INTERNATIONAL EQUITY U.S. MUTUAL FUNDS AND DIVERSIFICATION BENEFITS

Jonathan Fletcher

University of Strathclyde

Key words: Diversification Benefits, Resampled Portfolio Efficiency^{TM}, Mutual Funds

JEL classification: G11, G12

Helpful comments received from an anonymous reviewer. I am grateful for permission to use the resampled portfolio efficiency by Richard Michaud. I am grateful for the use of return data from Ken French's Data Library.

This draft: February 2021

Address correspondence to Professor J. Fletcher, Department of Accounting and Finance, University of Strathclyde, Stenhouse Wing, 199 Cathedral Street, Glasgow, G4 0QU, United Kingdom, phone: +44 (0) 141 548 4963, fax: +44 (0) 552 3547, email: j.fletcher@strath.ac.uk

INTERNATIONAL EQUITY U.S. MUTUAL FUNDS AND DIVERSIFICATION BENEFITS

ABSTRACT

This study examines whether U.S. international equity mutual funds provide out-of-sample diversification benefits in the presence of no short selling constraints. Ignoring the economic state, international mutual funds do not provide out-of-sample diversification benefits. However international mutual funds do provide out-of-sample diversification benefits in certain economic states. The benefits are concentrated when the lag one-month U.S. Treasury Bill return is lower than normal.

I Introduction

Ever since the studies by Grubel(1968), Lessard(1973), and Solnik(1974), there has been a large number of studies that have examined the benefits of international diversification, including Li, Sarkar and Wang(2003), Hodrick and Zhang(2014), and Liu(2016) among many others. One way in which investors, especially retail investors, can access the benefits of international diversification is through investing in managed funds with international objectives, such as international equity funds. If these managed funds enhance the risk/return trade-off for domestic investors, then these international funds deliver diversification benefits.

Studies by Bekaert and Urias(1996), Errunza, Hogan and Huang(1999), Cao, Fu and Jin(2017), and Fletcher(2020) among others have examined whether international managed funds deliver diversification benefits and provide mixed evidence. Bekaert and Urias find that U.S. emerging market closed-end funds do not provide diversification benefits, but U.K. emerging market funds do. As well as looking at diversification benefits, a number of studies examine the performance of international managed funds including Cumby and Glen(1990), Busse, Goyal and Wahal(2014), and Jagannathan, Jiao and Wermers(2019) among others and again find mixed evidence of the performance ability of international fund managers.

This study revisits the issue of international diversification and examine whether U.S. international equity open-end mutual funds provide out-of-sample diversification benefits for a domestic U.S. equity investor. The study use the simulation approach, made popular by Jobson and Korkie(1981), to examine the out-of-sample diversification benefits of the international mutual funds, and examines whether these benefits exist in the presence of no short selling constraints. The domestic equity returns are adjusted for trading costs following the approach of Luttmer(1996) and De Roon, Nijman and Werker(2001).

To build the optimal portfolios, this study uses Michaud optimization which is based on resampled (RE) portfolio efficiency^{TM} (Michaud and Michaud(2008)). Michaud optimization provides a way to deal with the estimation error in the inputs to mean-variance optimization by simulating a large number of statistically equivalent efficient frontiers for a given set of mean-variance inputs. The optimal RE portfolios are then calculated by taking the average of the optimal portfolio weights across the simulated efficient frontiers. The use of Michaud optimization results in more diversified portfolios compared to the mean-variance optimal portfolios[1], and the optimal portfolio weights are a lot less sensitive to changes in the mean-variance inputs.

This study examines two main research questions. First, do U.S. international mutual funds provide out-of-sample diversification benefits? The study uses six size/book-to-market (BM) portfolios or six size/momentum portfolios as the benchmark assets in the domestic equity universe. The U.S. international equity mutual funds are sorted into seven portfolios based on the Morningstar category. The sample period runs from January 1994 and April 2019. Second, do the out-of-sample diversification benefits of international mutual funds vary across economic states? The dummy variable approach of Ferson and Qian(2004) and the lagged one-month U.S. Treasury Bill returns (Fama and Schwert(1977), Ferson(1989)) is used to allocate each month during the sample period to one of three economic states. Ferson and Qian find that the conditional performance of domestic equity funds varies across economic states[2].

This study has four main findings. First, the lagged one-month U.S. Treasury Bill return has significant predictive ability of the benchmark assets and the international mutual funds. The predictability is stronger in the international mutual funds. Second, when ignoring the economic state, international mutual funds fail to provide diversification benefits. Third,

[1] See Michaud and Michaud(2008).

[2] See also Ferson, Henry and Kisgen(2006a) and Ferson, Kisgen and Henry(2006b) on U.S. bond funds.

the out-of-sample diversification benefits vary across economic states. The benefits are concentrated when the lag one-month U.S. Treasury Bill return is lower than normal. Fourth, different international fund portfolios drive the diversification benefits of the optimal portfolios.

This study makes three main contributions to the literature. First, the paper complements studies that have examined the diversification benefits of international funds including Bekaert and Urias(1996), Errunza et al(1999), Cao et al(2017), and Fletcher(2020) among others by focusing on international equity open-end mutual funds. Second, the study extends prior literature looking at international mutual fund performance such as Cumby and Glen(1990), Busse et al(2014), and Jagannathan et al(2019) among others, by looking at the diversification benefits of international mutual funds. Third, the study complements a number of studies that find time-variation in conditional mutual fund performance such as Ferson and Qian(2004), Ferson et al(2006a) and Ferson et al(2006b), and De Souza and Lynch(2012) among others. This literature is extended by documenting that diversification benefits of international mutual funds varies across economic states.

The paper is organized as follows. Section II presents the research method. Section III describes the data used in the study. Section IV reports the empirical results. The final section concludes.

II Research Method

The Markowitz(1952) approach to optimal portfolio selection with no short selling constraints[3], using the notation of Best and Grauer(1990), assumes that the investor with a given risk tolerance (t) selects the optimal weights (x_i) for the N risky assets to:

[3] Markowitz(2005) discusses the importance of portfolio constraints as does the empirical evidence in Grauer(2009). Best and Grauer(1990) provide the mathematics of the mean-

$$\text{Max } tx'u - (1/2)x'Vx \tag{1}$$

$$\text{subject to } x'e = 1, \text{ and } x_i \geq 0 \text{ for } i=1,\ldots,N.$$

where x is the (N,1) vector of optimal weights, V is the (N,N) covariance matrix, u is a (N,1) vector of expected gross returns (1+returns), and e is a (N,1) vector of ones. When t=0, the optimal portfolio is the Global Minimum Variance (GMV) portfolio. As t increases, we move further up the mean-variance frontier until the point of the asset with the maximum expected return. The solution in equation (1) is equivalent to an alternative formulation of the mean-variance problem given by the maximum of $x'u - (\gamma/2)x'Vx$, where γ is the risk aversion level of the investor, with $t=1/\gamma$ (Grauer(2009)).

Mean-variance analysis has often been used to examine issues related to international diversification (Bekaert and Urias(1996), Hodrick and Zhang(2014), and Liu(2016) among others). In this study, the focus is on the diversification benefits of international equity mutual funds. The benchmark investment universe consists of domestic equity portfolios, and then international equity mutual fund portfolios are added to form an augmented investment universe, and then whether there is a significant shift in the mean-variance frontier is examined.

A number of empirical tests are available to examine issues of mean-variance intersection and mean-variance spanning[4] in the presence of portfolio constraints. Classical tests of mean-variance intersection include Basak, Jagannathan and Sun(2002), Briere, Drut,

---

variance problem in the presence of general linear constraints. See Best(2010) for a textbook treatment.

[4] Mean-variance intersection occurs when two mean-variance frontiers intersect at one point only. Mean-variance spanning occurs when the two frontiers are the same. Kan and Zhou(2012) provide a review of mean-variance spanning tests when only the budget constraint is imposed.

Mignon, Oosterlinck and Szafarz(2013), and De Roon et al(2001) develop the test for mean-variance spanning[5]. An alternative Bayesian approach to test mean-variance efficiency in the presence of portfolio constraints includes Wang(1998) and Li, Sarkar and Wang(2003).

The classical and Bayesian tests provide a test of in-sample diversification benefits. In this study, the simulation approach originally pioneered by Jobson and Korkie(1981)[6] is used to evaluate the out-of-sample diversification benefits provided by international mutual funds. An alternative approach would be to use back-testing portfolio strategies[7]. Michaud, Esch and Michaud(2020) argue against back-testing as results are sample specific. Using the simulation approach, a large number of different investment scenarios can be considered. A related issue is that we can require a large number of observations to find statistical significance when back-testing portfolio strategies.

The Monte Carlo simulation runs as follows. Define M as the number of assets in the benchmark investment universe and K is the number of international mutual fund portfolios, where N=M+K. First, the sample moments of the N risky assets are estimated and assumed to be the true values u_{true} and V_{true}. Second, T simulated returns are drawn from the $MVN(u_{true},V_{true})$ distribution[8], where MVN is the multivariate normal distribution and the sample moments u_s and V_s are calculated from the simulated returns. Third, the optimal

[5] Karehnke and De Roon(2020) develop a test of mean-variance skewness spanning and use the test to evaluate hedge fund performance.

[6] The simulation approach has been used in a number of studies exploring alternative ways of improving the out-of-sample performance of mean-variance strategies including Kan and Zhou(2007), Tu and Zhou(2011), and Kan, Wang and Zhou(2021) among others.

[7] The two approaches can be viewed as being complementary with one another.

[8] A multivariate t-distribution (Kan and Zhou(2017)) or bootstrapping could also be used.

portfolios in the benchmark investment universe and augmented investment universe are solved using Michaud optimization. Fourth, using the optimal weights from step 3, the mean gross return (u_{pout}), and standard deviation (σ_{pout}) of the optimal portfolios are calculated using the u_{true} and V_{true} inputs, which captures the out-of-sample return and risk of the optimal portfolios.

The optimal portfolios using Michaud optimization are solved as follows. First, T simulated returns for the N risky assets are drawn from a multivariate normal distribution using the inputs u_s and V_s. Second, the sample moments of the simulated returns are calculated, and the mean-variance frontiers are estimated based on these inputs subject to the budget constraint and no short selling constraints for both the benchmark (B) and augmented investment universes. Three different optimal portfolios from the two mean-variance frontiers are selected assuming different risk tolerance level of investors, including t=0 (GMV_B, GMV), t=0.2 ($RT1_B$, RT1), and t=0.5 ($RT2_B$, RT2). Third, steps 1 and 2 are repeated 1000 times, and the optimal GMV_B, $RT1_B$, $RT2_B$, GMV, RT1, and RT2 portfolios are given by the average weights of the 1000 corresponding simulated portfolios.

The use of Michaud optimization differs from traditional mean-variance optimization. Michaud and Michaud(2008) argue that traditional mean-variance optimization assumes that the inputs are known with certainty. Michaud optimization tries to deal with the uncertainty of the mean-variance inputs by using resampling methods to generate a large number of statistically equivalent efficient frontiers relative to the original mean-variance frontier. Rank association methods can then used to associate portfolios across the different simulated frontiers. Michaud and Michaud use expected returns for the rank association ranging from the GMV portfolio to the asset with the maximum return portfolio. An alternative approach as used in this study is to use the risk tolerance of the investor. By simulating a large number of statistically equivalent efficient portfolios, and assuming that all are equally likely from the original mean-variance inputs, the optimal RE portfolios are selected by taking the average weights of the relevant rank associated portfolios across the simulated efficient frontiers.

The optimal RE portfolios moderate the extreme weights that can arise from the original mean-variance frontier. An illustration of this is the maximum return portfolio on the original frontier is 100% in one asset. Michaud and Michaud(2008) show that the maximum return portfolio on the RE portfolio is a lot more diversified. Additional advantages of RE optimal portfolio is that they are a lot more stable and less sensitive to both the inputs used for the mean-variance optimization and to any changes in these inputs. Best and Grauer(1991a) highlight the extreme sensitivity of optimal mean-variance portfolio weights to changes in expected returns, even when there are no short selling constraints[9]. The optimal RE frontier provides a narrower range of expected return and volatility and will lie below the original mean-variance frontier implying a lower in-sample performance. Similar results will hold when compared with maximizing expected utility with higher moments as in Harvey, Liechty, Liechty and Mueller(2010) in terms of in-sample performance.

Michaud and Michaud(2008) argue that the key issue is whether RE optimization leads to better out-of-sample performance. Studies which examine the out-of-sample performance of RE optimal portfolios include Markowitz and Usmen(2003), Michaud and Michaud(2008), Harvey, Liechty and Liechty(2008), and Michaud and Esch(2018) among others with conflicting findings. Given the robustness and diversification properties of RE optimization, the tests in this study can be viewed as a conservative test of out-of-sample diversification benefits in cases where Markowitz optimization outperforms Michaud optimization (Michaud and Esch(2018)).

[9] Best and Grauer(1991b, 1992) derive the analytics of the sensitivity analysis of mean-variance problems.

Two different performance measures are used to evaluate the out-of-sample diversification benefits of the international mutual funds, which are commonly used in the performance of mean-variance trading rules (Tu and Zhou(2011), Kan et al(2020)). The first measure is based on the Sharpe(1966) measure, and is calculated as the change in Sharpe performance (DSharpe) between the optimal portfolios in the augmented universe (A) and the benchmark universe (B). The DSharpe measure is given by:

$$\text{DSharpe} = [(u_{pout} - Rf)/\sigma_{pout}]_A - [(u_{pout} - Rf)/\sigma_{pout}]_B \tag{2}$$

where Rf is the risk-free return. The mean return of the one-month U.S. Treasury Bill is used as Rf.

The second measure is based on the CER and is calculated as the change in CER performance (DCER) between the optimal portfolios in the augmented and benchmark universes. The DCER measure is given by:

$$\text{DCER} = [(u_{pout} - Rf) - (\gamma/2)\sigma^2_{pout}]_A - [(u_{pout} - Rf) - (\gamma/2)\sigma^2_{pout}]_B \tag{3}$$

where γ is the risk aversion of the investor. The DCER measure is calculated using $\gamma = 2$ and $\gamma = 5$.

The performance of the GMV and GMV_B, RT1 and $RT1_B$, and RT2 and $RT2_B$ portfolios is compared using both performance measures. The average out-of-sample DSharpe and DCER measures estimates the diversification benefits of international mutual funds. Under the null hypothesis of no diversification benefits, the average DSharpe and average DCER measure equals 0. The null hypothesis is rejected in favour of the alternative hypothesis that the measures are positive whenever the 5% and 10% percentiles of the DSharpe (DCER) measures are greater than zero. The mean out-of-sample average return (r_p) and volatility ($\sigma(r_p)$) are also calculated. Ledoit and Wolf(2017) argue that the out-of-sample volatility is a more relevant metric when comparing the performance of GMV portfolios. The simulation analysis produces the out-of-sample distribution of the optimal portfolio weights. If the international mutual funds provide diversification benefits, then would expect the international mutual fund portfolios to have the dominant weights.

The analysis so far examines the out-of-sample unconditional diversification benefits of international mutual fund portfolios. To examine whether international mutual portfolios provide conditional diversification benefits, the dummy variable approach of Ferson and Qian(2004) is used, which allocates each month in the sample period to one of a number of economic states. The out-of-sample diversification benefits of international mutual funds could vary across economic states due to time variation in the expected returns and risks of the underlying assets the funds invest in. An alternative reason is that the benefits might vary due to time variation in the skill of fund managers in international mutual funds. Additionally the impact of cash flows into and out of the fund, might require fund managers to engage in liquidity trading which tends to hurt performance (Edelen(1999)).

Ferson and Qian(2004) point out that the dummy variable approach has a number of advantages. First, it avoids the use of a linear functional form of the lagged information variables. Second, it identifies each month to a given state ex ante by only using information available prior to that month. This approach differs from using NBER recession and expansion states which are only known ex post. Third, it reduces the spurious regression bias of highly persistent lagged information variables highlighted in Ferson, Sarkissian and Simin(2003).

The economic states are identified as follows. Define z_{t-1} as the value of the lagged information variable at time t-1, x_{t-1} is z_{t-1} minus the prior 60-month average of z_{t-1}, and $\sigma(z_{t-1})$ is the standard deviation of z_{t-1} over the prior sixty months. Each month in the sample period is allocated to one of the three states depending on the value of $x_{t-1}/\sigma(z_{t-1})$. A Low state month is when $x_{t-1}/\sigma(z_{t-1}) < -1$. A High state month is when $x_{t-1}/\sigma(z_{t-1}) > 1$. A Normal state month is when $-1 < x_{t-1}/\sigma(z_{t-1}) < 1$. Three dummy variables[10] are formed, which equals 1 if a month is in the relevant state and 0 otherwise. Monte Carlo simulation approach with the Michaud optimization is then used to evaluate the out-of-sample diversification benefits in the different economic states. The sample moments from each state are used as the true values u_{true} and V_{true} in these tests.

III Data

A) Sample of International Equity Mutual Funds

The sample period runs from January 1994 and April 2019. The start date is chosen so that there is a reasonable number of mutual funds in each fund portfolio. The initial sample of international equity mutual funds is identified from Morningstar Direct using the U.S. mutual fund list with a Morningstar international equity category. The sample includes dead funds and the primary share class for each mutual fund is used. Monthly returns and Morningstar categories are collected for each fund. All fund returns prior to the fund's inception date are removed to minimize the incubation bias of Evans(2010).

Each mutual fund is allocated to one of seven investment sector categories:

1. World/Large – this sector includes the Morningstar categories of World Large Stock, Foreign Large-Value, Foreign Large-Blend, and Foreign Large-Growth.
2. World/Small – this sector includes the Morningstar categories of World Small/Mid Stock, Foreign Small/Mid-Value, Foreign Small/Mid-Blend, and Foreign Small/Mid-Blend.
3. Europe – this sector includes the Morningstar category of Europe Stock.
4. Japan – this sector includes the Morningstar category of Japan Stock.

[10] Using more than three states would reduce the number of observations in each state.

5. Asia Pacific – this sector includes the Morningstar categories of Diversified Pacific/Asia and Pacific/Asia ex-Japan Stock.

6. Emerging Markets (EM) – this sector includes the Morningstar categories of Diversified Emerging Markets, and Latin American Stock.

7. Other – this sector includes the Morningstar categories of Miscellaneous Region, China Region, and India Equity.

For each year between 1994 and 2019, an equal weighted (EW) portfolio of funds is formed within each investment sector. At the start of each year, all funds are grouped on the basis of their Morningstar category at the end of the previous year. For each group, the monthly return is then calculated on the portfolio as the average return of all funds with a return observation in that month. The use of EW portfolios is common in mutual fund performance studies e.g. Ferson and Qian(2004) and Ferson and Mo(2016). Using fund portfolios also minimizes the impact of survivorship bias and look-ahead bias (Brown, Goetzmann, Ibbotson and Ross(1992) and Carhart, Carpenter, Lynch and Musto(2002)). De Souza and Lynch(2012) point out that fund portfolios reduces the impact of the reverse survivorship bias of Linnainmaa(2013).

B) Benchmark Assets

The benchmark assets are six size/BM portfolios used in the formation of the SMB and HML factors in the Fama and French(1993) factor model. The study also uses six size/momentum portfolios later in the paper as the benchmark investment universe. The monthly returns of these portfolios are collected from Ken French's Data Library. Table 1 reports summary statistics of the benchmark assets and the mutual fund portfolios. The summary statistics include the mean, standard deviation (Std Dev,%), minimum, and maximum monthly gross returns (1+returns) of the size/BM portfolios (panel A) and the mutual fund portfolios (panel B). Panel C reports the correlations between the international mutual fund portfolios and size/BM portfolios.

Table 1 here

Panel A of Table 1 shows that the mean returns of the size/BM portfolios range between 0.798% (Small/Growth) and 1.139% (Small/Value). There is a value effect in the Small portfolios (Fama and French(2015)) and there is a size effect in the Value portfolios. The mean returns of the fund portfolios in panel B of Table 1 range between 0.270% (Japan) and 0.679% (World/Small). The EM and Other fund portfolios have the largest volatility. The mean returns of the fund portfolios are lower than the benchmark assets. The correlations between the fund portfolios and benchmark assets in panel C of Table 1 are above 0.414. The Japan fund portfolio has the lowest correlations with the benchmark assets, followed by the Other fund portfolio. The World/Large, World/Small, and Europe fund portfolios tend to have the largest correlation with the benchmark assets.

C) Lagged Information Variable

The lagged one-month U.S. Treasury Bill return is used to form the economic states, The one-month Treasury Bill return has a strong predictive ability of future returns (Fama and Schwert(1977) and Ferson(1989), Dimson, Nagel and Staunton(2016))[11], and is widely used in conditional fund performance studies such as Ferson and Schadt(1996) and Ferson and Qian(2004) among others. The Treasury Bill returns are collected from Ken French's Data Library. Table 2 reports summary statistics of the benchmark assets (panel A) and mutual fund portfolios (panel B) across the three economic states. The summary statistics include the mean

[11] See also Perez-Quiros and Timmermann(2000) and Gulen, Xing and Zhang(2011).

and standard deviation of returns within each state. A statistical test of Ferson et al(2006b) is used to examine if the mean returns between two states are equal to one another and the final three columns report *t*-statistics of the null hypothesis of the mean returns between two states are equal to one another. The *t*-statistics are calculated for the High and Low states (HL), High and Normal states (HN), and Low and Normal states (LN).

Table 2 here

Panel A of Table 2 shows that the patterns in predictability are different between the Small and Big portfolios. The Small portfolios have their highest mean returns and volatility in the Low state. The Big portfolios have their highest mean returns in the High state and highest volatility in the Low state. Perez-Quiros and Timmermann(2000) find that the relation between mean returns across states is a lot stronger in small companies as they are more affected by economic downturns. The statistical significance of the predictability in the benchmark assets is relatively weak. The mean returns of the Big/Value portfolio is significantly higher in the High state compared to the Low state. The average returns between High and Normal states are statistically significant for the Small/Value and Big/Value portfolios. At the 10% level, there is a significant difference in the mean returns between the Low and Normal states for all the Small portfolios.

Panel B of Table 2 shows that there is substantial predictability in the fund portfolios across economic states using the lag one-month Treasury Bill return. The predictability is stronger in the fund portfolios compared to the benchmark assets. The mean returns and volatility of the fund portfolios is highest in the Low state. The fund portfolios have their poorest average returns in the Normal state, except for the Japan fund portfolio. The magnitude of the predictability is greatest for the Asia Pacific, EM, and Other fund portfolios. The statistical significance of the predictability in mean returns across states is strong for the fund portfolios. The difference in mean returns between the Low and Normal states are significant for all fund portfolios. All of the fund portfolios have a significant difference in mean returns between the High and Normal states, except for the Japan fund portfolio. The difference in mean returns between the High and Low states is significant for the World/Small, Japan, and Asia Pacific, and EM fund portfolios. The predictive ability of the lag one-month Treasury Bill return is consistent with Fama and Schwert(1977), Ferson(1989), and Ferson and Qian(2004), and Dimson et al(2016) among others.

IV Empirical Results

The empirical analysis is begun by running the Michaud optimization across the whole sample period. The number of simulated return observations is set equal to the number of observations in the sample period. The mutual fund returns in this study are net of management expenses and trading costs, but the benchmark asset returns are not. The benchmark asset returns are adjusted for trading costs following the approach of Luttmer(1996) and De Roon et al(2001) assuming a proportional cost per transaction of 10 basis points for the size/BM portfolios (see Novy-Marx and Velikov(2016), Detzel, Novy-Marx and Velikov(2019))[12]. Table 3 reports summary statistics of the out-of-sample performance (panel A) and optimal portfolio weights (panel B) of the optimal RE portfolios. The summary statistics of performance in panel A include the mean, standard deviation (Std Dev), 5% and 10% percentiles, and the median of the DSharpe and DCER (%) measure. The final rows in panel A report the mean out-of-sample average return (r_p) and volatility ($\sigma(r_p)$). The summary statistics of the optimal portfolio weights in panel B include the mean and 10% percentile of

[12] Patton and Weller(2020) and Frazzini, Israel and Moskowitz(2018) provide alternative approaches to estimating trading costs.

the optimal weights in the GMV, Middle, and High portfolios. To conserve space, the Table only reports the DCER measure when $\gamma = 5$, but will discuss in the text where there are any differences.

Table 3 here

Table 3 shows that the international mutual funds do not provide out-of-sample diversification benefits. The GMV, RT1, and RT2 portfolios all have small negative mean DSharpe and DCER measures. The GMV portfolio has the largest underperformance relative to the GMV_B portfolio, which is driven by a lower mean out-of-sample average return. The GMV portfolio does however provide some portfolio risk reduction benefits with a lower mean out-of-sample volatility than the GMV_B portfolio (Ledoit and Wolf(2017)). The risk reduction benefits of the GMV portfolio is driven by the Japan fund portfolio, which has a significant positive mean weight of 0.279. The dominant benchmark assets in the GMV portfolio are the Big/Growth and Big/Neutral portfolios.

The RT1 and RT2 portfolios have similar out-of-sample performance to the $RT1_B$ and $RT2_B$ portfolios. None of the international fund portfolios have significant positive mean weights in the RT1 and RT2 portfolios in panel B of Table 3. The mean weights on all the fund portfolios are below 0.039 in the RT1 and RT2 portfolios, which explains the similar performance with the $RT1_B$ and $RT2_B$ portfolios. The Big/Growth portfolio has the largest mean weight in the RT1 portfolio but there is an increasing mean weight on the Small/Value portfolio, which has the largest mean weight in the RT2 portfolio.

Table 3 shows that international mutual funds do not provide out-of-sample diversification benefits when ignoring the economic state. The next question examined is whether the international mutual funds provide out-of-sample diversification benefits across the economic states. Given the summary statistics in Table 2, the tests are only run for the Low and High states. Tables 4 and 5 report the out-of-sample performance (Table 4) of the strategies and optimal portfolio weights (Table 5) in both the Low state (panel A) and the High state (panel B) of the optimal RE portfolios.

Table 4 here

Table 5 here

Panel A of Table 4 shows that international mutual funds provide large out-of-sample diversification benefits in the Low state. This result holds for the GMV, RT1, and RT2 portfolios. The GMV, RT1, and RT2 portfolios deliver both significantly positive mean DSharpe and DCER measures. The DCER measures are large in economic terms. The superior performance of the GMV, RT1, and RT2 portfolios are driven by both a higher mean out-of-sample average return and lower mean out-of-sample volatility compared to the GMV_B, $RT1_B$, and $RT2_B$ portfolios.

Panel A of Table 5 shows that the diversification benefits of the GMV portfolio is driven by the Japan fund portfolio with a significant positive mean weight of 0.270. The Big/Growth portfolio has the largest mean weight in the GMV portfolio. The pattern of mean weights in the RT1 and RT2 portfolios are more widely diversified than the GMV portfolio, which is a feature of Michaud optimization (Michaud and Michaud(2008)). For the RT1 and RT2 portfolios, there are number of investment sectors of mutual funds, which drive the diversification benefits. In the RT1 portfolio, there is a significant positive mean weight in World/Small, Japan, Asia Pacific, EM, and Other fund portfolios. In the RT2 portfolio, there is a significant positive mean weight in the World/Small, Asia Pacific, EM, and Other fund portfolios. The EM fund portfolio plays the biggest role in both the RT1, and RT2 portfolios.

In contrast, to the Low state, in panel B of Table 4, there is limited out-of-sample diversification benefits provided by the international mutual funds in the High state. The GMV portfolio has both a negative mean DSharpe and DCER measures but it does have a mean lower out-of-sample volatility than the GMV_B portfolio. The RT1 and RT2 portfolios have a positive mean DSharpe and DCER measures, but only the mean DCER measure for the RT1 portfolio is significant at 10% percentile. When $\gamma = 2$, the mean DCER measures for both the RT1 and RT2 portfolios at 0.0677 (RT1) and 0.081 (RT2) are significant at the 10% percentile. The optimal weights in panel B of Table 5 show that the majority of the weights are in the benchmark assets. The Japan fund portfolio has a significant positive mean weight in the GMV portfolio, and the Other fund portfolio has a large significant positive mean weight in the RT1 and RT2 portfolios.

Tables 4 and 5 show that the out-of-sample diversification benefits of international mutual funds varies across economic states. The benefits are concentrated when the lagged one-month U.S. Treasury Bill return is lower than normal. This finding might be due to the performance of the underlying assets that the fund managers invest in or due to time variation in the skill of fund managers. The analysis so far has used the size/BM portfolios as the benchmark assets. The robustness of the results is examined by using six size/momentum portfolios as the benchmark assets. The size/momentum portfolio returns are adjusted for trading costs assuming a proportional cost per transaction of 50 basis points (Novy-Marx and Velikov(2016)).

The tests in Tables 3 to 5 are repeated using the size/momentum portfolios as the benchmark assets. For the overall sample period, the international mutual fund portfolios do not provide significant out-of-sample diversification benefits. The mean DSharpe and DCER measures are now positive but are small in economic terms. Tables 6 and 7 report the summary statistics of the out-of-sample performance (Table 6), and optimal portfolio weights (Table 7) in the Low and High economic states of the optimal RE portfolios.

Table 6 here

Table 7 here

Table 6 shows that the out-of-sample diversification benefits of international mutual funds continues to vary across economic states, when the benchmark investment universe is the size/momentum portfolios. As with panel A of Table 3, the GMV, RT1, and RT2 portfolios all provide large significant out-of-sample diversification benefits in the Low state. The mean DSharpe and DCER measures are all statistically significant at the 5% percentile. The magnitude of the DCER measures is large in economic terms. The superior performance of the GMV, RT1, and RT2 portfolios relative to the GMV_B, $RT1_B$, and $RT2_B$ portfolios is driven by both a higher mean out-of-sample average return and a lower average volatility.

Panel A of Table 7 shows that different international mutual fund portfolios drive the diversification benefits across the GMV, RT1, and RT2 portfolios. The Japan fund portfolio drives the diversification benefits of the GMV portfolio, with a significant positive mean weight of 0.198. Moving to the RT1 and RT2 portfolios, the vast majority of the optimal weights are in the international mutual fund portfolios. In the RT1 portfolio, the largest mean weights are in the EM and World/Small fund portfolios, but there are also significant mean weights on Japan, Asia Pacific, and Other fund portfolios. In the RT2 portfolio, the EM fund portfolio plays a more dominant role, but there are still positive mean weights on the Japan, Asia Pacific, and Other fund portfolios above 0.1.

Panel B of Table 6 shows that the international mutual fund portfolios also deliver significant out-of-sample diversification benefits in the High state for the optimal RE portfolios with higher levels of risk tolerance. This result stands in sharp contrast to the results in Panel B of Table 4. The RT1 and RT2 portfolios deliver large significant positive mean DSharpe and DCER measures in the High state. The superior performance relative to the $RT1_B$ and $RT2_B$ portfolios is driven by the higher average out-of-sample mean return. Panel B of Table 7 shows that it is the Other fund portfolio, which drives the diversification benefits of the RT1 and RT2 portfolios with a significant positive mean weight of 0.433 (RT1), and 0.503 (RT2). There is also a significant mean weight on the EM fund portfolio. Tables 6 and 7 show that there is stronger support of the out-of-sample diversification benefits by the international mutual funds when the benchmark investment universe is the size/momentum portfolios.

V Conclusion

This study examines whether U.S. international equity mutual funds provide out-of-sample diversification benefits in the presence of no short selling constraints, and whether these benefits vary across economic states. There are four main findings from the study.

First, the lag one-month U.S. Treasury Bill return has significant predictive ability of both the benchmark assets and international mutual fund returns. The predictability is a lot stronger in the international mutual funds. The mean returns of the international mutual funds and volatility are highest in the Low state and the mean returns are lowest in the Normal state. The predictive ability of the Treasury Bill return is consistent with Fama and Schwert(1977), Ferson(1989), Ferson and Qian(2004), and Dimson et al(2016) among others.

Second, when ignoring the economic state, international mutual funds do not provide out-of-sample diversification benefits. The mean DSharpe and DCER measures are all negative for the GMV, Middle, and High portfolios using the size/BM benchmark assets. In the optimal RT1 and RT2 portfolios, there is little exposure to the international mutual fund portfolios. The only evidence of some diversification benefits is that the GMV portfolio has a lower mean out-of-sample volatility than the GMV_B portfolio, which is driven by a significant positive mean weight on the Japan fund portfolio.

Third, the out-of-sample diversification benefits of the international mutual funds varies across economic states. The international mutual funds provide large significant diversification benefits in the Low state. The GMV, RT1, and RT2 portfolios all provide significant positive mean DSharpe and DCER measures. This result holds for both benchmark investment universes. This result is consistent with Fletcher(2020) for international equity CEF. The diversification benefits are more mixed in the High state and depends upon the benchmark investment universe used. When the size/momentum portfolios are the benchmark universe, the RT1 and RT2 portfolios do provide significant positive mean DSharpe and DCER measures.

Fourth, the out-of-sample diversification benefits of the GMV, RT1, and RT2 portfolios rely on different portfolios of international mutual funds. The diversification benefits of the GMV portfolio in the Low state is driven by the Japan fund portfolio. The benefits of the RT1 and RT2 portfolios in the Low state are driven by the World/Small, Japan, Asia Pacific, EM, and Other fund portfolios, with the EM fund portfolio having the largest mean weight. For the diversification benefits in the High state of the RT1 and RT2 portfolios, using the size/momentum portfolios as the benchmark universe, the benefits are driven by the Other fund portfolio with a smaller role played by the EM fund portfolio.

This study suggests that international mutual funds do provide out-of-sample diversification benefits in the presence of no short selling constraints when the lag one-month U.S. Treasury Bill return is lower than normal. The study has used international equity open-end mutual funds. The analysis could be extended to consider alternative international managed funds such as Exchange-Traded Funds (ETF) or hedge funds. The lagged one-month Treasury Bill return has been used to form the economic states. The analysis could be extended to use alternative lagged information variables such as combination forecasts of Rapach, Strauss and Zhou(2010) or using investor sentiment indexes (Huang, Jiang, Tu and Zhou(2015)). These issues are left to future research.

Table 1 Summary Statistics of Benchmark Assets and Mutual Fund Portfolios

Panel A: Benchmark	Mean	Std Dev	Minimum	Maximum
Small/Growth	1.00798	6.68682	0.75529	1.27077
Small/Neutral	1.01092	5.17191	0.80802	1.16642
Small/Value	1.01139	5.44616	0.79502	1.17287
Big/Growth	1.00949	4.23350	0.85013	1.10047
Big/Neutral	1.00878	4.24045	0.82040	1.12386
Big/Value	1.00867	5.14361	0.77733	1.17658

Panel B: Funds	Mean	Std Dev	Minimum	Maximum
World/Large	1.00513	4.40030	0.79257	1.11995
World/Small	1.00679	4.62784	0.76882	1.13594
Europe	1.00620	5.04350	0.75026	1.16149
Japan	1.00270	5.00892	0.81511	1.14904
Asia Pacific	1.00498	5.66620	0.76525	1.18069
EM	1.00585	6.29077	0.71341	1.18126
Other	1.00594	6.51059	0.75953	1.24178

Panel C	World/Large	World/Small	Europe	Japan	Asia/Pacific	EM	Other
Small/Growth	0.771	0.788	0.716	0.500	0.675	0.695	0.618
Small/Neutral	0.795	0.800	0.752	0.475	0.670	0.697	0.626
Small/Value	0.760	0.767	0.722	0.455	0.628	0.656	0.581
Big/Growth	0.852	0.772	0.785	0.528	0.711	0.715	0.635
Big/Neutral	0.819	0.741	0.776	0.479	0.677	0.696	0.622
Big/Value	0.757	0.676	0.717	0.414	0.590	0.611	0.539

The table reports summary statistics of the benchmark assets (panel A), international mutual fund portfolios (panel B), and correlations between the benchmark assets and the mutual fund portfolios (panel C). The summary statistics include the mean, standard deviation (Std Dev) (%), minimum, and maximum monthly gross returns (1+returns) between January 1994 and April 2019. The benchmark assets include six size/BM portfolios. The international mutual funds are sorted into seven investment sectors, World/Large, World/Small, Europe, Japan, Asia Pacific, Emerging Markets (EM), and Other.

Table 2 Summary Statistics across Economic States

Panel A	Low		Normal		High				
Benchmark	Mean	Std dev	Mean	Std dev	Mean	Std dev	t(HL)	t(HN)	t(LN)
Small/Growth	1.01408	7.58840	1.00492	6.44016	1.00835	6.22351	-1.16	0.76	1.84[2]
Small/Neutral	1.01585	6.30465	1.00760	4.95541	1.01285	4.32148	-0.78	1.59	2.05[1]
Small/Value	1.01759	7.27801	1.00741	4.88703	1.01345	4.34817	-0.97	1.84[2]	2.32[1]
Big/Growth	1.00855	4.95202	1.00918	4.11632	1.01093	3.71101	0.76	0.63	-0.19
Big/Neutral	1.00858	5.07910	1.00754	4.33208	1.01135	3.03985	0.93	1.43	0.31
Big/Value	1.00277	7.12375	1.00875	4.58863	1.01389	3.76683	2.75[1]	1.73[2]	-1.41
Panel B	Low		Normal		High				
Funds	Mean	Std dev	Mean	Std dev	Mean	Std dev	t(HL)	t(HN)	t(LN)
World/Large	1.00968	5.43560	1.00177	4.37037	1.00750	3.18224	-0.69	2.11[1]	2.26[1]
World/Small	1.01626	5.46440	1.00176	4.58983	1.00790	3.61684	-2.55[1]	2.10[1]	4.06[1]
Europe	1.01338	6.17203	1.00200	5.06587	1.00780	3.55326	-1.56	1.87[2]	2.85[1]
Japan	1.01088	5.87215	1.00077	4.89471	0.99898	4.23082	-3.28[1]	-0.55	2.64[1]
Asia Pacific	1.01749	6.14775	0.99703	5.95053	1.00897	4.20895	-2.28[1]	3.27[1]	4.78[1]
EM	1.01935	6.69689	0.99630	6.52945	1.01202	5.00332	-1.75[2]	3.82[1]	4.92[1]
Other	1.01570	6.85443	0.99700	6.85229	1.01437	5.12134	-0.31	4.06[1]	3.86[1]

[1] Significant at 5%
[2] Significant at 10%

The table reports summary statistics of the benchmark assets (panel A) and international mutual fund portfolios (panel B) across economic states between January 1994 and April 2019. The economic states are when the lagged one-month U.S. Treasury Bill return is lower than normal (Low), Normal, and higher than normal (High). The summary statistics include the mean and standard deviation (Std Dev) of monthly gross returns (1+returns). The final three columns report *t*-statistics of the null hypothesis of the mean returns between two states are equal to one another. The *t*-statistics are calculated for the High and Low states (HL), High and Normal states (HN), and Low and Normal states (LN).

Table 3 Out-of-Sample Performance and Optimal Portfolio Weights

Panel A	Mean	Std Dev	5%	10%	Median
DSharpe					
GMV	-0.033	0.007	-0.047	-0.043	-0.033
RT1	-0.006	0.008	-0.025	-0.017	-0.003
RT2	-0.007	0.010	-0.032	-0.020	-0.003
DCER	Mean	Std Dev	5%	10%	Median
GMV	-0.121	0.029	-0.174	-0.157	-0.118
RT1	-0.027	0.035	-0.102	-0.071	-0.015
RT2	-0.033	0.056	-0.153	-0.088	-0.012
	r_p		$\sigma(r_p)$		
	Benchmark	Augmented	Benchmark	Augmented	
GMV	1.00820	1.00660	4.06570	3.86570	
RT1	1.00900	1.00870	4.39360	4.36620	
RT2	1.00930	1.00900	4.68620	4.67650	

Panel B	GMV		RT1		RT2	
	Mean	10%	Mean	10%	Mean	10%
Small/Growth	0	0	0.001	0	0.013	0
Small/Neutral	0.007	0	0.140	0.007	0.165	0.009
Small/Value	0.002	0	0.212	0.023	0.349	0.052
Big/Growth	0.304	0.175	0.360	0.090	0.232	0.020
Big/Neutral	0.385	0.264	0.155	0.007	0.082	0.001
Big/Value	0.001	0	0.031	0.000	0.049	0.000
World/Large	0.002	0	0	0	0	0
World/Small	0.014	0.000	0.034	0.000	0.021	0
Europe	0.000	0	0.010	0	0.014	0
Japan	0.279	0.215	0.024	0.000	0.012	0
Asia Pacific	0	0	0.002	0	0.002	0
EM	0	0	0.005	0	0.016	0
Other	0.001	0	0.021	0.000	0.039	0.000

The table reports summary statistics of the out-of-sample performance (panel A) and optimal portfolio weights (panel B) using Michaud optimization. Two sets of RE portfolios are estimated. The first set uses a benchmark investment universe consisting of the returns of six size/BM portfolios. The benchmark asset returns are adjusted for a proportional cost per transaction of 10 basis points. The second set is the augmented investment universe, which adds seven international equity mutual fund portfolios to the benchmark investment universe. Three RE portfolios are formed from the benchmark investment universe and from the augmented investment universe for risk tolerance levels of 0 (GMV, GMV_B), 0.2 (RT1, $RT1_B$), and 0.5 (RT2, $RT2_B$). The DSharpe (DCER) measure in panels A (B) is the change in Sharpe (CER,%) performance between the GMV and GMV_B, RT1 and $RT1_B$, and RT2 and $RT2_B$ strategies. The summary statistics of performance in panel A includes the average, standard deviation (Std Dev), 5% and 10% percentiles, and median of the DSharpe and DCER measures. The final rows of panel A include the average out-of-sample average return (r_p) and volatility ($\sigma(r_p)$) for the three strategies in the benchmark and augmented investment universes. The summary statistics in panel B of the optimal portfolio weights includes the mean and 10% percentile of each asset in the GMV, RT1, and RT2 portfolios. The simulation uses the sample moments of the benchmark assets and international mutual fund portfolios over the January 1994 and April 2019 sample period as the true u and V. The number of simulation trials is 1000. The DCER performance assumes a risk aversion level of 5.

Table 4 Out-of-Sample Performance across Economic States

Panel A:					
Low DSharpe	Mean	Std Dev	5%	10%	Median
GMV	0.032	0.010	0.014	0.018	0.031
RT1	0.054	0.021	0.018	0.025	0.053
RT2	0.047	0.020	0.013	0.020	0.047
DCER	Mean	Std Dev	5%	10%	Median
GMV	0.170	0.049	0.083	0.106	0.171
RT1	0.343	0.103	0.161	0.205	0.349
RT2	0.348	0.114	0.151	0.199	0.356
	r_p		$\sigma(r_p)$		
	Benchmark	Augmented	Benchmark	Augmented	
GMV	1.00760	1.00887	4.88801	4.70787	
RT1	1.01316	1.01580	6.13985	5.89271	
RT2	1.01443	1.01627	6.64510	6.13975	
Panel B:					
High DSharpe	Mean	Std Dev	5%	10%	Median
GMV	-0.074	0.028	-0.123	-0.113	-0.072
RT1	0.013	0.015	-0.015	-0.005	0.016
RT2	0.014	0.018	-0.020	-0.009	0.017
DCER	Mean	Std Dev	5%	10%	Median
GMV	-0.231	0.080	-0.368	-0.339	-0.226
RT1	0.055	0.045	-0.015	0.004	0.058
RT2	0.062	0.062	-0.031	0.000	0.062
	r_p		$\sigma(r_p)$		
	Benchmark	Augmented	Benchmark	Augmented	
GMV	1.01037	1.00772	2.99372	2.76086	
RT1	1.01173	1.01249	3.52596	3.63584	
RT2	1.01172	1.01266	3.66985	3.82659	

The table reports summary statistics of the out-of-sample performance across economic states using Michaud optimization. The economic states are when the lagged one-month U.S. Treasury Bill return is lower (Low) than normal (panel A) and higher (High) than normal (panel B). Two sets of RE portfolios are estimated. The first set uses a benchmark investment universe consisting of the returns of six size/BM portfolios. The benchmark asset returns are adjusted for a proportional cost per transaction of 10 basis points. The second set is the augmented investment universe, which adds seven international equity mutual fund portfolios to the benchmark investment universe. Three RE portfolios are formed from the benchmark investment universe and from the augmented investment universe for risk tolerance levels of 0 (GMV, GMV_B), 0.2 ($RT1$, $RT1_B$), and 0.5 ($RT2$, $RT2_B$). The DSharpe (DCER) measure is the change in Sharpe (CER,%) performance between the GMV and GMV_B, RT1 and $RT1_B$, and RT2 and $RT2_B$ strategies. The summary statistics of performance include the average, standard deviation (Std Dev), 5% and 10% percentiles, and median of the DSharpe and DCER measures. The final rows of each panel include the average out-of-sample average return (r_p) and volatility ($\sigma(r_p)$) for the three strategies in the benchmark and augmented investment universes. The simulation uses the sample moments of the benchmark assets and international mutual fund portfolios across the relevant economic state during the January 1994 and April 2019 sample period as the true u and V. The number of simulation trials is 1000. The DCER performance assumes a risk aversion level of 5.

Table 5 Optimal Portfolio Weights across Economic States

Panel A	GMV		RT1		RT2	
Low	Mean	10%	Mean	10%	Mean	10%
Small/Growth	0	0	0.030	0	0.055	0
Small/Neutral	0.002	0	0.062	0.000	0.035	0
Small/Value	0	0	0.114	0.003	0.176	0.007
Big/Growth	0.492	0.272	0.024	0	0.007	0
Big/Neutral	0.184	0.021	0.003	0	0.000	0
Big/Value	0	0	0.000	0	0.000	0
World/Large	0.001	0	0	0	0	0
World/Small	0.006	0	0.157	0.010	0.075	0.001
Europe	0.001	0	0.029	0	0.029	0
Japan	0.270	0.127	0.130	0.003	0.102	0.000
Asia Pacific	0.000	0	0.121	0.005	0.097	0.002
EM	0	0	0.207	0.017	0.277	0.036
Other	0.040	0.001	0.118	0.004	0.140	0.004
Panel B	GMV		RT1		RT2	
High	Mean	10%	Mean	10%	Mean	10%
Small/Growth	0	0	0.012	0	0.030	0
Small/Neutral	0.001	0	0.050	0.000	0.050	0.000
Small/Value	0.012	0	0.132	0.006	0.138	0.005
Big/Growth	0.030	0.000	0.082	0.001	0.060	0.000
Big/Neutral	0.563	0.422	0.058	0.000	0.028	0
Big/Value	0.026	0.000	0.294	0.044	0.262	0.023
World/Large	0.035	0.000	0.000	0	0	0
World/Small	0.000	0	0.003	0	0.002	0
Europe	0.062	0.004	0.014	0	0.008	0
Japan	0.217	0.114	0.003	0	0.002	0
Asia Pacific	0.031	0.001	0.004	0	0.002	0
EM	0.001	0	0.074	0.001	0.088	0.001
Other	0.016	0.000	0.269	0.029	0.325	0.033

The table reports summary statistics of the optimal portfolio weights across economic states using Michaud optimization. The economic states are when the lagged one-month U.S. Treasury Bill return is lower (Low) than normal (panel A), and higher (High) than normal (panel B). Two sets of RE portfolios are estimated. The first set uses a benchmark investment universe consisting of the returns of six size/BM portfolios. The benchmark asset returns are adjusted for a proportional cost per transaction of 10 basis points. The second set is the augmented investment universe, which adds seven international equity mutual fund portfolios to the benchmark investment universe. Three RE portfolios are formed from the benchmark investment universe and from the augmented investment universe for risk tolerance levels of 0 (GMV, GMV_B), 0.2 (RT1, $RT1_B$), and 0.5 (RT2, $RT2_B$). The summary statistics of the optimal portfolio weights includes the mean and 10% percentile of each asset in the GMV, RT1, and RT2 portfolios. The simulation uses the sample moments of the benchmark assets and international mutual fund portfolios across the relevant economic state and during the January 1994 and April 2019 sample period as the true u and V. The number of simulation trials is 1000. The DCER performance assumes a risk aversion level of 5.

Table 6 Out-of-Sample Performance across Economic States: Size/Momentum Portfolios

Panel A:					
Low DSharpe	Mean	Std Dev	5%	10%	Median
GMV	0.045	0.019	0.014	0.020	0.044
RT1	0.090	0.030	0.040	0.053	0.090
RT2	0.072	0.027	0.022	0.035	0.074
DCER	Mean	Std Dev	5%	10%	Median
GMV	0.216	0.085	0.075	0.103	0.216
RT1	0.588	0.161	0.324	0.393	0.588
RT2	0.653	0.249	0.292	0.369	0.621
	r_p		$\sigma(r_p)$		
	Benchmark	Augmented	Benchmark	Augmented	
GMV	1.00308	1.00507	4.541078	4.465444	
RT1	1.01074	1.01555	6.265293	5.939844	
RT2	1.01288	1.01606	7.409988	6.504697	
Panel B:					
High DSharpe	Mean	Std Dev	5%	10%	Median
GMV	-0.011	0.023	-0.048	-0.041	-0.010
RT1	0.106	0.031	0.040	0.058	0.113
RT2	0.115	0.033	0.044	0.063	0.125
DCER	Mean	Std Dev	5%	10%	Median
GMV	-0.025	0.066	-0.132	-0.111	-0.024
RT1	0.403	0.119	0.165	0.233	0.422
RT2	0.478	0.140	0.197	0.288	0.495
	r_p		$\sigma(r_p)$		
	Benchmark	Augmented	Benchmark	Augmented	
GMV	1.00506	1.00467	2.95920	2.86320	
RT1	1.00574	1.01057	3.72187	4.12699	
RT2	1.00573	1.01117	4.07413	4.39219	

The table reports summary statistics of the out-of-sample performance across economic states using Michaud optimization. The economic states are when the lagged one-month U.S. Treasury Bill return is lower (Low) than normal (panel A) and higher (High) than normal (panel B). Two sets of RE portfolios are estimated. The first set uses a benchmark investment universe consisting of the returns of six size/momentum portfolios. The benchmark asset returns are adjusted for a proportional cost per transaction of 50 basis points. The second set is the augmented investment universe, which adds seven international equity mutual fund portfolios to the benchmark investment universe. Three RE portfolios are formed from the benchmark investment universe and from the augmented investment universe for risk tolerance levels of 0 (GMV, GMV_B), 0.2 (RT1, $RT1_B$), and 0.5 (RT2, $RT2_B$). The DSharpe (DCER) measure is the change in Sharpe (CER,%) performance between the GMV and GMV_B, RT1 and $RT1_B$, and RT2 and $RT2_B$ strategies. The summary statistics of performance include the average, standard deviation (Std Dev), 5% and 10% percentiles, and median of the DSharpe and DCER measures. The final rows of each panel include the average out-of-sample average return (r_p) and volatility ($\sigma(r_p)$) for the three strategies in the benchmark and augmented investment universes. The simulation uses the sample moments of the benchmark assets and international mutual fund portfolios across the relevant economic state during the January 1994 and April 2019 sample period as the true u and V. The number of simulation trials is 1000. The DCER performance assumes a risk aversion level of 5.

Table 7 Optimal Portfolio Weights across Economic States: Size/Momentum Portfolios

Panel A	GMV		RT1		RT2	
Low	Mean	10%	Mean	10%	Mean	10%
Small/Losers	0	0	0.047	0.000	0.152	0.003
Small/Neutral	0.000	0	0.029	0	0.016	0
Small/Winners	0.000	0	0.092	0.001	0.086	0.000
Big/Losers	0	0	0.001	0	0.001	0
Big/Neutral	0.472	0.262	0.003	0	0.000	0
Big/Winners	0.287	0.095	0.010	0	0.003	0
World/Large	0.000	0	0	0	0	0
World/Small	0.003	0	0.181	0.015	0.087	0.002
Europe	0.000	0	0.031	0	0.029	0
Japan	0.198	0.073	0.133	0.003	0.107	0.000
Asia Pacific	0.000	0	0.131	0.007	0.102	0.003
EM	0.000	0	0.219	0.018	0.273	0.031
Other	0.033	0.001	0.118	0.003	0.139	0.003
Panel B	GMV		RT1		RT2	
High	Mean	10%	Mean	10%	Mean	10%
Small/Losers	0	0	0.013	0	0.030	0
Small/Neutral	0.042	0.001	0.103	0.004	0.071	0.001
Small/Winners	0.000	0	0.038	0.000	0.056	0.000
Big/Losers	0.003	0	0.067	0.001	0.065	0.000
Big/Neutral	0.619	0.463	0.040	0.000	0.016	0
Big/Winners	0.024	0.000	0.036	0	0.027	0
World/Large	0.032	0.000	0.013	0	0.003	0
World/Small	0.000	0	0.030	0	0.017	0
Europe	0.070	0.004	0.074	0.000	0.045	0
Japan	0.161	0.062	0.007	0	0.005	0
Asia Pacific	0.019	0.000	0.011	0	0.005	0
EM	0.001	0	0.128	0.004	0.149	0.004
Other	0.022	0.000	0.433	0.087	0.503	0.106

The table reports summary statistics of the optimal portfolio weights across economic states using Michaud optimization. The economic states are when the lagged one-month U.S. Treasury Bill return is lower (Low) than normal (panel A), and higher (High) than normal (panel B). Two sets of RE portfolios are estimated. The first set uses a benchmark investment universe consisting of the returns of six size/momentum portfolios. The benchmark asset returns are adjusted for a proportional cost per transaction of 50 basis points. The second set is the augmented investment universe, which adds seven international equity mutual fund portfolios to the benchmark investment universe. Three RE portfolios are formed from the benchmark investment universe and from the augmented investment universe for risk tolerance levels of 0 (GMV, GMV_B), 0.2 (RT1, $RT1_B$), and 0.5 (RT2, $RT2_B$). The summary statistics of the optimal portfolio weights includes the mean and 10% percentile of each asset in the GMV, RT1, and RT2 portfolios. The simulation uses the sample moments of the benchmark assets and international mutual fund portfolios across the relevant economic state and during the January 1994 and April 2019 sample period as the true u and V. The number of simulation trials is 1000. The DCER performance assumes a risk aversion level of 5.

References

Basak, G., Jagannathan, R. and G. Sun, 2002, A direct test for the mean-variance efficiency of a portfolio, *Journal of Economic Dynamics and Control*, 26, 1195-1215.

Bekaert, G. and Urias, M.S., 1996, Diversification, integration and emerging market closed-end funds, *Journal of Finance*, 51 835-869.

Best, M.J., 2010, Portfolio optimization, *Chapman and Hall (CRC)*, New York.

Best, M.J. and R.R. Grauer, 1990, The efficient set mathematics when mean-variance problems are subject to general linear constraints, *Journal of Economics and Business*, 42, 105-120.

Best, M.J. and R.R. Grauer, 1991a, On the sensitivity analysis of mean-variance efficient portfolios to changes in the means: Some analytical and computational results, *Review of Financial Studies*, 4, 315-342.

Best, M.J. and R.R. Grauer, 1991b, Sensitivity analysis for mean-variance portfolio problems, *Management Science*, 37, 980-989.

Best, M.J. and R.R. Grauer, 1992, The analytics of sensitivity analysis for mean-variance portfolio problems, *International Review of Financial Analysis*, 1, 17-37.

Briere, M., Drut, B, Mignon, V., Oosterlinck, K. and A. Szafarz, 2013, Is the market portfolio efficient? A new test of mean-variance efficiency when all assets are risky, *Finance*, 34, 7-41.

Brown, S. J., Goetzmann, W.N., Ibbotson, R. and S.A. Ross, 1992, Survivorship bias in performance studies, *Review of Financial Studies*, 5, 553–580.

Busse, J.A., Goyal, A. and S. Wahal, 2014, Investing in a global world, *Review of Finance*, 561-590.

Cao, J, Fu, R. and Y. Jin, 2017, International diversification through iShares and their rivals, *Journal of Risk*, 19, 25-55.

Carhart, M.M., Carpenter, J.N., Lynch, A.W. and D.K. Musto, 2002, Mutual fund survivorship, *Review of Financial Studies*, 15, 1439-1463.

Cumby, R.E. and J.D. Glen, 1990, Evaluating the performance of international mutual funds, *Journal of Finance*, 45, 497-521.

De Roon, F.A., Nijman, T.E. and B.J.M. Werker, 2001, Testing for mean-variance spanning with short sales constraints and transaction costs: The case of emerging markets, *Journal of Finance*, 56, 721-742.

De Souza, A. and A.W. Lynch, 2012, Does mutual fund performance vary over the business cycle, *Working Paper*, New York University.

Detzel, A.L., Novy-Marx, R. and M. Velikov, 2019, Model selection with transaction costs, *Working Paper*, University of Rochester.

Dimson, E., Marsh, P.R. and M. Staunton, 2016, Cycling for the good of your wealth, *Global Investment Returns Yearbook 2016*, Credit Suisse.

Errunza, V., Hogan, K. and M.W. Hung, 1999, Can the gains from international diversification be achieved without trading abroad?, *Journal of Finance*, 2075-2107.

Evans, R.B., 2010, Mutual fund incubation, *Journal of Finance*, 65, 1581-1611.

Fama, E.F. and K.R. French, 1993, Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3-56.

Fama, E.F. and K.R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics*, 116, 1-22.

Fama, E.F. and G.W. Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics*, 5, 115-146.

Ferson, W.E., 1989, Changes in expected security returns, risk and the level of interest rates, *Journal of Finance*, 44, 1191-1217.

Ferson, W.E., Henry, T. and D. Kisgen, 2006a, Evaluating government bond fund performance with stochastic discount factors, *Review of Financial Studies*, 19, 423-456.

Ferson, W.E., Kisgen D. and T. Henry, T., 2006b, Fixed income fund performance across economic states, *Research in Finance*, 23, 1-62.

Ferson, W.E. and H. Mo, 2016, Performance measurement with selectivity, market and volatility timing, *Journal of Financial Economics*, 121, 93-110.

Ferson, W.E. and M. Qian, 2004, Conditional performance evaluation revisited, *Research Foundation Monograph*, CFA Institute.

Ferson, W.E., Sarkissian S. and T. Simin, 2003, Spurious regressions in financial economics, *Journal of Finance*, 58, 1393-1414.

Ferson, W.E. and R.W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance*, 51, 425-462.

Fletcher, J., 2020, Exploring the diversification benefits of U.S. international equity closed-end funds, *Working Paper*, University of Strathclyde.

Frazzini, A., Israel, R. and T.J. Moskowitz, 2018, Trading costs, *Working Paper*, New York University.

Grauer, R.R., 2009, Extreme mean-variance solutions: Estimation error versus modelling error, *Applications of Management Science*, 13, 19-51.

Grubel, H., 1968, Internationally diversified portfolios: Welfare gains and capital flows, American Economic Review 58, 1299-1314.

Gulen, H., Xing, Y. and L. Zhang, 2011, Value versus growth: Time-varying expected stock returns, *Financial Management*, 40, 381-407.

Harvey, C.R., Liechty, J. and M. Liechty, 2008, Bayes vs resampling: A rematch, *Journal of Investment Management*, 6, 29-45.

Harvet, C.R., Liechty, J., Liechty, M. and P. Mueller, 2010, Portfolio selection with higher moments, *Quantitative Finance*, 10, 469-485.

Hodrick, R.J. and X. Zhang, 2014, International diversification revisited, *Working Paper*, University of Columbia.

Huang, D., Jiang, F., Tu, J. and G. Zhou, 2015, Investor sentiment aligned: A powerful predictor of stock returns, *Review of Financial Studies*, 28, 791-837.

Jagannathan, M., Jiao, W. and R. Wermers, 2019, International characteristic-based asset pricing, *Working Paper*, University of Maryland.

Jobson, J.D. and R.M. Korkie, 1981, Putting Markowitz theory to work, *Journal of Portfolio Management*, 7, 70-74.

Kan, R, Wang, X. and G. Zhou, 2021, Optimal portfolio choice with estimation risk: No risk-free asset case, *Working Paper*, University of Toronto.

Kan, R. and G. Zhou, 2007, Optimal portfolio choice with parameter uncertainty, *Journal of Financial and Quantitative Analysis*, 42, 621-656.

Kan, R. and G. Zhou, 2012, Tests of mean-variance spanning, *Annals of Economics and Finance*, 13, 145-193.

Kan, R, and G. Zhou, 2017, Modeling non-normality using multivariate *t*: Implications for asset pricing, *China Finance Review*, 7, 2-32.

Karehnke, P. and F. De Roon, 2020, Spanning tests for assets with option-like payoffs: The case of hedge funds, *Management Science*, 66, 5969-5989.

Ledoit, O. and M. Wolf, 2017, Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks, *Review of Financial Studies*, 30, 4349-4388.

Lessard, D., 1973, World, national and industry factors in equity returns, Journal of Finance 29, 379-391.

Li, K., Sarkar, A. and Z. Wang, 2003, Diversification benefits of emerging markets subject to portfolio constraints, *Journal of Empirical Finance*, 10, 57-80.

Linnainmaa, J., 2013, Reverse survivorship bias, *Journal of Finance*, 68, 789-813.

Liu, E.X., 2016, Portfolio diversification and international corporate bonds, *Journal of Financial and Quantitative Analysis*, 51, 959-983.

Luttmer, E., 1996, Asset pricing in economies with frictions, *Econometrica*, 64, 1439-1467.

Markowitz, H., 1952, Portfolio selection, *Journal of Finance*, 7, 77-91.

Markowitz, H.M., 2005, Market efficiency: A theoretical distinction and so what?, *Financial Analysts Journal*, 61, 17-30.

Markowitz, H. and N. Usmen, 2003, Resampled frontiers versus diffuse bayes: An expirment, *Journal of Investment Management*, 1, 9-25.

Michaud, R.O., 1989, The Markowitz optimization enigma: Is ‘optimized’ optimal, *Financial Analysts Journal*, 45, 31-42.

Michaud, R.O. and D.N. Esch, 2018, When Michaud optimization fails, *Working Paper*, New Frontier Advisors.

Michaud, R.O., Esch, D.N. and R.O. Michaud, 2020, Estimation error and the “fundamental law of active management”: Is quant fundamentally flawed?, *Journal of Investing*, forthcoming.

Michaud, R.O., and R.O. Michaud, 2008, Efficient asset management: A practical guide to stock portfolio optimization and asset allocation 2nd Edition, Oxford University Press.

Novy-Marx, R. and M. Velikov, 2016, A taxonomy of anomalies and their trading costs, *Review of Financial Studies*, 29, 104-147.

Patton, A.J. and B. Weller, 2020, What you see is not what you get: The costs of trading market anomalies, *Journal of Financial Economics*, 137, 515-549.

Perez-Quiros, G. and A. Timmermann, 2000, Firm size and cyclical variations in stock returns, *Journal of Finance*, 55, 1229-1262.

Rapach, D., Strauss, J.K. and G. Zhou, 2010, Out-of-sample equity prediction: Combination forecasts and links to the real economy, *Review of Financial Studies*, 23, 821-862.

Sharpe, W.F., 1966, Mutual fund performance, *Journal of Business*, 39, 119-138.

Solnik, B.H., 1974, Why not diversify internationally rather than domestically?, *Financial Analysts Journal*, 30, 48-54.

Tu, J. and G. Zhou, 2011, Markowitz meets Talmud: A combination of sophisticated and naïve diversification strategies, *Journal of Financial Economics*, 99, 204-215.

Wang, Z., 1998, Efficiency loss and constraints on portfolio holdings, *Journal of Financial Economics*, 48, 359-375.