The Sensitivity of Multifactor Efficiency and Performance Evaluation

Yoon K. Choi*

Revised: April 21, 2025

Abstract

In this study, we examine the sensitivity of portfolio efficiency to the choice of test assets and factor portfolios and discuss its implications for portfolio performance evaluation. We confirm that the efficiency tests are very sensitive to the choice of test assets, but relatively robust to the specification of factor portfolios. Notably, both the original Fama-French factors and their component three factors consistently identify the same test portfolios which show significant abnormal returns (i.e., non-zero alphas). Additionally, our results indicate that portfolio efficiency varies over time and declines during periods of financial crises. These findings underscore the importance of aligning performance benchmarks with both the structural characteristics of mutual funds and prevailing market conditions.

JEL classification: G10; G12

Keywords: Portfolio efficiency; performance evaluation; component factors; financial crises

*Department of Finance, College of Business Administration
University of Central Florida, Orlando, FL 32816
Email: ychoi@ucf.edu; Phone: 407-823-5023

The Sensitivity of Multifactor Efficiency and Performance Evaluation

1. Introduction

A central tenet of modern portfolio theory is that asset returns are linearly related to their risk exposures when the market portfolio is mean-variance efficient. In multifactor models, this implies a linear relationship between asset (or portfolio) returns and multiple factor loadings, as formalized by Merton's (1973) Intertemporal Capital Asset Pricing Model (ICAPM). Thus, a key requirement for using such models in performance evaluation is portfolio efficiency — i.e., whether factor portfolios span the efficient frontier. This requirement underlies many applications in mutual fund performance analysis, yet its validity remains empirically contentious.[1]

This paper examines the sensitivity of portfolio efficiency to the choice of test assets and factor portfolio construction, and consequently, the validity of using standard factor models — such as the Fama-French (FF) three-factor model — as performance benchmarks.[2] These questions are essential for mutual fund evaluation, where estimated intercepts from factor regressions (alphas) are widely interpreted as measures of abnormal performance. However, if factor models are

[1] Fama (1996) also shows the relation between multifactor portfolio efficiency and the ICAPM.

[2] The three factors are the market risk premium (MKT), size premium (SMB: small portfolio returns minus big portfolio returns), and value premium (HML: high B/M portfolio returns minus low B/M portfolio returns). Fama and French (1993) argue that HML and SMB seem to explain the test assets returns very well because those two factors probably capture the important state variables, thus playing as good hedging portfolios. In fact, Liew and Vassalou (2000) contend that Fama-French two factors such as HML and SMB are good proxies for future economic growth productivity. This is consistent with the assumption that HML and SMB may serve as good proxies for underlying state variables. We also examine five-factor model and find almost the same results. We discuss more details in a subsequent section.

inefficient or mis-specified, positive (or negative) alphas may simply reflect pricing biases rather than true managerial skill.

Our empirical framework builds on the multivariate F-test developed by Gibbons, Ross, and Shanken (GRS, 1989). A high GRS statistic implies that the null hypothesis of portfolio efficiency is rejected, indicating that the benchmark model fails to span the efficient frontier.[3] Using this methodology, we test the efficiency of the FF three-factor model and compare it against several alternative specifications based on its underlying component portfolios — such as small (S) and big (B) size-sorted portfolios and high (H) and low (L) book-to-market portfolios. Based on the tests with these component factor portfolios, we can re-examine the validity of SMB and HML as true risk factors (The literature section will discuss the potential pricing biases arising from the misspecification of SMB and HML in detail).[4]

We employ three sets of test assets: (1) the FF 25 size–value sorted portfolios, (2) cash-to-price (C/P) decile portfolios, and (3) FF 30 industry portfolios. This variety allows us to assess whether portfolio efficiency test results are driven by similarities between test assets and the factors, especially when both are constructed using firm characteristics. Additionally, we segment our data (1986–2022) into 5-year rolling windows to capture potential time variation in portfolio efficiency, particularly around major crises such as the Dot-com bubble, Global Financial Crisis, and the COVID-19 pandemic.

---

[3] The GRS statistic can be interpreted as the difference between the Sharpe ratio of a benchmark factor model with that of an expanded portfolio including the test assets (see also Jobson and Korkie, 1989; Huberman and Kandel, 1987).

[4] Refer to Ferson and Harvey (1999), Berk (2000), and Pukthuanthong et al. (2019) for a discussion of the potential issues in testing asset pricing when characteristics-sorted (size or value) test portfolios are used with the risk factors constructed using the same set of test portfolios.

In sum, this paper extends the foundational work of Fama and French (1993, 1996) by addressing several underexplored issues in portfolio efficiency testing and mutual fund performance evaluation. First, we show that the outcomes of GRS efficiency tests are highly sensitive to the choice of test assets, particularly when the portfolios are not constructed using the same firm characteristics that define the factors themselves. Cremers et al. (2012) show that the FF three-factor model yields significantly positive alphas for passive and big portfolios like the S&P 500, largely due to construction issues in the SMB factor. This misalignment can lead to spurious inferences about performance. Second, we build on this argument by examining whether component-based can reduce such distortions. Specifically, our results suggest that component factors, used in place of traditional constructs like SMB and HML, offer comparable explanatory power while potentially mitigating measurement biases. Notably, these alternative factors also help identify specific test portfolios with significant abnormal returns, which in turn contribute to overall portfolio inefficiency. For example, we find significant alphas in the same test portfolios when we use component factors such as M, S, and H in place of SMB and HML as factors. This observation suggests that the source of pricing bias may not be solely due to the construction of SMB, as argued by Cremers et al. (2012), but could stem from other misspecifications.

Third, we document substantial time variation in portfolio efficiency, showing that standard factor models tend to break down during major market disruptions – including the 1998 Asian crisis and Dot-com crash, the 2008 financial crisis, and the COVID-19 pandemic. Specifically, GRS efficiency statistics spike during these episodes, leading to strong rejection of portfolio efficiency. This pattern is consistent with the notion that HML and SMB (or their component factors) may proxy for underlying state variables. For example, Liew and Vassalou

(2000) find that HML and SMB are predictive of future economic growth, while Harvey and Liu (2021) provide further support for the FF three-factor model using individual stocks as test assets.

Moreover, the literature shows that financial crises are characterized by heightened contagion and co-movements across sectors and firms, thus reducing the benefits of portfolio diversification. These dynamics can distort investor expectations and confound perceptions of risk and return, leading to portfolio inefficiency during such periods. Finally, we demonstrate that model misspecification can generate spurious alpha estimates, reenforcing the role of aligning benchmarks with both portfolio characteristics and prevailing market conditions. Collectively, these findings advance our understanding of when and how factor models serve as reliable tools for evaluating mutual fund performance—an issue that remains only partially addressed in the original Fama-French framework.

Section 2 introduces the GRS tests of portfolio efficiency using alternative test assets. We also examine the relation between the GRS statistic which is a function of abnormal returns (regression alphas) and portfolio efficiency. It also presents a literature review of portfolio efficiency and performance measures. Section 3 outlines the data and methodology used in our analyses, while Section 4 presents and discusses the empirical results. We discuss the implications of our findings for mutual fund performance evaluation in Section 5. Section 6 summarizes and concludes the paper.

2. Literature Review and GRS tests of Portfolio Efficiency

We follow the methodology of Gibbons, Ross, and Shanken (GRS, 1989) who develop a test of mean-variance efficiency for a given portfolio, using a multivariate F-test based on small-sample distributions. This approach offers several advantages. First, the GRS F-test is robust to departures from the normality assumption in asset returns (MacKinlay, 1985). Second, as shown in Gibbons and Shanken (1987), GRS statistics retain strong test power when efficiency is evaluated across subperiods, making it more appropriate than simply aggregating data over the entire sample. Third, the F-test statistic can be interpreted in terms of Sharpe ratios – specifically, the difference between the Sharpe ratios of the test assets and the combined portfolio of test and factor assets - which provides additional insight into the sources of portfolio inefficiency.

The GRS F-test is based on the following regression:

$$R_{it} - R_{ft} = a_i + b_i(R_{mt} - R_{ft}) + s_i SMB_{it} + h_i HML_{it} + e_{it.} \quad i = 1, \ldots\ldots, N \qquad (1)$$

This regression model states that, on average, excess returns on test assets can be explained by the risk premiums such as (*Rm –Rf), SMB,* and *HML,* and their sensitivities (loadings) to the three factors. If this three-factor model is correct, then we should expect the regression intercept estimate to be close to zero for all test assets. The GRS test evaluates whether the intercept estimates from time-series regressions of portfolio returns on factor models are jointly equal to zero. If the intercepts are statistically close to zero, it suggests that the factor model efficiently explains the returns of the test assets. As GRS (1989) demonstrate, a small value of the GRS statistic indicates that we cannot reject the null hypothesis of mean-variance efficiency for the given factor model – in our case, the FF three-factor model. Specifically, the GRS statistic is a function of the intercept estimates and the residual covariance matrix obtained from regressing each of the N test assets against the factor portfolios. In summary, the GRS F-test statistics is:

$$\frac{T}{N}\frac{T-N-1}{T-L-1}\left(\frac{\hat{a}'\hat{\Sigma}^{-1}\hat{a}}{1+\bar{\mu}'\hat{\Omega}^{-1}\bar{\mu}}\right) \sim F(N, T-N-L) \quad , \ldots\ldots\ldots\ldots\ldots (2)$$

where $\hat{a}$ is a vector of the intercept estimate from regression equation (1); $\bar{\mu}$ is the sample mean of test asset returns; $\hat{\Sigma}$ is the sample estimate of variance-covariance matrix of the residuals, and $\hat{\Omega}$ is the sample variance/covariance matrix of N-vector of test asset returns, in vector format. T, N, and L are the number of observations, the number of test assets, and the number of factor portfolios, respectively. This framework is widely adopted in asset pricing and performance evaluation due to its intuitive connection to the Sharpe ratio and its interpretability in mean-variance terms.

GRS efficiency statistics (equation 2) can be expressed in the context of Sharpe ratios as below:

$$\frac{T}{N}\frac{\mathrm{T-N-1}}{T-L-1}\,W, where\ \ W = \left(\frac{\sqrt{1+\hat{\theta}_{N+L}^2}}{\sqrt{1+\hat{\theta}_L^2}}\right)^2 - 1, \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$\hat{\theta}_L^2$ is the maximum sharpe ratio on the efficient frontier constructed with L factor portfolios, while $\hat{\theta}_{N+L}^2$ is the maximum sharpe ratio of the efficient frontier with N test assets and L factor portfolios. The statistic above, W, can also be expressed as the portfolio efficiency ratio, which is the Sharpe ratio of all assets including tests assets and factor portfolios divided by the Sharpe ratio of the tangent portfolio on the efficient frontier generated by factor portfolios. Thus, W and F will be zero, accepting the null hypothesis when the two Sharpe ratios are the same. In this case, the factor portfolios are said to span all assets, including test assets as well as the factor portfolios.

A popular way to measure the abnormal performance of mutual funds is to estimate their intercepts of the regression based on a factor model like equation (1). A problem with this approach for performance evaluation is that significant positive (or negative) intercepts can be obtained simply due to pricing bias, without any true abnormal performance from mutual funds. A notable concern in performance evaluation arises when non-managed portfolios, such as the S&P 500 index, yield statistically significant alphas under standard factor models. Cremers et al. (2012) demonstrate that the S&P500 index exhibits statistically significant abnormal performance when evaluated using the Fama-French three- or four-factor model (Fama and French, 1993; 1996). Cremers et al. (2012) argue that such findings undermine the validity of these factor models as appropriate benchmarks for portfolio evaluation. They attribute the pricing bias for the large test portfolio to the misspecification of SMB – a long-short portfolio.

To address this, we investigate alternative factor constructions. Fama and French (1996, 2015) note that their factors (SMB and HML) are essentially linear combinations of underlying component portfolios such as H, L, S, and B — each of which may also serve as multifactor minimum variance (MMV) portfolios – a point we elaborate on later. We replicate and extend their framework by comparing the performance of alternate component-based factor models with the traditional FF three-factor model across various test asset groups. We acknowledge the limitations of the GRS test, which relies on constant parameters and uses unconditional moments – assumptions that may not hold in the presence of time-varying risk premia and covariances. Prior studies such as Ferson, Kandel and Stambaugh (1987) and Ferson and Harvey (1999) have proposed conditional approaches to address these issues. While we do not pursue a fully conditional model, we adopt a rolling 5-year estimation window to partially account for changing market dynamics.

2.1. Sensitivity to the Choice of Test Assets

Fama and French (1993, 1996, 1997) evaluate portfolio efficiency using a variety of test portfolios, including the 25 size-value sorted portfolios, 48 industry portfolios and decile portfolios sorted on C/P, E/P, or B/M.[5] Based on the 1963 - 1991 sample, FF (1993) find that, out of the 25

[5] B/M, E/P, and C/P represent book-to-market, earnings/price, and cash flow/price, respectively.

portfolios, only three exhibit statistically significant alphas.[6] For instance, portfolio P (1,1) has an alpha of -0.34 (t = -3.16), while P(5,1) - the largest test portfolio in the sample - has a significantly positive alpha of 0.21 (t=3.27) as shown in Model (iv) of Table 9a. This result aligns with Cremers, et al. (2012), who argue that portfolio such as the SP&500 can exhibit positive alpha due to mis-specified factor construction.

In contrast, when using decile portfolios sorted by E/P or C/P, FF (1996) show that none of the intercepts are statistically significant, and the GRS test fails to reject the null hypothesis, implying that the three-factor model adequately explains these portfolio returns. Further, FF (1997) examine the efficiency of their three-factor model, using 48 industry test portfolios and find that seven portfolios exhibit significant intercepts, indicating that portfolio efficiency depends on the choice of test portfolios.

To investigate whether this sensitivity persists in more recent data, we replicate and extend these analyses using updated samples. Our study examines whether the rejection of portfolio efficiency under the FF three-factor model continues to depend on the choice of test assets and factor construction. This line of inquiry contributes to ongoing debates over whether the apparent success of FF factor models is partially driven by the alignment between factor construction and test asset formation – both of which are often based on similar firm characteristics (e.g., size, value).

2.2. Multifactor-Minimum-Variance (MMV) portfolios

[6] They report similar results on the alpha estimates in FF (1996), except that four test portfolios produce significant abnormal intercepts including the same portfolios for slightly different sample period of 1963 through 1993. For example, P (1,1) has an alpha estimate of -0.45 (t = -4.19), while P (5,1) has an intercept estimate of 0.20 (t=3.14).

Fama and French (1996) argue that their three factors (MKT, SMB, and HML) are close to multifactor minimum variance (MMV) portfolios and that their component portfolios – M, H, L, S, B – can serve as viable substitutes. Building on this insight, we propose using these MMV component portfolios to evaluate the FF 25 portfolios. This approach may help mitigate potential pricing biases arising from misspecifications in the construction of SMB and HML, thereby offering a more robust assessment of portfolio efficiency.

Concerns about the validity and interpretations of FF factors have been widely discussed in the literature. For example, Berk (1995) notes that ratios such as the book-to-market – central to HML – may spuriously explain cross-sectional returns due to their mechanical relation with price. Daniel and Titman (1997) argue that the success of the FF model may stem from firm characteristics embedded in both the factors and test assets, rather than from priced risk. Similarly, Lewellen, Nagel, and Shanken (2010) and Wallmeier and Tauscher (2012) highlight that using test portfolios sorted on size and B/M, while constructing factors from the same criteria, may create artificial model validation due to overlapping construction mechanism. Similarly, Ferson, Sarkissian, and Simin (2003) challenge the validity of SMB and HML as true risk factors. Ferson and Harvey (1999) further show that macroeconomic variables reflecting time-varying expected returns can explain the cross-section of stock returns well. In contrast, Liew and Vassalou (2000) present evidence that both HML and SMB are associated with future GDP growth, suggesting their relevance as proxies for systematic risk. More recently, Fama and French (2015; 2018) expand their framework to a five-factor model by including profitability and investment factors, which improve the model’s explanatory power relative to the original three-factor specification.

2.3. Time-Varying Portfolio Efficiency and Market Crises

Portfolio efficiency is not static. We find that GRS statistics spike during crisis periods — such as 2000, 2008, and 2020 — indicating breakdowns in factor model efficiency. These periods are characterized by heightened market volatility driven by macroeconomic uncertainty, policy interventions, and shifting investor sentiment. During financial crises, volatility tends to spike, and correlations among asset classes often increase as investors simultaneously shift their exposure to riskier assets. This dynamic has been documented in the literature (e.g., Liew and Vassalou, 2000; Longin and Solnik, 2001). These dynamics can cause dramatic shifts in the shape and position of the efficient frontier, rendering previously efficient portfolios inefficient. Muir (2017) further highlights that risk premia—such as expected returns or dividend yields relative to price—typically rise during crisis periods, altering the underlying risk-return tradeoff.

Further, traditional diversification strategies become less effective as asset correlations rise and return distributions become more skewed (Aloui, Aïssa, and Nguyen, 2011; Beine, Cosma, and Vermeulen, 2010). In response, alternative investment strategies have been proposed to enhance portfolio efficiency during turbulent periods. For instance, incorporating additional individual stocks or alternative assets—such as commodities or non-U.S. equities—into traditional equity portfolios or Fama-French factor portfolios can help mitigate contagion effects and restore diversification benefits (Akhtaruzzaman, Boubaker, and Sensoy, 2021; Ao, Yingying, and Zheng, 2019; Tronzano, 2022; Wen, Wei, and Huang, 2012).

To investigate whether standard factor models remain robust across varying market regimes, we include these crisis periods in our analysis. This allows us to explore how portfolio efficiency evolves under extreme market stress and whether the Fama-French three-factor model retains its explanatory power. We recognize the limitations of the F-test of portfolio efficiency as used in this paper, particularly its reliance on the assumption of stable parameters and constant factor loadings over time. As Lettau and Ludvigson (2001) emphasize, risk-premia can vary significantly with changing economic conditions, making traditional efficiency tests potentially unreliable. To mitigate this issue, we adopt a rolling five-year window approach to estimate portfolio efficiency. This method allows the model to dynamically adjust to evolving market conditions, thereby reducing the influence of structural breaks and improving the robustness of our results.[7]

The rolling window technique is widely supported in the literature as a means to address parameter instability and time variation in expected returns and covariances. Pesaran and Timmermann (1995) examine forecast performance under changing environments; Ang and Bekaert (2002) explore regime shifts in international equity markets; and Ferson and Harvey (1999) model conditional asset pricing using time-varying instruments. By applying this approach, we aim to provide more reliable and time-sensitive assessments of the efficiency of the Fama-French factor models, particularly during periods of systemic financial stress.[8]

3. Data and Methodology

We obtain the Fama-French three-factor portfolios from Kenneth French's online data library to evaluate the portfolio efficiency of the three-factor model over the period from 1986 to

[7] Further, Harvey and Liu (2021) argue that the use of *portfolios,* rather than *individual* stocks, helps alleviate concerns about the assumption of constant factor loadings, as portfolios tend to exhibit more stable exposures over time.

[8] Pesaran and Timmermann (1995) discuss the benefits of using rolling window analysis in predicting stock returns, emphasizing its ability to adapt to changing market conditions and improve model robustness. Ferson and Harvey (1999) highlight the role of conditioning variables in asset pricing models, advocating for methods that account for time-varying risk factors and suggesting rolling window analysis as a practical approach. Ang and Bekaert (2002) explore the impact of regime shifts on international asset allocation and demonstrate the effectiveness of rolling window techniques in capturing time-varying correlations and factor loadings.

2022. As test assets, we employ three distinct sets: (1) the 25 Fama-French size and value-sorted portfolios, (2) 10 FF portfolios sorted by cash flow-to-price (C/P) ratios, and (3) 30 FF industry portfolios. The GRS test, which we use to assess portfolio efficiency, is highly sensitive to the number of time-series observations (T). By construction, the GRS statistics tend to increase with T, potentially overstating inefficiency in longer samples. Gibbons, Ross, and Shanken (1989) themselves caution against this issue, recommending a sample of approximately 60 monthly observations for reliable inference for 20 to 30 test assets.

Accordingly, we compare the results of the GRS test over the full 444-month sample—mirroring the approach in Fama and French (1996)—with those obtained from 5-year subperiods (i.e., 60-month windows). This allows us to evaluate whether the efficiency of the three-factor model remains stable across different time horizons and economic conditions. Further, this approach, as discussed above, allows the factor premia (and portfolio efficiency) to be time-varying. In addition to the standard FF factors, we further explore the portfolio efficiency of the three-factor model by replacing SMB and HML with their component portfolios—specifically, M (market), S (small), B (big), H (high book-to-market), and L (low book-to-market). One advantage of using these component portfolios is that this approach imposes fewer constraints on the factor loadings. Specifically, it allows the model to estimate separate sensitivities to each component factor portfolio, providing greater flexibility in capturing the return dynamics of the test assets. Consider the following three-factor regression:

$$R_{it} - R_{ft} = a_i + b_i(R_{mt} - R_{ft}) + ca_i(Coma_{it} - R_{ft}) + cb_i(Comb_{it} - R_{ft}) + e_{it.} \quad i = 1, \ldots, N, \quad (2)$$

where *Coma and Comb* represent any two component factors from the set of three components – H, S, and L. The big portfolio factor (B) is omitted from the regression equation due to its high correlation with the market factor. In Equation (2), both the high book-to-market (H) and low book-to-market (L) portfolios have independent factor loadings, h_i and l_i , respectively. In addition to Cremers et al. (2012)'s concerns regarding the SMB and HML construction,[9] the traditional Fama-French three-factor regression (Equation 1) implicitly assumes that the test assets' sensitivities to H and L are equal in magnitude but opposite in sign, with the HML loading represented as (i.e., hHML = hH – hL). This constraint may oversimplify the true structure of returns. By estimating h_i and l_i, separately, Equation (2) allows for more accurate and flexible modeling of test asset exposures, which can lead to better explanatory power.

To further assess the robustness of these models, we extend our sample period to include the most recent market environments, covering the full span from 1980 to 2022. This expanded horizon captures several major financial crises and structural shifts. We estimate portfolio efficiency using the GRS test within rolling 5-year windows throughout the sample period, thereby accounting for time variation in market conditions and model performance.

4. Empirical Results

4.1. Choice of Test Assets and Portfolio Efficiency

Table 1 shows the regression estimates for each of the 25 FF size and value-sorted portfolios over the period 1986 to 2022, using the Market risk premium, SMB and HML explanatory variables. Where appropriate, we compare our results with those of Fama and French (1996) to highlight differences in factor loadings and interpret the results in light of more recent data. Overall, the three-factor model explains average returns well for most portfolios, with the

[9] According to Cremers et al. (2012), a large-cap portfolio such as the S&P500 tends to be highly correlated with the BIG portfolio, which leads to a significantly negative loading on SMB. Coupled with a positive average SMB premium, this results in an artificially high alpha for the S&P500 – a misleading indicator of performance.

exception of six portfolios - P(1,1), P(2,1), P(1,4), P(1,5), P(5,1) and P(5,4) - which exhibit significant intercepts of -0.62, -0.17, 0.23, 0.25, 0.17, and -0.27, respectively.[10] Interestingly, the estimated market betas are close to 1.0 across all 25 portfolios, regardless of firm size or book-to-market characteristics. This result reflects the ability of the SMB and HML factors to absorb variation in market exposure attributable to size and value effects. The SMB factor loadings show systematically different patterns across portfolio size: at each book-to-market level, the loading decreases as portfolio size increases. Specifically, the smallest portfolios exhibit SMB beta greater than one, ranging from 1.38 to 1.04, while the biggest portfolios show even negative factor loadings, ranging from -0.26 to -0.19. This pattern is quite consistent with the original findings of FF (1993), based on their 1963 – 1991 sample. Given the strong factor loadings of SMB and HML, with very large t-values, it is not surprising to observe large R^2 values.

We extend our analysis by examining alternative sets of test portfolios, including the 30 industry portfolios and 10 C/P decile portfolios, as shown in Table 2. It is noteworthy that six out of the 30 industry portfolios show significant intercept estimates, implying that the FF three-factor model explains the average returns of most industry portfolio fairly well. For the 10 C/P decile portfolios, the model performs even more robustly: eight portfolios have insignificant intercepts, with only CP2 and CP3 showing t-values greater than 2.0. Compared to the 25 FF size-value portfolios, these two test portfolios offer an opportunity to assess the FF three factors under different test portfolios. Notably, the industry and CP-sorted portfolios are not formed using the same firm characteristics that define the FF factors, thus reducing the likelihood of spurious

[10] We follow the matrix convention in naming each portfolio in the matrix format in the table. For example, P(1,1) stands for the portfolio in the first row and first column. That is, P (1,1) stands for the size-value sorted portfolio being the smallest in size and lowest in B/M ratio. The FF five factor model yields similar results (not reported here).

correlation between the test assets and the factors. This independence allows for a more credible assessment of whether the FF factors truly capture systematic risk premia, rather than merely reflecting mechanical overlaps in portfolio construction.

There appears to be a monotonic relation between cash flows relative to market price and value/growth (HML) factor loadings as shown in Panel A of Table 2. Specifically, portfolio returns sorted by on the cash flows per market price increase as value/growth factor loadings increase. However, we do not find strong evidence of covariance between industry portfolio returns and factor returns. This result seems reasonable given that industry portfolios aggregate heterogeneous firms - varying in size, valuation, and performance – thereby diluting their sensitivity to size and value factors. For example, within a given industry, a small, weak-performing firm may exhibit positive sensitivity to SMB and HML, while a large, strong-performing firm in the same industry may display negative sensitivity to these factors. When these diverse exposures are aggregated into an industry-level portfolio, the net loading on SMB and HML may be substantially attenuated. As a result, the explanatory power of the three-factor model is weaker for industry portfolios than for portfolios sorted directly on firm characteristics. This is reflected in the average R^2 reported in Panel B of Table. For the 30 industry portfolios, R^2 value range widely – from as low as 0.22 (e.g., Tobacco products or Coal products) to as high as 0.86 (Finance industry). However, the CP decile and 25 FF size/value test assets exhibit uniformly higher explanatory power, with R^2 values ranging from 0.80 to 0.96.

We notice very high HML factor loadings for certain industries. For instance, the Finance industry shows an HML loading of 0.60 with a t-value of 18.51, implying its high sensitivity to financial distress, which is proxied by the value factor. Similarly, the Oil and Textiles industries show high positive HML loadings - 0.72 (t = 9.80) and 0.86 (t = 10.18), respectively. Meanwhile, two industries - Services and Business Equipment - show significantly negative HML factor loadings (-0.55 and -0.52, with t-values of -14.60 and -9.29, respectively). These negative exposures imply that firms in these industries are more closely associated with strong growth prospects.

Table 3 reports the results of the GRS test for portfolio efficiency under on the FF three-factor model, using the 25 size/value portfolios over the period from 1986 to 2022. We reject the portfolio efficiency, with a GRS statistic of 4.327 (p value = 0.000) and consistently large absolute intercepts (in Panel A). The average intercept estimate is 0.117, further underscoring the model's inability to fully explain the cross-section of returns during this period. Panel B of Table 3 presents GRS test results for seven non-overlapping sub-periods. As expected, portfolio efficiency changes over time. While the multifactor efficiency is rejected in the first two sub-periods (p-values of 0.002 and 0.006), it is not rejected for the remaining five periods. Notably, the most recent two sub-periods (2011-2014 and 2016-2020) yield the smallest GRS statistics of 1.110 and 0.962, respectively with corresponding p values of 0.386 and 0.534. These periods also coincide with the smallest average absolute intercept estimates, suggesting that the FF three-factor model performs relatively better in explaining returns during more recent years.

In addition to the 25 FF portfolios as test assets, we use ten decile portfolios sorted by cash flow/price (C/P) as test assets. Interestingly, the results differ markedly from those obtained using the FF 25 portfolios. As shown in Panel A of Table 4, the GRS statistic is very small (1.333 with p = 0.210) over the full sample period (1986 – 2022, or 444 monthly observations), indicating that we cannot reject the efficiency of the FF three-factor model. The average absolute value of the regression intercepts is also small, at just 0.075. These findings are consistent with those of Fama and French (1996), who failed to reject portfolio efficiency for earlier sample periods. Overall, the FF three-factor model appears to explain the average returns of the C/P decile portfolios quite well.

Panel B of Table 4 presents the result for seven sub-periods. In each case, the GRS statistic remains insignificant, with the smallest value (0.544) observed in the 2001 – 2005 period. The only exception is the 2011 to 2015 sub-period, during which the GRS statistic rises to 2.30 with a p-value of 0.027, indicating a rejection of portfolio efficiency at conventional significant levels. Table 5 shows another portfolio efficiency result using the 30 FF industry portfolios as test assets for seven sub-periods. In this case, the FF three-factor portfolio efficiency is rejected in two sub-periods (1996 to 2000 and 2016 to 2020) with corresponding p-values of 0.001 and 0.004, respectively. Overall, these findings highlight the sensitivity of efficiency test outcomes to the underlying test asset selection and test periods. A more thorough investigation is warranted to resolve these discrepancies, but it is beyond the scope of this paper.

4.2. Multifactor Portfolio Efficiency with alternative factors

Fama and French (1996) demonstrate that their three-factor portfolios can be derived as a linear combinations of component portfolios – specifically, the market portfolio (M), high BM portfolio (H), low BM portfolio (L), and small-cap portfolio (M). They argue that each of these component portfolios can be viewed as a multifactor minimum variance (MMV) portfolio. Their mean-variance spanning test shows that excess returns on any three of M, S, H, and L almost completely explain the excess returns on the remaining component portfolios. We confirm their results using updated data from 1986 to 2022. As shown in Table 6, any one of the L, H, S, and M factors is nearly fully explained by the other three, with insignificant intercepts and very high R-squared values. This suggests that the original FF three-factor model is not unique in capturing systematic risks; rather, any three MMV component factors can substitute for the original factors. Table 7 shows the results of using alternative three-factor combinations – such as MHL, MSL, and MSH – to evaluate portfolio efficiency using the 25 FF size-value portfolios as test assets. Consistent with FF (1996), we find strong rejection of portfolio efficiency across all combinations. The GRS test statistics are consistently around 4.4 for the whole sample period from 1986 to 2022, nearly identical to the statistic obtained when using the original FF factors (see Table 7). This indicates that the FF three factors may not be unique in mimicking underlying state variables. Any triplet of M, S, H, and L as factors performs similarly, consistent with the results shown in Fama and French (2016).

Even the four-factor model (with S, B, H, and L) produces a GRS statistic comparable to the three-factor model, with the smallest average intercept (0.109) and the highest R^2 (0.927) across the 25 FF test portfolios. However, its performance deteriorates when applied to the 10 CP test asset returns. As shown in Panel B of Table 7, the four-factor model yields a relatively high GRS statistic with a marginally significant p-value of 0.063, indicating a portfolio efficiency rejection at the 10% level. In this case, we include B and exclude M, using S, H, L, and B as the four component MMV factors. This framework allows us to examine the implications of the structural restrictions embedded in the SMB and HML factors in the traditional Fama-French three-factor model. Specifically, in the FF model, the loading h_i on HML—defined as H minus L—implicitly imposes equal and opposite loadings on the high and low book-to-market portfolios.

Table 8 shows the estimated factor loadings on these component MMV factors such as M, S, H, and L (and B in a later section) for the CP decile test portfolios. As shown in Table 7, these component factors and the FF three factors produce similar GRS statistics and R^2 in most cases. However, we observe quite distinct patterns in the behavior of market and book-to-market (BM) risk exposures across the alternative factor specifications. In Panel A of Table 8, we find a clear monotonic decline in market beta as the cash flow-to-price ratio increases, while the loading on the H (high B/M) factor rises. This suggests that firms with higher cash flow-to-price ratios—typically associated with weaker performance—exhibit lower market risk and higher exposure to value-related risk. In contrast, Panel B, which uses M, S, and L as factors, reveals that high-C/P portfolios are associated with increasingly negative loadings on the L (low B/M) factor.

A particularly noteworthy finding is the contrasting behavior of market betas across these specifications. With the M, S, H factor set (Panel A), there is a negative relationship between the C/P ratio and market beta, while with the M, S, L set (Panel B), the relationship turns positive. This stands in contrast to the results from the standard three-factor model (M, SMB, HML), shown in Table 2, where market betas remain relatively stable — hovering around one — across the full spectrum of C/P portfolios.

These findings highlight the additional interpretive flexibility afforded by decomposing the SMB and HML factors into their underlying MMV components. Although the models perform similarly in terms of overall fit (as indicated by GRS and R^2 values), the factor loadings reveal economically meaningful variations in how different types of risk are priced across portfolios. In particular, the results in Panel B are consistent with the view that portfolios with high cash flow-to-price ratios—often associated with value traps or distressed firms—should exhibit elevated market risk, as reflected in their higher betas under the M, S, L specification.

Panels C and D of Table 8 report the estimated factor loadings on H and L in alternative component-factor models, which do not impose the stringent constraints in HML. In Panel C, where the market factor is included alongside H and L, we observe that the factor loadings on H and L generally have opposite signs, except for CP2 and CP3, which is partly consistent with the implied restraint on the HML factor. In contrast, as shown in Panel D, where the market factor is substituted by S and B instead, the factor loadings on H and L are quite different from each other, often exhibiting the same sign across most CP decile portfolios. This behavior contrasts with the typical expectation under the HML construction, where H and L are assumed to have equal and opposite effects. Similarly, the factor loadings on S and B also tend to share the same sign, indicating that these two size-related factors may be capturing overlapping risk dimensions in this context.

Furthermore, we observe that the H factor loading is strongly associated with average CP decile returns: as the loading on H increases, so do portfolio returns. This pattern aligns with the earlier findings on HML in Table 2 and reinforces the interpretation that both H and HML effectively capture distress-related risks. Overall, the results in Table 8 suggest that while both the original Fama-French three-factor model and alternative models based on component MMV factors explain CP decile portfolio returns well, the underlying mechanisms differ depending on the factor specification. These differences highlight that multiple paths to achieving portfolio efficiency exist, and they are sensitive to the choice of component portfolios used to construct the factor model.[11] We discuss the implications of these findings with component factor models for mutual fund performance evaluation in Section 5.

4.2. The Behavior of Portfolio efficiency around financial crises

[11] We also estimated three-factor models using the component factor portfolios for the 25 FF size/value sorted portfolios. We obtained similar results that the market betas are close to one for all 25 portfolios, while the market betas estimated with the component factors appear more random. They range from -1.95 to 1.44 for the MHL factors, while they range from -0.50 to 1.46 for MSH factors.

Finally, we present portfolio efficiency statistics over the extended period from 1980 to 2022 to examine the behavior of portfolio efficiency around major financial crises relative to more stable periods. Earlier, as shown in Tables 3 through 5, GRS F-statistics varied over time across the five-year sub-periods from 1986 to 2022 and are highly sensitive to the choice of test portfolios. In this section, we extend the analysis to rolling five-year periods beginning in 1980, with a particular focus on how GRS test results behave around major economic crises. As we discussed before, the efficient frontier can shift substantially during financial crises, reflecting changes in the return-risk relationship. Muir (2017) finds that risk premia such as dividend yields and credit spreads tend to rise significantly during financial downturns, driven by increased uncertainty and compensation for bearing risk.

Figure 1 illustrates conspicuously high GRS statistics around the years 1991, 2000, 2008, and 2020 when using the 30 FF industry portfolios as test assets in conjunction with the FF three-factor model.[12] These elevated GRS values suggest that the three FF factors struggle to explain average returns during financial crises, indicating portfolio inefficiency. Such inefficiency may be attributed to investors' shifting expectations and heightened uncertainty about future economic conditions during turbulent times. Table 9 provides further details. Panel A of Table 9 reports GRS statistics exceeding 2.5 in the years 1998, 1999, 2000, and 2020 – each aligned with periods of financial distress. These results suggest that the GRS test is effective in flagging episodes when the FF factor model fails to adequately explain returns.

Panel B of Table 9 presents analogous results using the 25 FF size-value portfolios as test assets. Interestingly, the GRS statistics derived from industry portfolios appear to be more

---

[12] We also can match three major financial crises - the Asian Crisis (1998), the Dot-Com Bubble (2000), the Global Finance crisis (2008), and the COVID crisis (2020) - with highest GRS statistics when the 30 industry portfolios as test assets are employed with a single market factor.

responsive in identifying financial crises than those from the size/value portfolios. One possible explanation is that industry portfolios better capture sector-level economic dynamics, which are directly influenced by macroeconomic shocks and investor behavior. This finding is consistent with Liew and Vassalou (2000), who show that the SMB and HML factors possess predictive power for future economic growth. In sum, our results underscore the importance of selecting appropriate test assets and accounting for time variation when evaluating the portfolio efficiency of factor models, particularly during periods of economic stress.

5. Implications for mutual fund performance evaluation

We discuss the implications of our empirical findings for mutual fund performance evaluation through the lens of *portfolio efficiency,* which can be interpreted by the concept of *potential performance* introduced by Jobson and Korkie (1989). Potential performance refers to the best achievable return-to-risk tradeoff within a given set of assets or portfolios. A mutual fund manager demonstrates abnormal performance only if they surpass this potential. Our results highlight the sensitivity of portfolio efficiency to the choice of factor models, underscoring the importance of selecting an appropriate benchmark. For instance, as shown in Table 1, the FF three-factor model effectively explains the average returns of the 25 FF size-value portfolios. Specifically, 76% (19 out of 25) of the test portfolios yield insignificant alpha estimates, supporting the model's use as a benchmark for funds with exposures driven by firm characteristics such as size and book-to-market ratios.

However, six portfolios - P(1,1), P(2,1), P(5,1), P(1,4), P(1,5), and P(5,4) - exhibit statistically significant intercepts (refer to Table 10). These alphas are likely due to pricing biases rather than managerial skill. Accordingly, performance evaluations based on these portfolios should be adjusted to avoid misattributing spurious abnormal returns. In particular, P (5,1) - the largest size and lowest BM portfolio - displays a strong positive alpha of 0.17% per month (Panel A), aligning with Cremers et al. (2012) who find similar positive performance in S&P500. Conversely, P(1,1) - the smallest portfolio among 25 FF portfolios – shows a large negative alpha of -0.62% per month, reinforcing concerns about pricing biases at the extremes of the size-value spectrum. We also observe persistent abnormal returns in CP2 and CP3 portfolios, as shown in Table 8, regardless of whether standard FF factors or their underlying components are employed. These findings call for caution when evaluating mutual funds with exposures similar to these six FF portfolios and the CP2 and CP3 portfolios.

Interestingly, the component MSH factor model (Panel B) identifies the exact same six portfolios for their significant alphas as the original FF three-factor model. MSL and MHL factors also identify the exact five portfolios for their significant alphas. For these two component factor models, we observe the significant alphas for P (4,1) instead of P (2,1). This observation suggests that the source of pricing bias may not be solely due to the construction of SMB or HML, as argued by Cremers et al. (2012), but could stem from other misspecifications. Furthermore, we show in Tables 3, 4, 5, and 9 that portfolio efficiency is time-varying. Given that intercept estimates - and thus the evidence of abnormal performance - can vary significantly over time, it is essential to evaluate portfolio efficiency prior to adopting any factor model as a benchmark. A time-sensitive approach ensures that the selected factor model offers a valid and unbiased standard for assessing mutual fund performance during the specific evaluation period.

6. Conclusion and Future Research

We explore the multifactor efficiency of the FF three-factor model, employing various test assets and alternative factor constructions. Our analysis highlights the complexity of evaluating portfolio efficiency, driven by two primary sources of portfolio inefficiency in factor models. First, we find that the outcomes of efficiency tests are highly sensitive to the choice of test assets. We employ three distinct sets: the 25 FF size/value-sorted portfolios, the cash flow-to-price (C/P) decile portfolios, and the 30 FF industry portfolios. Each set reflects different economic constructs - firm characteristics, fundamental valuation, and sectoral composition, respectively. This variation highlights the importance of aligning the nature of the test assets with the specific evaluation context.

Second, we observe that efficiency test results are largely consistent across different factor constructions, including those using component portfolios rather than the aggregated SMB and HML factors. This suggests that the Fama-French factor portfolios may not be unique representations of the underlying state variables. The persistence of significant pricing errors in the almost same portfolios even under the component factor models offer two major future research opportunities. One suggestion is that such biases documented in the literature may not stem solely from the mismeasurement of SMB or HML, but from broader model misspecification. The other is that the original FF factors, MKT, SMB and HML still remain effective in explaining test portfolios returns.

These insights have important implications for mutual fund performance evaluation. Specifically, analysts should (1) carefully select appropriate test assets that reflect the investment style of the fund, (2) examine the robustness of the factor specification, and (3) adjust for pricing errors to avoid attributing spurious alpha to managerial skill. This is particularly critical for funds whose exposures resemble test portfolios known to produce significant intercepts, such as P (1,1), P (5,1), or CP2 and CP3.

Moreover, our rolling-window time-series analysis reveals that portfolio efficiency is not static but varies across market regimes. We identify substantial inefficiencies during major financial crises - such as the Dot-com bubble (2000), the Global Financial Crisis (2008), and the COVID-19 pandemic (2020) - particularly when industry portfolios are used as test assets. This pattern likely reflects shifts in risk premia, investor sentiment, and macroeconomic uncertainty, all of which distort the risk-return tradeoff.

Together, these findings reinforce the importance of a dynamic and context-sensitive approach to evaluating portfolio efficiency. Rigid application of standard factor models—without regard for test asset alignment, factor robustness, and temporal variation—may yield misleading conclusions in both academic research and practical performance evaluation.

References

Akhtaruzzaman, M., Boubaker, S., Sensoy, A., 2021. Financial contagion during COVID–19 crisis, Financial Research Letters, 101604.

Ang, A., Bekaert, G., 2002. International asset allocation with regime shifts. Review of Financial Studies 15, 1137-1187.

Aloui, R., Aissa, M., Nguyen, K., 2011. Global financial crisis, extreme interdependences, and contagion effects: the role of economic structure? Journal of Banking & Finance 35, 130 – 141.

Ao, M., Yingying, L., Zheng, X., 2019. Approaching mean-variance efficiency for large portfolios. Review of Financial Studies 32, 2890 – 2919.

Beine, M., Cosma, A., Vermeulen, R., 2010. The dark side of global integration: Increasing tail dependence. Journal of Banking and Finance 34, 184–192.

Berk, J.B., 2000. Sorting out sorts. Journal of Financ. 55, 407–427.

Cremers, M., Petajisto, A., Zitzewitz, E., 2012. Should benchmark indices have alpha? Revisiting performance evaluation, Critical Financial Review 2, 1 – 48.

Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. Journal of Finance 52, 1 – 33.

Fama, E., 1996. Multifactor portfolio efficiency and multifactor asset pricing. Journal of Financial and Quantitative Analysis 31, 441 – 465.

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

Fama, E., French, K., 1996. Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55-84.

Fama, E., French, K., 1997. Industry cost of equity. Journal of Financial Economics 43, 153 – 193.

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

Fama, E., French, K., 2018, Choosing factors, Journal of Financial Economics 128, 234–252.

Ferson, W., Harvey, C., 1999. Conditioning variables and the cross section of stock returns. Journal of Finance 54, 1325–60.

Ferson, W., Kandel, S., Stambaugh, R., 1987. Tests of asset pricing with time-varying expected risk premiums and market betas. Journal of Finance, 42, 201-220.

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

Gibbons, M., Ross, S., Shanken, J., 1989. A test of the efficiency of a given portfolio, Econometrica 57, 1121 – 1152.

Gibbons, M., Shanken, J., 1987. Subperiod aggregation and the power of multivariate tests of portfolio efficiency. Journal of Financial Economics 19, 389 – 394.

Harvey, C., Liu, Y., 2021. Lucky factors. Journal of Financial Economics 141, 413–435.

Huberman, G., Kandel, S., 1987. Mean-variance spanning, Journal of Finance 42, 873 – 888.

Jobson, J., Korkie, B., 1989. A Performance interpretation of multivariate tests of asset set intersection, spanning, and mean-variance efficiency, Journal of Financial and Quantitative Analysis 24, 185 – 204.

Lewellen, J., Nagel, S., & Shanken, J., 2010. A skeptical appraisal of asset pricing tests. Journal of Financial Economics, 96(2), 175–194.

Lettau, M., Ludvigson, S., 2001. Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. Journal of Political Economy 110, 777-804.

Liew, J., Vassalou, M., 2000. Can book-to-market, size and momentum be risk factors that predict economic growth? Journal of Financial Economics 57, 221 – 245.

Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets. Journal of Finance 56, 649-676.

Mackinlay, A., 1985. An analysis of multivariate financial tests. Ph.D. Dissertation, Graduate School of Business, University of Chicago, Chicago, IL.

Merton, R., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867 - 887.

Muir, T., 2017. Financial crises and risk premia. The Quarterly Journal of Economics, 765 – 809.

Pesaran, M., Timmermann, A., 1995. Predictability of stock returns: robustness and economic significance. Journal of Finance 50, 1201-1228.

Pukthuanthong, K. , Roll, R. , Subrahmanyam, A. , 2019. A protocol for factor identification. Review of Financial Studies 32, 1573–1607 .

Tronzano, 2022. Optimal portfolio allocation between global stock indexes and safe haven assets: Gold versus the Swiss Franc (1999–2021), Journal of Risk and Financial Management 15, 241.

Wallmeier, M., & Tauscher, K., 2012. Portfolio Overlapping Bias in Tests of the Fama-French Three-Factor Model. European Financial Management, 18(1), 1-29.

Wen, X., Wei, Y., Huang, D., 2012. Measuring contagion between energy market and stock market during financial crisis: a copula approach. Energy Econ. 34, 1435–1446.

Table 1: Regressions coefficients of excess 25 FF stock portfolio returns on monthly excess market returns (Rm – Rf) and the size (SMB) and book-to-market equity (HML) factors for the period from 1986 to 2022. SMB is the difference between the average of the returns on the small-stock portfolios and the average of the returns on the big-stock portfolios. HML is the difference between the average of the returns on the high-BE/ME portfolios and the average of the returns on the low-BE/ME portfolios. BE/ME is the book-to-market equity ratio. Coefficient estimates, their t-stats, R^2, and residual standard errors are reported. We estimate the regression coefficients from the following three-factor model.

$$R - Rf = a + b(Rm - Rf) + sSMB + hHML + e$$

Book-to-Market quintiles

	Low	2	3	4	High	Low	2	3	4	High
		a					t(a)			
Small	-0.62	0.07	0.00	0.23	0.25	-4.79	0.67	-0.05	3.37	2.27
2	-0.17	0.07	0.05	0.04	-0.02	-2.03	1.02	0.73	0.62	-0.28
3	-0.09	0.12	-0.01	0.06	0.05	-1.10	1.49	-0.16	0.83	0.47
4	0.15	0.07	-0.03	0.02	-0.06	1.95	0.87	-0.34	0.27	-0.61
Big	0.17	0.07	0.05	-0.27	-0.19	3.68	0.96	0.61	-3.47	-1.50
		b					t(b)			
Small	1.11	0.96	0.91	0.88	0.93	38.65	42.71	61.97	57.93	38.22
2	1.13	1.01	0.96	0.93	1.09	59.08	63.78	58.59	66.85	73.23
3	1.10	1.02	0.97	1.00	1.09	62.61	57.87	54.56	57.61	50.88
4	1.07	1.04	1.03	1.01	1.11	62.88	58.70	52.82	51.83	48.75
Big	0.99	0.95	0.93	1.03	1.17	95.49	61.37	53.84	58.42	40.42
		s					t(s)			
Small	1.38	1.33	1.04	1.06	1.06	31.88	39.19	46.93	46.50	28.98
2	1.03	0.91	0.71	0.75	0.93	35.83	38.22	28.69	35.49	41.34
3	0.76	0.57	0.39	0.43	0.52	28.87	21.44	14.53	16.37	16.09
4	0.45	0.21	0.14	0.20	0.26	17.46	8.03	4.93	6.82	7.50
Big	-0.26	-0.20	-0.19	-0.21	-0.20	-16.62	-8.41	-7.20	-7.98	-4.48
		h					t(h)			
Small	-0.26	-0.01	0.28	0.50	0.72	-6.26	-0.41	13.36	22.77	20.36
2	-0.30	0.14	0.42	0.60	0.84	-10.90	6.02	17.69	29.69	38.77
3	-0.38	0.17	0.44	0.65	0.86	-14.84	6.54	17.04	25.64	27.83
4	-0.36	0.25	0.45	0.56	0.84	-14.69	9.68	16.07	19.67	25.51
Big	-0.33	0.10	0.33	0.71	0.86	-21.88	4.30	13.41	27.76	20.50
		R2					s(e)			
Small	0.89	0.91	0.94	0.94	0.87	2.68	1.78	1.63	1.58	0.96
2	0.94	0.94	0.92	0.94	0.95	2.10	1.47	1.65	1.65	1.45
3	0.94	0.91	0.89	0.91	0.89	1.37	1.53	1.65	1.82	1.61
4	0.93	0.90	0.88	0.88	0.87	1.41	1.30	1.62	1.82	1.64
Big	0.96	0.90	0.87	0.90	0.81	2.26	1.39	1.99	2.13	2.69

Table 2: Regression estimation of the FF three-factor model with alternative test portfolios - FF 10 cashflow/price (CP) decile portfolios and 30 FF industry portfolios for the period from 1986 to 2022.

Panel A: Regression coefficient estimates with their t-values and R^2 for 10 FF CP portfolios

	a	b	s	h	t(a)	t(b)	t(s)	t(h)	R^2
CP1	0.06	1.12	0.00	-0.41	0.79	62.53	0.01	-15.85	0.92
CP2	0.15	0.97	-0.13	-0.12	2.12	62.17	-5.59	-5.16	0.90
CP3	0.18	0.96	-0.21	-0.02	2.43	57.66	-8.52	-0.69	0.89
CP4	0.04	0.96	-0.14	0.17	0.55	55.40	-5.42	6.88	0.88
CP5	-0.03	0.96	-0.08	0.22	-0.42	53.27	-2.91	8.57	0.87
CP6	-0.02	0.95	-0.10	0.37	-0.30	54.40	-3.92	14.81	0.87
CP7	0.13	0.86	-0.01	0.39	1.22	37.67	-0.15	11.67	0.77
CP8	0.06	0.95	-0.08	0.47	0.77	52.08	-2.97	17.72	0.87
CP9	-0.02	0.95	0.00	0.50	-0.23	46.46	0.02	17.00	0.84
CP10	-0.05	1.08	0.16	0.58	-0.38	39.87	3.84	14.92	0.80

Panel B: Regression coefficient estimates with their t-values and R^2 for 30 FF industry portfolios

	a	b	s	h	t(a)	t(b)	t(s)	t(h)	R^2
Food	0.30	0.67	-0.31	0.15	2.02	20.59	-6.33	3.10	0.50
Beer	0.50	0.68	-0.40	-0.03	2.69	16.29	-6.29	-0.47	0.39
Smoke	0.61	0.69	-0.35	0.27	2.12	10.83	-3.59	2.97	0.22
Games	-0.04	1.25	0.30	0.11	-0.23	28.87	4.58	1.81	0.69
Books	-0.44	1.10	0.19	0.43	-2.82	31.87	3.60	8.50	0.72
Hshld	0.19	0.72	-0.22	0.04	1.28	21.76	-4.43	0.90	0.52
Clths	0.01	1.11	0.08	0.22	0.06	23.70	1.18	3.22	0.58
Hlth	0.36	0.77	-0.21	-0.19	2.52	24.07	-4.44	-4.21	0.59
Chem	-0.07	1.11	-0.02	0.39	-0.43	32.74	-0.43	7.97	0.72
Txtls	-0.39	1.20	0.62	0.86	-1.48	20.53	7.06	10.18	0.57
Cntr	-0.14	1.19	0.23	0.47	-0.91	34.67	4.52	9.38	0.75
Steel	-0.42	1.43	0.53	0.42	-1.72	26.13	6.48	5.35	0.65
FabPr	-0.03	1.22	0.39	0.28	-0.18	35.41	7.54	5.62	0.77
ElcEq	0.03	1.27	0.03	0.05	0.19	35.81	0.54	1.02	0.76
Autos	-0.27	1.40	0.21	0.41	-0.96	22.38	2.21	4.51	0.55
Carry	0.01	1.10	0.01	0.51	0.05	25.72	0.10	8.23	0.62
Mines	-0.06	0.88	0.32	0.40	-0.20	12.08	2.92	3.78	0.29
Coal	-0.01	1.08	0.59	0.62	-0.02	9.41	3.43	3.74	0.21
Oil	0.02	0.92	0.08	0.72	0.11	18.08	1.06	9.80	0.48
Util	0.24	0.50	-0.25	0.27	1.51	14.18	-4.73	5.29	0.34
Telcm	-0.10	0.93	-0.19	0.04	-0.72	28.57	-3.85	0.77	0.65
Servs	0.20	1.14	0.12	-0.55	1.71	43.71	2.95	-14.60	0.85
BusEq	0.09	1.24	0.29	-0.52	0.55	32.29	5.07	-9.29	0.76
Paper	-0.12	0.95	-0.05	0.32	-0.82	30.13	-1.03	6.92	0.68
Trans	-0.07	1.01	0.06	0.35	-0.48	29.87	1.17	7.14	0.68
Whlsl	-0.06	0.93	0.26	0.26	-0.53	35.80	6.67	6.79	0.77
Rtail	0.20	0.97	-0.05	-0.10	1.32	29.20	-1.09	-2.12	0.68
Meals	0.18	0.89	-0.06	0.19	1.12	25.01	-1.20	3.68	0.59
Fin	-0.16	1.14	-0.11	0.60	-1.55	51.04	-3.27	18.51	0.86
Other	-0.35	1.02	-0.04	0.27	-2.24	28.98	-0.71	5.22	0.66

Table 3: Portfolio efficiency with 25 FF size/value portfolios for all observations from 1986 t0 2022, given the FF 3-factor model. GRS is the F-statistic of Gibbons, Ross, and Shanken (1989), testing the hypothesis that the regression intercepts for test assets are all zero. p(GRS) is the p-value of GRS. Avg |a| is the average of the absolute value of the intercepts.

Panel A: All 444 monthly observations for the period from 1986 to 2022

	GRS	p(GRS)	Avg |a|	R^2
1986-2022	4.327	0.000	0.117	0.907

Panel B: 5-year sub-periods of 60 monthly return observations

Periods	GRS	p(GRS)	Avg |a|	R^2
1986 - 1990	2.913	0.002	0.205	0.962
1991 - 1995	2.597	0.006	0.178	0.884
1996 - 2000	1.364	0.202	0.355	0.889
2001 - 2005	1.499	0.139	0.204	0.894
2006 - 2010	1.478	0.148	0.222	0.950
2011 - 2015	1.110	0.386	0.140	0.930
2016 - 2020	0.962	0.534	0.143	0.944

Table 4: Portfolio efficiency with 10 FF CP portfolios for all observations from 1986 to 2022, given the FF three-factor model.

Panel A: All 444 monthly observations for the period from 1986 to 2022

	GRS	p(GRS)	Avg |a|	R^2
1986 - 2022	1.333	0.210	0.075	0.861

Panel B: 5-year subperiods of 60 monthly return observations

Periods	GRS	p(GRS)	Avg |a|	R^2
1986 - 1990	1.146	0.350	0.129	0.956
1991 - 1995	0.885	0.554	0.097	0.864
1996 - 2000	1.125	0.365	0.252	0.806
2001 - 2005	0.544	0.849	0.113	0.829
2006 - 2010	1.154	0.345	0.203	0.904
2011 - 2015	2.300	0.027	0.140	0.887
2016 - 2020	1.305	0.256	0.252	0.915

Table 5: Portfolio efficiency with 30 FF industry portfolios for all observations from 1986 to 2022, given the FF three-factor model. Refer to Table 3 for the definitions of the variables ans statistics.

Panel A: All 444 monthly observations for the period from 1986 to 2022

	GRS	p(GRS)	Avg |a|	R^2
1986 - 2022	1.718	0.012	0.189	0.603

Panel B: 5-year sub-periods of 60 monthly return observations

Periods	GRS	p(GRS)	Avg |a|	R^2
1986 - 1990	1.589	0.114	0.436	0.796
1991 - 1995	0.888	0.626	0.365	0.572
1996 - 2000	3.586	0.001	0.738	0.519
2001 - 2005	0.601	0.912	0.344	0.567
2006 - 2010	1.123	0.382	0.439	0.736
2011 - 2015	1.374	0.204	0.651	0.662
2016 - 2020	2.838	0.004	0.386	0.693

Table 6: Regression results of the component factors against the remaining component factors using the sample period of 1986 - 2022. L, H, S, B, and M indicate the excess returns of the low book-to-market portfolio, high book-to-market portfolio, small portfolio, big portfolio, and market portfolio, respectively. The numbers in the parentheses are t-values below the estimated regression coefficients.

							R^2
L	=	-0.06	0.84 M	0.75 S	-0.53 H		0.99
		(-1.83)	(56.43)	(53.24)	(-35.31)		
H	=	-0.04	1.26 M	1.2 S	-1.38 L		0.96
		(-0.87)	(37.23)	(49.03)	(-35.31)		
S	=	0.05	-0.91 M	0.71 H	1.15 L		0.98
		(1.38)	(-30.21)	(49.03)	(53.24)		
M	=	0.07	-0.74 S	0.61 H	1.05 L		0.98
		(2.02)	(-30.21)	(37.23)	(56.43)		
M	=	0.03	-0.02 S	0.77 B	-0.18 H	0.39 L	0.99
		(1.11)	(-0.56)	(22.8)	(-4.93)	(12.57)	

Table 7: GRS F-tests with original and component factors: (M, SMB, HML), (M, S, H), (M, S, L), (M, H, L), and (S, B, H, L) for 1986 – 2022. Refer to Table 3 for the definitions of factors. P(GRS) is the p-value of GRS.

Panel A: GRS tests with 25 size/value portfolios as test portfolios

	Factors	GRS	p(GRS)	Avg |a|	R^2
25 Size/Value	M SMB HML	4.327	0.0	0.117	0.907
25 Size/Value	M S H	4.430	0.0	0.126	0.907
25 Size/Value	M S L	4.396	0.0	0.111	0.904
25 Size/Value	M H L	4.432	0.0	0.131	0.900
25 Size/Value	S B H L	4.211	0.0	0.109	0.927

Panel B: GRS F-tests with 10 CP portfolios as test assets

	Factors	GRS	p(GRS)	Avg |a|	R^2
10 C/P	M SMB HML	1.333	0.210	0.075	0.861
10 C/P	M S H	1.215	0.279	0.069	0.856
10 C/P	M S L	1.502	0.136	0.082	0.857
10 C/P	M H L	1.201	0.288	0.073	0.859
10 C/P	S B H L	1.775	0.063	0.080	0.880

Table 8: Regression results of FF 10 CP test portfolios against the component factors of the three-factor model for the period 1986 - 2022.

Panel A: The components factors are M, S, and H.

	a	b	s	h	t(a)	t(b)	t(s)	t(h)	R^2
CP1	0.04	1.45	0.29	-0.59	0.43	35.32	7.30	-14.21	0.91
CP2	0.14	1.15	-0.07	-0.10	1.99	32.64	-2.07	-2.82	0.90
CP3	0.18	1.10	-0.24	0.10	2.36	29.32	-6.81	2.54	0.88
CP4	0.05	0.89	-0.28	0.33	0.68	22.83	-7.49	8.42	0.87
CP5	-0.02	0.81	-0.22	0.36	-0.22	19.82	-5.79	8.76	0.86
CP6	0.00	0.69	-0.38	0.62	-0.01	17.43	-10.17	15.37	0.87
CP7	0.15	0.50	-0.23	0.57	1.45	9.69	-4.69	10.80	0.77
CP8	0.09	0.59	-0.42	0.75	1.08	14.14	-10.65	17.92	0.86
CP9	0.01	0.48	-0.34	0.77	0.08	10.64	-7.77	16.56	0.84
CP10	-0.01	0.45	-0.20	0.79	-0.08	7.42	-3.38	12.67	0.80

Panel B: The component factors are M, S, and L.

	a	b	s	l	t(a)	t(b)	t(s)	t(l)	R2
CP1	0.09	0.55	-0.53	1.07	1.17	10.02	-13.51	16.88	0.92
CP2	0.16	0.96	-0.23	0.23	2.22	19.62	-6.44	4.02	0.90
CP3	0.19	1.18	-0.16	-0.06	2.41	22.03	-4.17	-0.91	0.88
CP4	0.04	1.33	0.14	-0.50	0.44	24.00	3.44	-7.73	0.87
CP5	-0.05	1.34	0.26	-0.62	-0.60	23.66	6.40	-9.46	0.87
CP6	-0.04	1.54	0.41	-0.97	-0.51	27.36	10.03	-14.87	0.87
CP7	0.10	1.33	0.53	-0.97	1.01	18.68	10.26	-11.73	0.78
CP8	0.04	1.62	0.54	-1.17	0.49	27.06	12.49	-16.90	0.85
CP9	-0.04	1.52	0.63	-1.18	-0.41	23.07	13.23	-15.35	0.83
CP10	-0.07	1.60	0.86	-1.34	-0.62	19.12	14.09	-13.77	0.81

Panel C: The component factors are M, H, and L.

	a	b	h	l	t(a)	t(b)	t(h)	t(l)	R2
CP1	0.07	1.05	-0.39	0.45	0.83	16.21	-12.64	9.83	0.92
CP2	0.14	1.15	-0.15	-0.03	2.00	20.02	-5.38	-0.80	0.90
CP3	0.17	1.28	-0.07	-0.24	2.22	20.37	-2.46	-5.32	0.88
CP4	0.04	1.16	0.14	-0.34	0.47	18.21	4.50	-7.37	0.87
CP5	-0.04	1.08	0.20	-0.32	-0.46	16.46	6.45	-6.78	0.87
CP6	-0.03	1.10	0.35	-0.50	-0.35	17.35	11.36	-10.93	0.87
CP7	0.13	0.82	0.40	-0.36	1.24	9.87	10.11	-6.06	0.77
CP8	0.06	1.04	0.46	-0.55	0.74	15.64	14.33	-11.45	0.87
CP9	-0.02	0.86	0.53	-0.45	-0.19	11.64	14.97	-8.43	0.84
CP10	-0.04	0.78	0.65	-0.36	-0.30	7.96	13.72	-5.05	0.80

Panel D: The component factors are S, B, H, and L.

	a	s	b	h	l	t(a)	t(s)	t(b)	t(h)	t(l)	R2
CP1	0.12	-0.70	0.23	0.08	1.44	1.48	-5.50	1.91	0.59	12.94	0.91
CP2	0.16	0.03	1.06	-0.47	0.34	2.34	0.27	10.50	-4.41	3.67	0.91
CP3	0.18	0.08	1.27	-0.51	0.12	2.63	0.75	12.17	-4.61	1.25	0.90
CP4	0.03	0.50	1.52	-0.67	-0.37	0.44	4.49	14.42	-5.98	-3.80	0.90
CP5	-0.06	0.85	1.72	-0.91	-0.67	-0.79	7.51	16.11	-7.97	-6.78	0.90
CP6	-0.03	0.53	1.46	-0.46	-0.57	-0.46	4.64	13.46	-3.97	-5.68	0.89
CP7	0.08	1.27	1.90	-1.07	-1.17	0.91	8.97	14.18	-7.54	-9.47	0.83
CP8	0.05	0.63	1.52	-0.44	-0.73	0.64	5.34	13.62	-3.71	-7.09	0.89
CP9	-0.03	0.62	1.35	-0.31	-0.68	-0.38	4.61	10.59	-2.33	-5.83	0.86
CP10	-0.06	0.92	1.42	-0.41	-0.85	-0.46	4.94	8.03	-2.15	-5.18	0.81

Table 9: Changes in portfolio efficiency during 1980 – 2022. We provide information on portfolio efficiency using five-year monthly observations to produce GRS statistics, absolute mean values of the intercept estimates, and R squared for each five-year rolling window. FF three-factor model is used here.

Panel A: GRS test results for 25 FF size/value portfolios

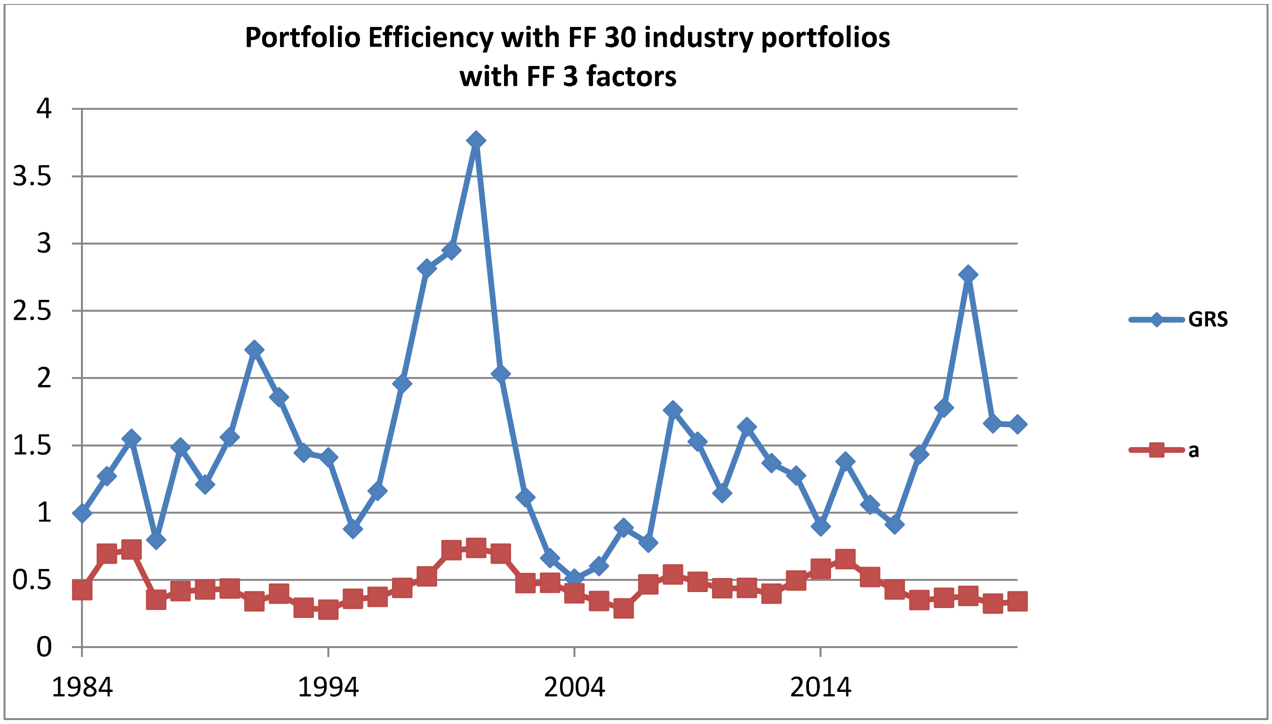
	start	end	GRS	p(GRS)	Avg|a|	R^2
Period 1	1980	1984	1.261	0.265	0.202	0.926
Period 2	1981	1985	2.069	0.027	0.295	0.924
Period 3	1982	1986	2.151	0.021	0.251	0.933
Period 4	1983	1987	1.585	0.109	0.202	0.955
Period 5	1984	1988	2.856	0.003	0.218	0.959
Period 6	1985	1989	2.163	0.020	0.179	0.960
Period 7	1986	1990	3.026	0.002	0.205	0.962
Period 8	1987	1991	2.377	0.011	0.176	0.957
Period 9	1988	1992	2.272	0.015	0.190	0.926
Period 10	1989	1993	1.768	0.064	0.162	0.922
Period 11	1990	1994	1.781	0.062	0.185	0.922
Period 12	1991	1995	2.447	0.009	0.179	0.885
Period 13	1992	1996	3.354	0.001	0.189	0.866
Period 14	1993	1997	2.520	0.007	0.212	0.882
Period 15	1994	1998	2.592	0.006	0.243	0.924
Period 16	1995	1999	1.814	0.056	0.332	0.906
Period 17	1996	2000	1.388	0.189	0.352	0.891
Period 18	1997	2001	1.395	0.186	0.301	0.876
Period 19	1998	2002	1.505	0.137	0.271	0.880
Period 20	1999	2003	1.978	0.035	0.334	0.873
Period 21	2000	2004	1.715	0.075	0.284	0.875
Period 22	2001	2005	1.476	0.149	0.210	0.895
Period 23	2002	2006	1.471	0.151	0.216	0.913
Period 24	2003	2007	1.687	0.081	0.136	0.909
Period 25	2004	2008	1.887	0.045	0.251	0.934
Period 26	2005	2009	1.223	0.292	0.224	0.947
Period 27	2006	2010	1.444	0.162	0.223	0.951
Period 28	2007	2011	1.901	0.044	0.240	0.953
Period 29	2008	2012	1.321	0.227	0.202	0.950
Period 30	2009	2013	2.021	0.031	0.196	0.947
Period 31	2010	2014	1.926	0.040	0.167	0.938
Period 32	2011	2015	1.110	0.386	0.140	0.930
Period 33	2012	2016	2.073	0.026	0.113	0.911
Period 34	2013	2017	2.386	0.011	0.133	0.913
Period 35	2014	2018	2.737	0.004	0.116	0.924
Period 36	2015	2019	1.393	0.187	0.120	0.929
Period 37	2016	2020	0.977	0.518	0.135	0.946
Period 38	2017	2021	1.386	0.190	0.155	0.930
Period 39	2018	2022	1.474	0.149	0.153	0.937

Panel B: GRS test results with thirty FF industry portfolios as test portfolios

	start	end	GRS	p(GRS)	Avg|a|	R^2
Period 1	1980	1984	0.992	0.511	0.423	0.681
Period 2	1981	1985	1.268	0.268	0.694	0.681
Period 3	1982	1986	1.548	0.127	0.724	0.709
Period 4	1983	1987	0.795	0.730	0.351	0.782
Period 5	1984	1988	1.483	0.152	0.415	0.797
Period 6	1985	1989	1.208	0.312	0.426	0.785
Period 7	1986	1990	1.558	0.124	0.435	0.795
Period 8	1987	1991	2.207	0.020	0.340	0.782
Period 9	1988	1992	1.855	0.054	0.395	0.686
Period 10	1989	1993	1.443	0.169	0.292	0.676
Period 11	1990	1994	1.409	0.186	0.278	0.667
Period 12	1991	1995	0.877	0.638	0.358	0.573
Period 13	1992	1996	1.160	0.351	0.373	0.513
Period 14	1993	1997	1.955	0.041	0.440	0.546
Period 15	1994	1998	2.813	0.004	0.525	0.625
Period 16	1995	1999	2.947	0.003	0.720	0.567
Period 17	1996	2000	3.764	0.000	0.736	0.520
Period 18	1997	2001	2.029	0.033	0.694	0.517
Period 19	1998	2002	1.112	0.392	0.475	0.522
Period 20	1999	2003	0.660	0.865	0.476	0.504
Period 21	2000	2004	0.507	0.964	0.399	0.517
Period 22	2001	2005	0.601	0.912	0.341	0.567
Period 23	2002	2006	0.887	0.626	0.286	0.562
Period 24	2003	2007	0.774	0.753	0.465	0.520
Period 25	2004	2008	1.759	0.071	0.538	0.636
Period 26	2005	2009	1.526	0.135	0.484	0.701
Period 27	2006	2010	1.142	0.366	0.437	0.736
Period 28	2007	2011	1.636	0.100	0.439	0.759
Period 29	2008	2012	1.367	0.208	0.397	0.751
Period 30	2009	2013	1.274	0.264	0.493	0.728
Period 31	2010	2014	0.896	0.617	0.581	0.695
Period 32	2011	2015	1.377	0.202	0.652	0.662
Period 33	2012	2016	1.056	0.445	0.519	0.612
Period 34	2013	2017	0.910	0.602	0.427	0.593
Period 35	2014	2018	1.431	0.175	0.348	0.590
Period 36	2015	2019	1.777	0.068	0.366	0.616
Period 37	2016	2020	2.768	0.005	0.380	0.694
Period 38	2017	2021	1.661	0.093	0.322	0.688
Period 39	2018	2022	1.655	0.095	0.338	0.713

Figure 1.

GRS stands for GRS multifactor efficiency test statistics. “a” represents the average of the intercept estimates. We employ as test assets 30 FF industry portfolios and 25 FF size/value portfolios.



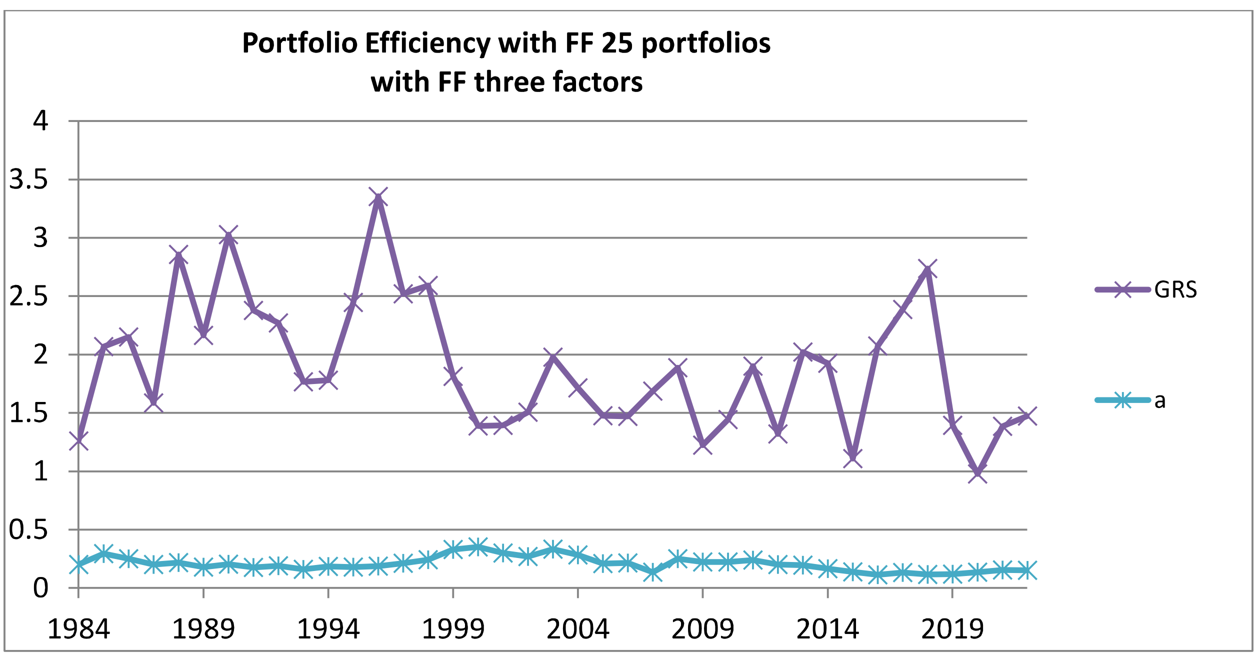


Table 10: The intercept estimates and their t - values of 25 size-value sorted portfolios for FF three factors and alternative MMV component factors.

	Low	2	3	4	High	Low	2	3	4	High
		a					t(a)			
Panel A: FF 3 factors – M, SMB, and HML										
Small	-0.62	0.07	0.00	0.23	0.25	-4.79	0.67	-0.05	3.37	2.27
2	-0.17	0.07	0.05	0.04	-0.02	-2.03	1.02	0.73	0.62	-0.28
3	-0.09	0.12	-0.01	0.06	0.05	-1.10	1.49	-0.16	0.83	0.47
4	0.15	0.07	-0.03	0.02	-0.06	1.95	0.87	-0.34	0.27	-0.61
Big	0.17	0.07	0.05	-0.27	-0.19	3.68	0.96	0.61	-3.47	-1.50
Panel B: MSH factors										
Small	-0.65	0.06	0.01	0.25	0.28	-4.79	0.55	0.14	3.45	2.42
2	-0.20	0.08	0.08	0.07	0.02	-2.48	1.29	1.15	1.16	0.33
3	-0.12	0.13	0.02	0.10	0.09	-1.44	1.76	0.19	1.32	0.96
4	0.12	0.09	0.00	0.06	-0.02	1.56	1.10	0.00	0.63	-0.23
Big	0.15	0.08	0.07	-0.24	-0.17	3.01	1.06	0.89	-2.97	-1.44
Panel C: MSL factors										
Small	-0.54	0.11	0.01	0.23	0.24	-4.42	1.03	0.15	3.15	2.05
2	-0.11	0.09	0.04	0.02	-0.03	-1.71	1.46	0.65	0.37	-0.39
3	-0.03	0.12	-0.03	0.04	0.03	-0.48	1.61	-0.41	0.57	0.28
4	0.19	0.06	-0.05	0.01	-0.07	2.51	0.75	-0.63	0.07	-0.66
Big	0.19	0.05	0.02	-0.29	-0.18	4.30	0.74	0.22	-3.07	-1. 12
Panel D: MHL factors										
Small	-0.51	0.16	0.07	0.30	0.32	-4.13	1.41	0.85	3.29	2.51
2	-0.09	0.14	0.10	0.09	0.05	-1.44	1.87	1.25	1.17	0.59
3	-0.03	0.16	0.01	0.10	0.08	-0.39	1.94	0.18	1.21	0.90
4	0.18	0.08	-0.02	0.04	-0.04	2.49	1.06	-0.21	0.47	-0.38
Big	0.16	0.05	0.03	-0.28	-0.19	3.06	0.77	0.44	-3.38	-1.41