *The Journal of Finance 70*(2), 577–614.

Cvitanić, J., L. Goukasian, and F. Zapatero (2003). Monte Carlo computation of optimal portfolios in complete markets. *Journal of Economic Dynamics and Control 27*(6), 971–986.

Cvitanić, J., V. Polimenis, and F. Zapatero (2008). Optimal portfolio allocation with higher moments. *Annals of Finance 4*(1), 1–28.

Damodaran, A. (2003). *Investment Philosophies: Successful Strategies and the Investors Who Made Them Work*, Volume 185. John Wiley & Sons.

de Athayde, G. M. and R. G. Flôres (2004). Finding a maximum skewness portfolio–a general solution to three-moments portfolio choice. *Journal of Economic Dynamics and Control 28*(7), 1335–1352.

Dittmar, R. F. (2002). Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *The Journal of Finance 57*(1), 369–403.

Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture 1*(2000), 1–32.

Drechsler, I. (2013). Uncertainty,time-varying fear, and asset prices. *The Journal of Finance 68*(5), 1843–1889.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *The Journal of Finance 58*(3), 1269–1300.

Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *The Quarterly Journal of Economics 127*(2), 645–700.

Grullon, G., S. Michenaud, and J. P. Weston (2015). The real effects of short-selling constraints. *The Review of Financial Studies 28*(6), 1737–1767.

Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller (2010). Portfolio selection with higher moments. *Quantitative Finance 10*(5), 469–485.

Idzorek, T. M. (2011). A step-by-step guide to the black-litterman model. In S. Satchell (Ed.), *Forecasting Expected Returns in the Financial Markets*, Chapter 2, pp. 17–38. Academic Press.

Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance 58*(4), 1651–1683.

Jobson, J. D. and B. Korkie (1982). Potential performance and tests of portfolio efficiency. *Journal of Financial Economics 10*(4), 433–466.

Jondeau, E. and M. Rockinger (2006). Optimal portfolio allocation under higher moments. *European Financial Management 12*(1), 29–55.

Jurczenko, E., B. Maillet, and P. Merlin (2006). Hedge fund portfolio selection with higher-order moments: A nonparametric mean–variance–skewness–kurtosis efficient frontier. In E. Jurczenko and B. Maillet (Eds.), *Multi-moment Asset Allocation and Pricing Models*, Chapter 3, pp. 51–66. John Wiley & Sons, Inc.

Kerstens, K., A. Mounir, and I. Van de Woestyne (2011). Geometric representation of the mean–variance–skewness portfolio frontier based upon the shortage function. *European Journal of Operational Research 210*(1), 81–94.

Levy, M. and G. Kaplanski (2015). Portfolio selection in a two-regime world. *European Journal of Operational Research 242*(2), 514–524.

Liu, J., F. A. Longstaff, and J. Pan (2003). Dynamic asset allocation with event risk. *The Journal of Finance 58*(1), 231–259.

Longin, F. and B. Solnik (2001). Extreme correlation of international equity markets. *The Journal of Finance 56*(2), 649–676.

Malmendier, U. and S. Nagel (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics 126*(1), 373–416.

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

Mencía, J. and E. Sentana (2009). Multivariate location–scale mixtures of normals and mean–variance–skewness portfolio allocation. *Journal of Econometrics 153*(2), 105–121.

Michaud, R. O. and R. O. Michaud (2008). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation* (2 ed.). Oxford University Press.

Poon, S.-H., M. Rockinger, and J. Tawn (2003). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *The Review of Financial Studies 17*(2), 581–610.

Porter, R. B. and R. P. Bey (1974). An evaluation of the empirical significance of optimal seeking algorithms in portfolio selection. *The Journal of Finance 29*(5), 1479–1490.

Santa-Clara, P. and S. Yan (2010). Crashes, volatility, and the equity premium: Lessons from S&P 500 options. *The Review of Economics and Statistics 92*(2), 435–451.

Simar, L. and P. W. Wilson (2000). Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis 13*(1), 49–78.

Stein, J. C. (2009). Presidential address: Sophisticated investors and market efficiency. *The Journal of Finance 64*(4), 1517–1548.

Todorov, V. (2009). Variance risk-premium dynamics: The role of jumps. *The Review of Financial Studies 23*(1), 345–383.

Tversky, A. and D. Kahneman (1986). Rational choice and the framing of decisions. *Journal of Business 59*(4/2), S251–S278.

Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *The Journal of Finance 68*(3), 987–1035.