JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 53, No. 6, Dec. 2018, pp. 2335–2354

doi:10.1017/S0022109018000376

When Factors Do Not Span Their Basis Portfolios

Mark Grinblatt and Konark Saxena*

Abstract

To price assets with a parsimonious set of factor-mimicking portfolios, one typically identifies and weights well-diversified basis portfolios. Traditional weightings lead to factor-mimicking portfolios that are unlikely to price even the basis portfolios from which they are formed. We offer a method to combine basis portfolios into a single factor-mimicking portfolio that is closely linked to the optimal portfolio. In practice, this method improves the pricing accuracy of parsimonious factor models, even for anomaly portfolios formed from characteristics that are distinct from those underlying the basis portfolios.

I. Introduction

Virtually every modern asset pricing theory can be viewed as a set of assumptions that ultimately identify an optimal investment portfolio or a linear space that contains an optimal portfolio. The asset pricing restrictions are merely first-order conditions for optimality. When theory is explicit about the optimal portfolio, as in the case of the Sharpe (1964)–Lintner (1965) capital asset pricing model (CAPM), empirical asset pricing tests are fairly straightforward. For example, Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) show how to conduct time-series and cross-sectional tests of the CAPM, respectively. Here, the definition of optimality is a maximum Sharpe ratio, and the CAPM empirical tests merely assess whether the market portfolio satisfies that optimization criterion.

The CAPM (and its multiperiod cousins) fail to explain several prominent cross-sectional return anomalies. This failing has led empirical asset pricing research toward multifactor models in hope of a better fit with the cross section

*Grinblatt, mark.grinblatt@anderson.ucla.edu, University of California at Los Angeles Anderson School of Management; Saxena (corresponding author), k.saxena@unsw.edu.au, University of New South Wales School of Banking and Finance. We thank Kenneth French for making the data available, as well as an anonymous referee, Hendrik Bessembinder (the editor), Tarun Chordia, Bhagwan Chowdhry, Robert Faff, David Feldman, Olivier Ledoit, Richard Roll, Avanidhar Subrahmanyam, Terry Walter, Chu Zhang, and seminar participants at Hong Kong Polytechnic University, University of South Australia, and the University of New South Wales for helpful comments on earlier drafts.

of average returns. Starting with Huberman and Kandel (1985), these multifactor models generally employ factor-mimicking portfolios tied to anomaly-related characteristics that the CAPM cannot explain. The most popular versions of these multifactor models, like the Fama and French (1993) 3-factor model,[1] combine a market factor with long–short spread portfolios that group stocks into long and short categories, referred to as "basis portfolios," from multidimensional rankings of their firms' attributes. Although weighting the longs and shorts equally is appealing for its simplicity, there is no statistical or theoretical foundation for the equal magnitude weighting of basis portfolios. We refer to multifactor models as "traditional" when they propose these kinds of spread portfolios as factors.

Our article shows that the pricing accuracy of most traditional multifactor models can be vastly improved by removing the restriction of equally weighting the longs and shorts within and across size groups. In particular, we propose and test (out of sample) a single-factor alternative to traditional factor models. The single-factor alternative applies the mathematics of mean–variance optimality to weight extreme basis portfolios using sample average returns and covariances. The single-factor alternative dominates its traditional counterpart at asset pricing. In short, our article shows that building a parsimonious set of priced factors that work well for asset pricing tests, valuation, and performance evaluation requires some deference to the mathematics of portfolio optimization.

The theoretical foundation of the traditional multifactor models is Ross's (1976) arbitrage pricing theory (APT). If the APT pricing equation holds exactly, as in Connor (1984), the alpha anomalies observed in the literature indicate that the optimal portfolio lies outside the span of the factor-mimicking portfolios implemented as benchmarks. The compelling logic of Ross's theory suggests this should not happen. In a K-factor return-generating model lacking arbitrage, the linear space spanned by K distinct[2] well-diversified benchmark portfolios contains the space spanned by any set of well-diversified test assets. By definition, this space includes the portfolio of the test assets with the highest Sharpe ratio.

So what went wrong with the empirical implementation? Ross's (1976) paper is vocal in suggesting that statistical techniques can identify both the number and identity of the factors. However, it is silent about pricing when the factor-mimicking portfolios are too few in number to span the factor space, a condition we refer to as "rank deficiency." When the implemented benchmarks are rank deficient, there may be a specific construction of the factor-mimicking benchmarks from basis portfolios that contains the optimal portfolio in its span. However, too many arbitrary restrictions on the construction of the factor benchmarks (e.g., the equal magnitude weightings of the long and short basis portfolios) are likely to prevent the factor-mimicking benchmarks from spanning the optimal portfolio. This leads to spurious nonzero alphas, which we derive in the Appendix. The empirical dilemma posed by the alpha anomalies cannot easily be resolved by supplementing the benchmark portfolios with additional factor-mimicking portfolios unless the number of distinct factor-mimicking portfolios is the same or

[1]Carhart (1997), Novy-Marx (2013), and Fama and French (2015) are other examples of multifactor models constructed using similar methods.

[2]Here, "distinct" means the factor beta vectors are linearly independent.

greater than the number of true factors. Empirical research tends to shun the latter solution as the number of true factors is large, making estimation of the many less important factors imprecise. Moreover, interpretation and practical implementation become a problem when the factor model contains large numbers of factors: A 20-factor model is unwieldy for both researchers and practitioners.

In the rank-deficient setting, one must be careful that restrictions on factor portfolio construction do not prevent the factor-mimicking portfolios from spanning the optimal portfolio. Some restrictions may be necessary for statistical power or practical implementation. However, equally weighting the longs and shorts within a size category or equally weighting across size categories are not well-thought-out restrictions. Such restrictions eliminate much of the influence of portfolio return data on the weights employed in the construction of rank-deficient factor mimicking portfolios.

For example, consider the Fama–French (1993) 3-factor model, which effectively forms its factors from 6 basis portfolios. The latter are value-weighted portfolios of firms in 2 size categories (above and below the New York Stock Exchange breakpoint median) intersected by 3 book-to-market categories. We refer to these 6 basis portfolios as small value, small neutral, small growth, large value, large neutral, and large growth. The market portfolio is a variation of combining all 6 portfolios. The high minus low (HML) factor portfolio is an equal weighting of 2 spread portfolios: large value minus large growth and small value minus small growth. Note that HML omits the 2 neutral portfolios and has the same magnitude weight on the remaining 4 portfolios: either $+1/2$ or $-1/2$. The small minus big (SMB) factor portfolio has the same magnitude weights on all 6 portfolios. The average returns of small growth and small value are the outliers of the 6 basis portfolios, but they have no greater weight in HML or SMB than the large-growth and large-value basis portfolios. The consequences of this equal weighting show up in empirical asset pricing tests: As Welch (2008) illustrates, the Fama–French 3-factor portfolios fail to price even the 6 basis portfolios from which they are derived. This failing of the 3-factor model is common to other factor models solely because of the weight restriction. Moreover, it is the weight restriction that accounts for many of the cross-sectional asset pricing anomalies we observe in the academic and practitioner literature.

The failures of various sets of traditional factor-mimicking portfolios to price even their basis portfolios has an unambiguous geometric interpretation: The optimal combination of the factor-mimicking portfolios has a significantly lower Sharpe ratio than the optimal combination of the basis portfolios from which they are created.[3] By proposing a better weighting scheme for factor portfolios, rooted in the mathematics of mean–variance optimality, we obtain factor portfolios that not only price their basis assets but also do an adequate job of pricing portfolios tied to a large number of cross-sectional return anomalies.

This article assesses the asset pricing efficacy of factor-mimicking portfolios with statistical tests developed by Gibbons, Ross, and Shanken (1989) (GRS)

[3]The proof is a small extension of a standard result. Because the basis portfolios span the factor-mimicking portfolios, the return space spanned by both the basis portfolios and the factor portfolios is the same as that spanned by the basis portfolios themselves.

(an F-statistic based on independent and identically distributed (i.i.d.) multivariate normal returns) and Ledoit and Wolf (2008) (LW) (a more robust bootstrap test). Both tests effectively compare the Sharpe ratios of ex post optimal combinations of 2 opportunity sets consisting of the set of factor-mimicking portfolios both with and without the test assets they are conjectured to price.[4] The more robust LW test allows for changing means and covariance matrices, as well as other sources of non-normally distributed unconditional returns.

We are careful to avoid a potential statistical bias in assessing the relative efficacy of our 1-factor alternative to traditional benchmarks. To understand the bias, consider what would happen if our 1-factor alternative to the 3-factor model consists of the ex post efficient combination of all 6 basis portfolios. In this case, it would price the 6 basis portfolios perfectly. However, there would be no statistical test of ex ante optimality, just a mathematical tautology relating ex post efficiency to its first-order condition. To conservatively address this issue, all comparisons are with a proposed single-factor alternative that is an out-of-sample ex post efficient combination of the basis portfolios.[5]

In an i.i.d. world, the out-of-sample estimates of the means, variance, and covariances that help determine the efficient combination of the basis portfolios are independent of the returns one is trying to price. This fact motivates a new approach that prices assets with efficient out-of-sample combinations of the basis portfolios. We refer to the benchmark produced with this approach as the "jackknife MVE benchmark" because the procedure uses the jackknife statistical methodology to estimate weights for a mean–variance efficient (MVE) benchmark. The jackknife MVE benchmark excludes only the time-series observation being priced from the information used to generate the benchmark. For instance, using monthly return data, we form a jackknife portfolio that prices assets in June 2005 with information gleaned from all months in the sample period, except for June 2005.[6] The 1-factor portfolio we implement as a benchmark thus has weights on the basis portfolios that vary each month, but only slightly for a long sample period.

We show that the jackknife MVE factor portfolio dominates the Fama–French (1993) 3-factor model for pricing assets. This finding generalizes to the other 3 traditional factor models we study. In all, there are 4 sets of traditional multifactor benchmarks, each with an associated jackknife MVE benchmark that is proposed as an alternative: the Fama–French 3-factor model based on size and book-to-market; the Carhart (1997) 4-factor model based on size, book-to-market, and momentum; the new Fama–French (2015) 5-factor model based on size, book-to-market, profitability, and investment; and a 6-factor model that augments the 5-factor model with the momentum factor.

[4]See Section 3 of GRS for this interpretation of their test.

[5]Kan and Smith (2008) show that the ex post frontier is generally on the left of the population frontier, providing an optimistic assessment of the population frontier. Hence, Sharpe ratio comparisons (in-sample) of out-of-sample optimized portfolios with in-sample optimized portfolios tend to find higher Sharpe ratios for the in-sample optimized portfolios.

[6]Jackknife estimates are common in statistics and have been used in finance, for example, in Basak, Jagannathan, and Ma (2009) who use them for estimating covariances to construct a minimum tracking error variance portfolio, but they have not been used to estimate asset pricing benchmarks.

https://doi.org/10.1017/S0022109018000376 Published online by Cambridge University Press

We illustrate the superior ability of the jackknife MVE benchmarks to price assets for several variations of the 4 traditional factor models. For example, we can form the 4 traditional factor benchmarks from basis portfolios distinguished by either coarse or finer groupings of firms ranked on their characteristics. We can also form the jackknife MVE portfolios from fine or coarse groupings of firms. However, when the grouping is fine, many MVE portfolio weights need to be estimated for basis portfolios that are highly multicollinear. Here, it is generally advisable to mitigate sampling error in estimating the jackknife MVE benchmarks' optimal basis portfolio weights by forcing some restrictions on the estimation. When the number of basis portfolios is large because the grouping of stocks is based on relatively fine distinctions, we force the jackknife MVE benchmark to place zero weight on basis portfolios with nonextreme amounts of a characteristic.

A benchmark that prices assets produces alphas of zero. Tests of zero alphas are essentially equivalent to Sharpe ratio comparisons, and we use both zero alpha tests and Sharpe ratio comparisons to make our point. We present this analysis in various tables described in Section II. Table 1 uses the t- and GRS F-tests of zero alpha to generalize the Welch (2008) finding: All 4 traditional factor models fail to price the basis portfolios from which they are constructed. Table 3, using both ordinary least squares (OLS) and Newey–West (1987) t-statistics, reveals that all of the jackknife MVE benchmarks have significant alphas when benchmarked against traditional factor portfolios. It also reveals that each jackknife MVE benchmark has a far higher Sharpe ratio than the optimal mix of the corresponding set of traditional factor portfolios. Table 4, using the GRS F-test, demonstrates that the jackknife MVE benchmarks, but not the traditional multifactor benchmarks (or variations of them), price 5×5 sets of basis portfolios that motivate and can be used to construct traditional factor-mimicking portfolios.

Table 5 effectively presents the same result as Table 4 using the LW Sharpe ratio comparison test. Table 6 uses the t-statistics of alphas and the GRS F-test to demonstrate the adequacy, as well as the superiority, of jackknife MVE benchmarks compared to their traditional factor-mimicking counterparts, at pricing decile spread portfolios associated with 10 anomalies. Finally, Table 7 uses the LW Sharpe ratio comparison test to show that the jackknife MVE benchmarks are far better than traditional factor-mimicking portfolios at pricing a set of 20 anomaly portfolios along with the basis portfolios.

II. Time-Series Tests of Various Asset Pricing Benchmarks

We now discuss time-series tests that address the ability of various sets of factor benchmarks to price assets. Data on monthly portfolio returns, obtained from the Kenneth French data library, are from Jan. 1973 to Dec. 2014. Descriptions of portfolio construction are found at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. One-month Treasury bill rates from the same data source are used to generate the excess returns of the basis portfolios needed for the pricing regressions. Table 1 studies the ability of 3 different factor models (3-, 4-, and 5-factor models) to price various sets of coarse basis portfolios that are the building blocks for factor portfolio construction. It reports

TABLE 1
Do Factors Span Their Basis Portfolios?

Table 1 reports alphas from regressing basis portfolio excess returns (above the risk-free rate) on associated factor portfolio returns. GRS is the F-statistic of Gibbons, Ross, and Shanken (1989), testing the hypothesis that all alphas for a group of portfolios are jointly zero. The GRS p-value is in square brackets. Panel A reports alphas of the 6 basis portfolios of the Fama–French (FF) (1993) 3-factor model (FF3). Panel B reports the alphas of the 12 basis portfolios of the Carhart (1997) 4-factor model. Panel C reports the alphas of the 18 basis portfolios of the Fama–French (2015) 5-factor model (FF5). Robust Newey and West (1987) t-statistics with 9 lags are reported in parentheses. MKT is market portfolio, HML is high minus low book-to-market (BTM) stocks, SMB is small minus big stocks, WML is winners minus losers portfolios, RMW is robust minus weak operating profitability (OP), CMA is conservative minus aggressive investments (INV), and MOM is momentum. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from July 1973 to Dec. 2014.

Panel A. Size and Book-to-Market Portfolios

	FF3 [MKT, SMB, HML]		
	BTM Rank		
Size Rank	Low	Medium	High
Small	−0.19***	0.09**	0.05**
	(−3.50)	(2.22)	(2.03)
Big	0.11***	−0.02	−0.13***
	(3.38)	(−0.28)	(−2.65)
GRS (all basis)		4.39***	
		[0.00]	

Panel B. Size, Book-to-Market, and Momentum

	Carhart Factors [MKT, SMB, HML, WML]					
	BTM Rank			MOM Rank		
Size Rank	Low	Medium	High	Low	Medium	High
Small	−0.15***	0.09***	0.06**	−0.12	0.11**	0.19***
	(−3.15)	(2.50)	(2.15)	(−1.62)	(2.04)	(3.42)
Big	0.11***	−0.02	−0.10**	0.25***	0.03	−0.06
	(3.55)	(−0.30)	(−2.07)	(3.22)	(0.49)	(−1.14)
GRS (subgroup)		3.65***			6.00***	
		[0.00]			[0.00]	
GRS (all basis)				4.96***		
				[0.00]		

Panel C. Size, Book-to-Market, Profitability, and Investment Portfolios

	FF5 [MKT, SMB, HML, RMW, CMA]								
	BTM Rank			OP Rank			INV Rank		
Size Rank	Low	Medium	High	Low	Medium	High	Low	Medium	High
Small	−0.09***	0.04	0.03	−0.08**	−0.00	−0.03	0.07*	0.08***	−0.10***
	(−2.38)	(1.31)	(1.15)	(−2.32)	(−0.05)	(−0.64)	(1.79)	(2.54)	(−2.92)
Big	0.03	−0.10*	−0.09**	0.07	−0.02	0.01	−0.06	−0.02	0.10***
	(0.92)	(−1.76)	(−1.97)	(1.54)	(−0.60)	(0.62)	(−1.32)	(−0.67)	(2.81)
GRS (subgroup)		2.23**			3.60***			3.81***	
		[0.04]			[0.00]			[0.00]	
GRS (all basis)					2.37***				
					[0.00]				

alphas from regressing excess returns (above the risk-free rate) of various basis portfolios on the factor portfolio returns.

Panel A of Table 1 studies the Fama and French (1993) 3-factor model and its ability to price the 6 basis portfolios that effectively combine to form the 3 Fama–French factors of market (MKT), HML, and SMB; Panel B studies the Carhart (1997) 4-factor alphas from the 12 basis portfolios that effectively form the 4 Carhart factors of MKT, HML, SMB, and winners minus losers (WML); and Panel C studies the Fama and French (2015) 5-factor model and its ability to price the 18 basis portfolios (or subsets of the 18) that effectively form the 5 factors, which consist of the traditional Fama–French factors MKT, HML, and SMB plus a robust minus weak (RMW) profitability factor and a conservative minus aggressive (CMA) investment factor. The t-statistics for the individual alphas test whether the alpha is significantly different from zero against the factors listed in the panel. GRS F-statistics test whether the alphas jointly differ from 0.

Panel A of Table 1 confirms the empirical finding of Welch (2008), indicating that the 3 Fama–French (1993) factors (MKT, HML, and SMB) generate significant alphas for the 6 basis portfolios from which they are formed.[7] The GRS F-statistic of 4.39, which is significant at the 1% level, indicates that the 6 alphas in Panel A jointly deviate from 0. Thus, even if the space spanned by all of the true factors is spanned by the 6 basis portfolios used to construct the 3 Fama–French 3-factor portfolios, the Fama–French factors do not span the optimal combination of their 6 basis portfolios.

Panel B of Table 1 confirms a similar result for the 4 Carhart (1997) factors. Moreover, the 4 Carhart factors cannot even price the 6 basis portfolios used to construct the 3-factor model benchmark. Of the 6 basis portfolios used for the 3-factor construction, 5 of the 6 have significant alphas against the 4 Carhart factors. The GRS test statistic on these portfolios, 3.65, strongly rejects zero alphas for all 6. Of the remaining 6 basis portfolios in Panel B, the 4-factor model again rejects zero alphas ($F = 6.00$), in part because large-cap losing stocks have a positive alpha of 25 basis points per month and small-cap winners have a positive alpha of 19 basis points per month. That is, the Carhart 4-factor model underpredicts the average return of both the large-cap losing stocks and small-cap winning stocks. All of the rejections of jointly zero alphas are at the 1% significance level or less, including the rejection ($F = 4.96$) of the 12 alphas of the Carhart model's basis portfolios in Panel B.

Panel C of Table 1 focuses on the Fama and French (2015) 5-factor model and its ability to price 18 basis portfolios: 6 based on size and book-to-market groupings, 6 on size and operating profitability (OP) groupings, and 6 on size and investment (INV) groupings. For each of the 3 subgroups of 6 basis portfolios, as well as for all 18 portfolios, the 5-factor model is rejected as an accurate asset pricing model. In particular, it has difficulty pricing small-cap firms, which are growth oriented or less profitable, as well as high-investment firms (both large and small cap). Combined with the results in Panels A and B, it appears that the ability of these parsimonious factor models to price assets is dismal.

Our key point here is that this failure is less an issue of the pricing ability of factor models per se, and more a consequence of the equal long–short weight restriction that is typically applied to construct factor portfolios. The question remains: Can we apply the mathematics of optimality (here, mean–variance efficiency) to obtain a better set of priced factors from the basis portfolios? The answer is yes. As we subsequently show, dispensing with the arbitrary restriction

[7]Moreover, Gerakos and Linnainmaa (2014) detect an unpriced component in the HML factor that distorts inferences, and Cremers, Petajisto, and Zitzewitz (2013) find that money managers earn nonzero alphas because of the weights used to construct the factors.

on portfolio weights in traditional constructions of factor-mimicking portfolios greatly improves the pricing accuracy of factor models.

Our new approach to factor construction begins with 4 sets of 5×5 basis portfolios. To contrast our approach with those of traditional factor design that use coarser basis portfolio groupings, we construct a set of traditional pseudofactors that mimic the traditional factors studied in Table 1. These are factors constructed from the returns of sets of 5×5 basis portfolios, as illustrated in Figure 1, used to more clearly illustrate that it is the weighting restrictions of the traditional factor models that account for their nonzero alphas. We cannot construct the actual factors used in the traditional factor models from 2-way sorts of quintile portfolios because the actual factor portfolios overlap cells. Note that the 8 symmetric cells with weights of $+1$ and -1 in the 5×5 arrays in Figure 1 are in the border rows and columns, indicating extreme quintiles of the attribute. Weighting extreme quintiles captures the characteristic while holding the other characteristic constant. The parsimony of having only 8 of 25 cells with nonzero weights adds statistical power to subsequent tests, although modest expansions of the number of nonzero cells to more interior cells does not alter our conclusions. We also note that the models employing pseudofactors lack a market factor. This is because the market factor rarely plays a significant alpha-influencing role once there are

FIGURE 1
Construction of Equal-Weighted Long–Short (Pseudo) FF Factors

Figure 1 illustrates the weighting scheme used to construct pseudo factors from sets of 5×5 basis portfolios. The 8 symmetric cells with weights +1 and −1 are in border rows and columns, and they indicate extreme quintiles of size and one other attribute. Graph A illustrates the weighting scheme for pseudo small minus big (SMB) stocks where, from top to bottom, the rows represent increasing size and, from left to right, columns represent increasing book-to-market. Graph B illustrates the weighting scheme for pseudo Fama–French (FF) factors high minus low (HML), robust minus weak (RMW), conservative minus aggressive (CMA), or winners minus losers (WML) where, from left to right, the columns represent increasing book-to-market, profitability, investments, or past returns respectively. As before, the rows, from top to bottom, represent increasing size.

Graph A. Pseudo SMB

Size	Low				High
Small	+1	+1		+1	+1
Big	–1	–1		–1	–1

Graph B. Pseudo HML/RMW/CMA/WML

Size	Low				High
Small	–1				+1
	–1				+1
	–1				+1
Big	–1				+1

factors for both size and book-to-market. As we see later, the market factor omission is usually innocuous.

Table 2 indicates that the returns of the pseudofactors are nearly proportional to those of the corresponding traditional factors studied. (The scale differs between the zero-cost pseudo and traditional factors, so means and standard deviations are not separately comparable.) The middle column of Panel A reports factor portfolio Sharpe ratios, which are the ratios of (zero-cost) portfolio means to their standard deviations. The Sharpe ratios of the factors in the 3-, 4-, and 5-factor models and those of the corresponding pseudo factors constructed from sets of 2, 3, and 4 groups of 5×5 portfolios, respectively, are similar. Moreover, the correlations between the factors in these models and their associated pseudo factors are generally above 0.9 and often above 0.95. Thus, the returns of our pseudo factors are largely scaled versions of their traditional counterparts.

TABLE 2

Descriptive Statistics

Table 2 presents descriptive statistics on various traditional factors and traditional pseudo factor portfolios. The traditional factors are the 5 Fama and French (2015) factors (market portfolio (MKT), small minus big stocks (SMB), high minus low book-to-market stocks (HML), robust minus weak operating profitability (RMW), and conservative minus aggressive investments (CMA)) and the Carhart (2014) factor (winners minus losers portfolio (WML)). The traditional pseudo factors are analogous zero-cost portfolios constructed from groups of 5×5 basis portfolios using the weighting scheme illustrated in Figure 1 (scaled by the number of nonzero weights). Panel A presents the means, standard deviations, Sharpe ratios, and minimum and maximum realizations for these (zero-cost) factor portfolios. Panel B presents their correlation matrix. The sample period is from July 1973 to Dec. 2014.

Panel A. Summary Statistics

Variable	Mean	Std. Dev.	Sharpe Ratio	Min.	Max.
SMB	0.2578	3.1106	0.0829	−16.4000	22.0200
Pseudo SMB	0.1299	2.2451	0.0579	−10.2850	17.1462
HML	0.3601	2.9981	0.1201	−12.6100	13.8800
Pseudo HML	0.2233	1.8055	0.1237	−10.5287	7.6475
WML	0.6403	4.4514	0.1438	−34.7200	18.3900
Pseudo WML	0.4541	2.6105	0.1739	−20.2850	11.4400
RMW	0.2793	2.2621	0.1235	−17.6000	12.2400
Pseudo RMW	0.1677	1.2737	0.1316	−9.3025	6.7488
CMA	0.3697	1.9554	0.1891	−6.7600	8.9300
Pseudo CMA	0.2219	1.1026	0.2013	−3.9475	−4.1275
MKT	0.5707	4.6214	0.1235	−23.2400	16.1000

Panel B. Correlation Matrix

	SMB	Pseudo SMB	HML	Pseudo HML	WML	Pseudo WML	RMW	Pseudo RMW	CMA	Pseudo CMA
Pseudo SMB	0.9545									
HML	−0.2258	−0.2551								
Pseudo HML	−0.3087	−0.3596	0.9652							
WML	0.007	0.0117	−0.1548	−0.1492						
Pseudo WML	0.0129	0.0134	−0.1429	−0.1376	0.9798					
RMW	−0.4543	−0.5078	0.1752	0.2607	0.0768	0.0971				
Pseudo RMW	−0.3529	−0.417	0.237	0.3048	0.0745	0.101	0.9479			
CMA	−0.1288	−0.127	0.6988	0.6754	0.0283	0.0466	−0.0086	0.0239		
Pseudo CMA	−0.1360	−0.1373	0.7067	0.6965	−0.0605	−0.0421	−0.0085	0.0132	0.9279	
MKT	0.2709	0.2478	−0.3213	−0.3197	−0.1288	−0.1672	−0.2269	−0.1994	−0.4037	−0.3144

Comparing Figure 2 with Figure 1 illustrates a less restrictive weighting scheme of the 8 relevant cells of the 5×5 arrays' basis portfolios. We now choose weights for nonzero cells to generate a singled priced factor that is judged based on its ability to price assets (i.e., generate zero alphas). The weights are simply those that maximize the Sharpe ratio of the excess returns of the 8-, 16-, 24-, or 32-cell combination. Thus, the optimization problem is over the 8 nonzero weights for a comparison with the 3-factor model, over 16 for the 4-factor model, and over 24 for the 5-factor model. Because the new Fama–French (2015) 5-factor model lacks a momentum factor, we create our own 6-factor model by tacking WML on to the Fama–French 5-factor model. The corresponding unconstrained portfolio weights 32 different cells among four 5×5 sorted portfolios.

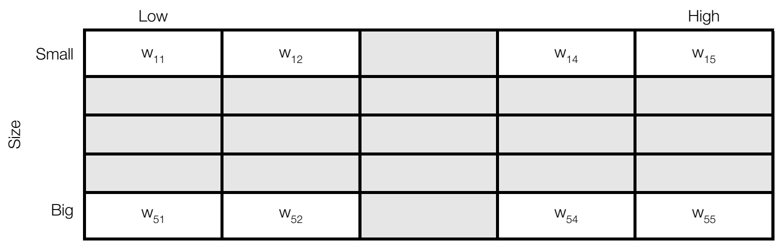
The mean–variance optimization of the cells used in our factor construction generates a maximum Sharpe ratio that is influenced by sampling error. The weighting scheme described previously generates a portfolio with the largest ratio of sample mean to sample standard deviation. This compromises in-sample pricing tests because of the correlation between sample means and sample standard deviations across assets, here, between the MVE factor portfolio and the basis portfolios used for testing. To address this issue, our mean–variance optimized factor portfolio does all of its optimization out of sample so that there is no correlation between the weights and returns used for testing the pricing efficacy of the proposed factor. The procedure is a jackknife: For each test month, we generate mean–variance optimal weights for nonzero cells in Figure 2 that maximize the Sharpe ratio of all observations in the time series that exclude that month. Thus, each month has a different factor that weights the basis portfolios differently, although the differences across months are small with our large sample.

Table 3 compares the jackknifed mean–variance optimized portfolios (referred to as the jackknife MVE(8), MVE(16), MVE(24), and MVE(32), respectively) with their traditional counterparts. Panel A reports alphas of the jackknifed MVE portfolios' returns from regressions against the returns of each of their corresponding traditional benchmarks.[8] The standard OLS t-statistics for the alphas

FIGURE 2

Construction of Optimally Weighted (MVE) Benchmarks

In Figure 2, we construct mean–variance efficient (MVE) benchmarks by choosing nonzero weights for the 8 nonshaded cells from 5×5 arrays of basis portfolios. From top to bottom, the rows represent increasing size and from left to right, the columns represent increasing book-to-market, profitability, investments, or past returns.



[8]The scale of these alphas (but not their t-statistics) depend on the weights used to construct our (zero-investment) jackknifed MVE benchmarks. These weights are given by $\hat{w}_t = \hat{\Sigma}_t^{-1}\hat{\mu}_t$, where $\hat{\Sigma}_t$

TABLE 3

Do Traditional Factors Price Jackknife Benchmarks?

Table 3 compares traditional factors with their associated jackknife benchmarks. The jackknife benchmarks are constructed from groups of 5×5 basis portfolios by optimally weighting the 8 extreme quintiles as illustrated in Figure 2. Panel A reports alphas (α) from regressing the excess returns from jackknife benchmarks on their associated traditional factors. Ordinary least squares (OLS) t-statistics and Newey–West (1987) (NW) corrected t-statistics with 9 lags are reported in parentheses. MVE stands for mean–variance efficient. Panel B reports the maximum Sharpe ratio obtained by optimally weighting a set of portfolios both in and out of sample using our jackknife procedure. The sets of optimized portfolios include sets of traditional factors as well as sets of basis portfolios portrayed in Figure 2. The FF3 column reports results for the Fama–French (1993) 3 factors; Carhart for the FF3 factors augmented with the Carhart (1997) momentum factor; FF5 for the Fama–French (2015) factors; and FF6 for FF5 factors augmented with the momentum factor. *** indicates statistical significance at the 1% level. RHS and LHS refer to right-hand side and left-hand side of the regression equation, respectively. The sample period is from July 1973 to Dec. 2014.

Panel A. Alphas of Jackknife MVE Benchmarks

Traditional Factors (RHS): Jackknife MVE Benchmark (LHS):	FF3 MVE(8)	Carhart MVE(16)	FF5 MVE(24)	FF6 MVE(32)
α_{MVE}	0.1066***	0.2055***	0.1521***	0.3249***
t-stat. (OLS)	(6.4131)	(8.3023)	(6.1727)	(9.2828)
t-stat. (NW)	(5.6221)	(8.1171)	(5.1134)	(7.9960)

Panel B. Maximum Sharpe Ratios

Sharpe Ratio	FF3	Carhart	FF5	FF6
Traditional factor benchmark	0.2234	0.2998	0.3774	0.4041
Jackknife traditional factor benchmark	0.1907	0.2609	0.3410	0.3595
	MVE(8)	MVE(16)	MVE(24)	MVE(32)
In-sample MVE benchmark	0.4223	0.5900	0.5847	0.7556
Jackknife MVE benchmark	0.3700	0.4908	0.4593	0.5926

along with Newey–West (1987) corrected t-statistics with 9 lags are reported in parentheses. All of the alphas in Panel A are highly significant, with the more conservative Newey–West t-statistics ranging from 5.11 (MVE(24) on the Fama–French (2015) 5-factor model) to 8.12 (the Carhart (1997) 4-factor model). These are impressive t-statistics in light of the fact that the MVE portfolio weights, which change each month, are estimated out of sample. Moreover, the alphas, which range from 11 to 32 basis points per month, are the same as those obtained from a single-factor ex post efficient combination of the traditional factors in sample. Hence, despite the out-of-sample handicaps placed on the MVE portfolio, the ability to relax the weights on the basis portfolios, even out-of-sample, has far outweighed the disadvantage of obtaining the MVE cell weights out of sample.

Panel B of Table 3, which reports the Sharpe ratios for the ex post MVE combination of the traditional factor portfolios and their corresponding MVE counterparts (both in sample and when jackknifed before optimizing), is highly consistent with Panel A. In all 4 cases, the Sharpe ratio of the jackknife MVE portfolio (bottom row of Panel B) is significantly larger than that of the in-sample optimal mix of the traditional factor portfolios (top row of Panel B). As a matter of mathematics, the larger Sharpe ratio for the jackknife MVE factor implies that it will generate positive alphas when regressed on the traditional factors.

Panel B of Table 3 also shows that the Sharpe ratios of the in-sample versions of efficient combinations of the traditional factors are about 10%–15% larger than the Sharpe ratios of jackknife-optimized versions of the traditional factors; for

and $\hat{\mu}_t$ are the basis portfolio covariance matrix and mean vector for excess returns (above the riskless rate) estimated using our jackknife procedure.

the 4 MVE portfolios in the bottom two rows, the Sharpe ratios of the in-sample optimized combination of the portfolios are about 15%–30% larger than Sharpe ratios of the jackknife MVE portfolios. Thus, inferences about the inefficiency of traditional factor portfolios are conservative when based on comparisons of in-sample traditional factor portfolios with jackknife MVE portfolios. The "apples-to-apples" comparison of the jackknife-optimized portfolios in the second and fourth rows shows that the weight restrictions of the second row can reduce Sharpe ratios by almost 50%.

To assess the ability to price assets more generally, Table 4 studies the pricing of groups of 5 × 5 basis portfolios that are associated with the various factor models. Panels A–D correspond to tests of the 3-, 4-, 5-, and 6-factor models, respectively. This is an extension of the GRS tests in Table 1 to a less coarse set of basis portfolios. For comparison purposes, the more comprehensive tests here study the pricing ability not only of the traditional factor portfolios, but also of the traditional factor portfolios stripped of a market factor, the pseudo-factor portfolios (which the basis portfolios directly construct), and the jackknife MVE portfolios.

TABLE 4

Do Factors Price Their 5 x 5 Basis Portfolios? Evidence from the GRS Test

Table 4 reports the F-statistic of the Gibbons, Ross, and Shanken (1989) (GRS) test for the mean–variance efficiency (MVE) of various factor models relative to their associated groups of 5 × 5 basis portfolios. The p-values are reported in square brackets. Panels A–D correspond to tests of groups of 5 × 5 basis portfolios associated with the Fama–French (1993) 3-factor model, the Carhart (1997) 4-factor model, the Fama and French (2015) 5-factor model, and a 6-factor model where the Fama–French (2015) factors are augmented with a momentum factor, respectively. The first 3 columns report statistics for the traditional factor models (W/ MKT), traditional factor models without the market portfolio (W/O MKT), and their traditional pseudo-factor models (Pseudo), respectively. The last column reports statistics for jackknife MVE benchmarks denoted as "Jackknife MVE(N)," where N represents the number of basis portfolios used to construct the jackknife benchmark. MKT is market portfolio, HML is high minus low book-to-market stocks, SMB is small minus big stocks, WML is winners minus losers portfolios, RMW is robust minus weak operating profitability, and CMA is conservative minus aggressive investments. The GRS statistic is calculated after taking into account the rank of the residual covariance matrix. * and *** indicate statistical significance at the 10% and 1% levels, respectively. The sample period is from July 1973 to Dec. 2014.

Panel A. 25 Basis Portfolios: Size and Book-to-Market

	SMB, HML			
	W/ MKT	W/O MKT	Pseudo	Jackknife MVE(8)
GRS	3.4282***	3.8902***	3.8164***	1.2569
p-value	[0.0000]	[0.0000]	[0.0000]	[0.2162]

Panel B. 50 Basis Portfolios: Size, Book-to-Market, and Momentum

	SMB, HML, WML			
	W/ MKT	W/O MKT	Pseudo	Jackknife MVE(16)
GRS	3.3927***	3.7526***	4.1999***	1.4180*
p-value	[0.0000]	[0.0000]	[0.0000]	[0.0628]

Panel C. 75 Basis Portfolios: Size, Book-to-Market, Profitability, and Investment

	SMB, HML, RMW, CMA			
	W/ MKT	W/O MKT	Pseudo	Jackknife MVE(24)
GRS	2.0195***	2.3399***	1.9361***	0.8943
p-value	[0.0000]	[0.0000]	[0.0000]	[0.6808]

Panel D. 100 Basis Portfolios: Size, Book-to-Market, Profitability, Investment, and Momentum

	SMB, HML, RMW, CMA, WML			
	W/ MKT	W/O MKT	Pseudo	Jackknife MVE(32)
GRS	2.7474***	3.1243***	2.9701***	1.2198
p-value	[0.0000]	[0.0000]	[0.0000]	[0.1260]

https://doi.org/10.1017/S0022109018000376 Published online by Cambridge University Press

Panel A of Table 4 shows that the 25 basis portfolios sorted into quintiles by size and book-to-market have jointly significant alphas whether benchmarked against the traditional Fama–French (1993) 3-factor model, the 3-factor model without the market, or the pseudo 3-factor model constructed from 8 of the extreme basis portfolios. By comparison, the corresponding GRS F-statistic of the MVE benchmark constructed from 8 basis portfolios with the jackknife procedure has a p-value of 0.22. The relatively high p-value indicates that the 25 jackknife MVE-benchmarked alphas of the 5×5 size and book-to-market portfolios cannot be statistically distinguished from 0.

The story in Panels B–D of Table 4 is similar. The 3 versions of the traditional factor models fail miserably at pricing their corresponding basis portfolios. By contrast, 2 of the 3 remaining jackknife MVE factor portfolios have insignificant GRS test statistics. The latter 2 have p-values of 0.68 (from alphas of 75 basis portfolios benchmarked against MVE(24)) and 0.13 (from alphas of 100 basis portfolios benchmarked against MVE(32)). The GRS F-statistic for MVE(16) is significant at the 10% level (from alphas of 75 basis portfolios) but, with a p-value of 0.06, is still orders of magnitude less significant than the F-statistics of the traditional benchmarks to which it is compared. One interpretation of the higher p-values with the MVE jackknife benchmarks is that the MVE benchmarks generate alphas estimated with less precision. Another is that lifting the weight restriction of traditional factor models improves their ability to price the basis portfolios used to construct them. Barillas and Shanken (2017), (2018) argue that the only relevant test of which of 2 factor models is superior comes from analyzing the alphas of one set of factors against the other. As we see in Panel A of Table 3, the jackknife benchmarks tend to have significant alphas when regressed on the traditional factors. However, the traditional factors (individually and jointly) all have insignificant alphas when regressed on the jackknife MVE counterpart (not shown in a table for brevity). Hence, the evidence in Table 4 is consistent with the weight restrictions of traditional models as the source of their nonzero alphas.

To address the possibility that the tests in Table 4 might be influenced by the strong assumptions of the GRS test, including i.i.d. normally distributed returns, Table 5 compares the Sharpe ratios of 2 portfolios using the LW test. The format is similar to Table 4, except that we report the difference in Sharpe ratios between i) the optimal combination of the test portfolios listed in the panel title, estimated with the same jackknife procedure used for the MVE portfolios described earlier, and ii) the jackknifed optimal combination of the set of benchmark portfolios in the column heading. The LW test requires that both Sharpe ratios being compared derive from portfolios that are prespecified; neither portfolio's weights can be endogenously determined by return data used to compute the Sharpe ratios. Optimizing both portfolios with data from the jackknife achieves this goal.

Table 5 obtains its asterisked significance levels and associated p-values for the Sharpe ratio differences from the LW circular block bootstrap algorithm.[9]

[9]Lahiri (1991) shows that the bootstrap estimate of the sampling distribution is more accurate than the normal approximation, provided it is centered around the bootstrap mean (not around the sample mean as is customary). Politis and Romano (1992) introduce a variant of this block-resampling bootstrap: The circular block-resampling bootstrap, which has the additional advantage of being automatically centered around the sample mean. "Circular" describes the resampling procedure, which amounts to "wrapping" the data around in a circle before blocking them.

TABLE 5

Do Factors Price Their 5 x 5 Basis Portfolios? Evidence from Differences in Sharpe Ratios

Table 5 reports differences in maximum Sharpe ratios obtained by jackknife-optimized test portfolios constructed from factors plus their associated groups of 5 × 5 basis portfolios and those obtained by optimally combining only the factors using our jackknife procedure. The *p*-values of the Sharpe ratio differences, reported in square brackets, use the Ledoit and Wolf (2008) (LW) robust algorithm with 5,000 simulated data sets. When the difference in maximum Sharpe ratios is negative, *p*-values are reported as [N/A]. Panels A–D correspond to tests of groups of 5 × 5 basis portfolios associated with the Fama–French (1993) 3-factor model, the Carhart (1997) 4-factor model, the Fama–French (2015) 5-factor model, and a 6-factor model where the Fama–French (2015) factors are augmented with a momentum factor, respectively. The first 3 columns report statistics for the traditional factor models (W/ MKT), traditional factor models without the market portfolio (W/O MKT), and their traditional pseudo-factor models (Pseudo), respectively. The last column reports statistics for jackknife mean–variance efficient benchmarks, denoted as "Jackknife MVE(*N*)" where *N* represents the number of nonzero weights used to construct the jackknife benchmark factor (as illustrated in Figure 2). MKT is market portfolio, HML is high minus low book-to-market stocks, SMB is small minus big stocks, WML is winners minus losers portfolios, RMW is robust minus weak operating profitability, and CMA is conservative minus aggressive investments. *, **, and *** indicate a positive and statistically significant difference at the 10%, 5%, and 1% levels, respectively. The sample period is from July 1973 to Dec. 2014.

Panel A. 25 Basis Portfolios: Size and Book-to-Market

	SMB, HML			
	W/ MKT	W/O MKT	Pseudo	Jackknife MVE(8)
Difference in Sharpe ratio	0.1569**	0.2076***	0.2074***	−0.0222
p-value	[0.0118]	[0.0006]	[0.0008]	[N/A]

Panel B. 50 Basis Portfolios: Size, Book-to-Market, and Momentum

	SMB, HML, WML			
	W/ MKT	W/O MKT	Pseudo	Jackknife MVE(16)
Difference in Sharpe ratio	0.1919***	0.2503***	0.2276***	−0.0381
p-value	[0.0008]	[0.0002]	[0.0002]	[N/A]

Panel C. 75 Basis Portfolio: Size, Book-to-Market, Profitability, and Investment

	SMB, HML, RMW, CMA			
	W/ MKT	W/O MKT	Pseudo	Jackknife MVE(24)
Difference in Sharpe ratio	0.0220	0.0983*	0.1034*	−0.0960
p-value	[0.6683]	[0.0852]	[0.0566]	[N/A]

Panel D. 100 Basis Portfolios: Size, Book-to-Market, Profitability, Investment, and Momentum

	SMB, HML, RMW, CMA, WML			
	W/ MKT	W/O MKT	Pseudo	Jackknife MVE(32)
Difference in Sharpe ratio	0.1739***	0.1733***	0.2436***	−0.0592
p-value	[0.0072]	[0.0012]	[0.0008]	[N/A]

This p-value assesses whether the difference in Sharpe ratios significantly differs from 0. The test accounts for the possibility that observed Sharpe ratios come from returns generated by changing population means, variances, and covariances, or other sources of non-normal distributions. The LW algorithm simulates 5,000 data sets by the circular block bootstrap and takes critical values as the empirical quantiles of the simulated data sets.

The negative numbers in the rightmost column of Table 5 indicate that each jackknife MVE benchmark has a larger Sharpe ratio than the jackknife-optimized combination of all of its basis portfolios, a phenomenon that cannot happen if the optimization is in sample.[10] The likely cause is estimation noise: The jackknifed basis portfolio weights are far from optimal for the population. The estimation noise leads to relatively low Sharpe ratios when tested on data not used for the

[10]The p-values are not reported when the Sharpe ratio difference is negative.

portfolio weight optimization. Sensible restrictions, like the jackknife MVE zero weight on some of the interior basis portfolios, thus improve the Sharpe ratio. What is surprising here is that the same principle does not apply for any of the traditional factor benchmarks, which have additional restrictions but lower Sharpe ratios than the jackknife-optimized basis portfolios. Thus, it appears to be the arbitrary equal magnitude weight restriction on the basis portfolios imposed in traditional factor portfolio construction that makes the Sharpe ratios of traditional factor portfolios so low.

We now investigate whether the asset pricing deficiencies generated by the weight restrictions of traditional factor construction carry over to the pricing of other portfolios. In particular, we study the alpha differences of portfolio pairings based on the decile extremes of the 10 anomalies listed in the Kenneth French data library: cash-flow-to-price, dividend yield, earnings-to-price, variance, residual variance, market beta, short-term reversal, long-term reversal, accruals, and net share issuance. For each anomaly, Table 6 reports the differences in the alphas (along with t-statistics) of equal-weighted portfolios of stocks ranked in deciles 1 (lowest) and 10 (highest) based on the anomaly-generating characteristic. The bottom of the table reports the GRS F-statistic testing whether all 10 alpha differences are zero. In almost every instance, the t-statistic for the traditional factor benchmark is larger than the t-statistic for its jackknife MVE counterpart. Moreover, although most of the t-statistics are insignificant at the 10% level on both the left (traditional) and right (jackknife MVE) halves of the table, the 4 jackknife MVE portfolios are especially adept at pricing the anomalies. Each of the jackknife MVE portfolios prices at least 8 of the anomalies and the jackknife MVE(32) prices all 10 anomalies, both individually and jointly. By contrast, many of the anomaly portfolios have extreme alphas with traditional factor benchmarks and none of the traditional factor benchmarks comes close to jointly pricing all anomaly portfolios.

Table 7 performs the LW test from Table 5 but focuses on a different test portfolio set. The set now includes both the 20 extreme decile anomaly portfolios and the 8, 16, 24, or 32 basis portfolios with nonzero weight in the MVE factor construction. We exclude 17 of the 25 basis portfolios with which the Fama and French (1993) 3-factor model and MVE(8) are associated, 34 of the 50 with which the Carhart (1997) 4-factor model and MVE(16) are associated, and so on, before jackknife optimizing. Excluding these "interior" basis portfolios reduces the effect of sampling error on the Sharpe ratio maximizing weights, making it harder to price the jackknife-optimized test portfolio (as noted in the Table 5 discussion). Table 7 shows that adding the anomaly portfolios and excluding the interior basis portfolio from the test assets reverses the sign of the Sharpe ratio difference between the jackknife MVE portfolio and the full set of test portfolios. Recall that these differences are negative in Table 5, but are now positive. Table 7 also differs from earlier tables by offering a more complete set of tests. It effectively compares the Sharpe ratios of each jackknife-optimized benchmark with those of test portfolios derived from all 4 sets of basis assets (and the 20 anomaly portfolios).

The p-values of the cells along the diagonal on the left of the two 4×4 arrays in Table 7 show that 3 of the 4 traditional benchmarks have virtually 0 p-values, with the remaining p-value below 2%. By contrast, among the p-values of the

TABLE 6

Do Factors Price Anomaly Portfolios? Evidence from Alphas and the GRS Test

Table 6 reports alphas (α) and corresponding test statistics for 10 long–short anomaly portfolios. These portfolios are long and short the two most extreme decile portfolios constructed using univariate sorts on cash-flow-to-price, dividend yield, earnings-to-price, variance, residual variance, market beta, short-term reversal, long-term reversal, accruals, and net share issues. Newey–West (1987) corrected t-statistics with 9 lags are reported in parentheses. The 4 columns in the left matrix correspond to statistics for the Fama–French (1993) 3-factor model (FF3), the Carhart (1997) 4-factor model (Carhart), the Fama–French (2015) 5-factor model (FF5), and a 6-factor model where the Fama–French (2015) factors are augmented with a momentum factor (FF6). The 4 columns in the right matrix correspond to statistics for their corresponding jackknife mean–variance efficiency (MVE) benchmarks: MVE(N), where N denotes the number of basis portfolios with nonzero weights in the jackknife MVE benchmark. The GRS row corresponds to the F-statistic of Gibbons, Ross, and Shanken (1989). Its corresponding p-values are reported below in square brackets. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from July 1973 to Dec. 2014.

	In Sample				Jackknife			
Anomaly Portfolios	FF3	Carhart	FF5	FF6	MVE(8)	MVE(16)	MVE(24)	MVE(32)
Cash-Flow-to-Price								
α	−0.1227	−0.0440	−0.1414	−0.0781	−0.1922	−0.2059	−0.3293	−0.2775
t-stat.	(−0.9440)	(−0.3535)	(−1.1425)	(−0.6320)	(−0.8917)	(−0.9773)	(−1.4680)	(−1.0998)
Dividend Yield								
α	0.0536	−0.0731	−0.0494	−0.1364	0.0662	−0.1076	0.0913	−0.1115
t-stat.	(0.2698)	(−0.3711)	(−0.2530)	(−0.7345)	(0.2186)	(−0.3599)	(0.2913)	(−0.3149)
Earnings-to-Price								
α	−0.0551	−0.0061	−0.0446	−0.0053	−0.0728	−0.0789	−0.2027	−0.1594
t-stat.	(−0.3626)	(−0.0430)	(−0.3492)	(−0.0432)	(−0.3400)	(−0.3617)	(−0.8953)	(−0.6146)
Variance								
α	1.0094***	0.7325***	0.3481	0.2008	0.5228	0.1628	0.3965	0.0463
t-stat.	(3.8517)	(2.8248)	(1.5187)	(0.8082)	(1.2084)	(0.3150)	(0.9056)	(0.0854)
Residual Variance								
α	1.1335***	0.8226***	0.4969***	0.3208*	0.5870	0.2632	0.4692	0.1603
t-stat.	(5.1179)	(3.7984)	(2.5997)	(1.6609)	(1.6020)	(0.5543)	(1.2749)	(0.3365)
Market Beta								
α	0.3392*	0.1228	0.0515	−0.0819	0.1326	−0.0931	−0.0782	−0.2845
t-stat.	(1.8795)	(0.6983)	(0.2809)	(−0.4430)	(0.3889)	(−0.2194)	(−0.2384)	(−0.6526)
Short-Term Reversal								
α	0.0257	0.3737	0.1307	0.3748	0.0599	0.2557	0.1039	0.1466
t-stat.	(0.1066)	(1.5245)	(0.4486)	(1.4717)	(0.2110)	(0.6737)	(0.3287)	(0.3529)
Long-Term Reversal								
α	−0.1341	0.0059	−0.1638	−0.0505	−0.1388	0.0417	−0.0737	0.1410
t-stat.	(−0.6330)	(0.0276)	(−0.8178)	(−0.2502)	(−0.5528)	(0.1374)	(−0.2605)	(0.3806)
Accruals								
α	0.4020***	0.3153***	0.4361***	0.3712***	0.4636***	0.3676***	0.3830***	0.2431
t-stat.	(3.2437)	(2.4938)	(2.9704)	(2.5151)	(3.3447)	(2.6596)	(2.4760)	(1.5806)
Net Share Issues								
α	0.5660***	0.4949***	0.2319***	0.2098	0.4220***	0.3700**	0.3343*	0.2688
t-stat.	(3.7517)	(3.4142)	(1.6825)	(1.5538)	(2.7195)	(2.0878)	(1.9278)	(1.3420)
All 10 Anomaly Portfolios								
GRS	6.2099***	4.5342***	3.0857***	2.4135***	2.4847***	2.0034**	2.2566**	1.5487
p-value	[0.0000]	[0.0000]	[0.0008]	[0.0083]	[0.0065]	[0.0313]	[0.0139]	[0.1191]

Sharpe ratio differences along the diagonal of the array on the right, we see that only the top left p-value is significant at the 5% level. Thus, when 20 anomaly portfolios are added to the mix, the MVE alternative to the traditional 3-factor model cannot price both the "restricted" basis portfolios formed from extreme size and book-to-market along with the 20 anomalies. There is 10% significance but not 5% significance for the 0.0625 difference between the Sharpe ratios of the MVE alternative to the Carhart (1997) 4-factor model and the optimal combination of the 16 Carhart basis portfolios and the 20 anomaly portfolios. Despite the MVE(16) benchmark's "on-the-fence" ability to price these 36 portfolios, it is still far better than the traditional alternative, which generates a Sharpe ratio difference of 0.2866, more than 4 times larger than its MVE counterpart. Even more impressive are the MVE alternatives to the Fama–French (2015) 5- and 6-factor

https://doi.org/10.1017/S0022109018000376 Published online by Cambridge University Press

TABLE 7

Do Factors Price Anomaly Portfolios? Evidence from Differences in Sharpe Ratios

Table 7 reports the differences between maximum Sharpe ratios obtained by a jackknifed optimal test portfolio constructed from 20 anomaly portfolios plus various sets of basis portfolios and Sharpe ratios obtained by optimally combining only the factors. The *p*-values of the Sharpe ratio difference, reported in square brackets, use the Ledoit and Wolf (2008) robust algorithm with 5,000 simulated data sets. The 20 anomaly portfolios include the two most extreme decile portfolios constructed using univariate sorts on 10 characteristics: cash-flow-to-price, dividend yield, earnings-to-price, variance, residual variance, market beta, short-term reversal, long-term reversal, accruals, and net share issues. In addition to these anomaly portfolios and factors, we use 8 extreme cells in various groups of 5 × 5 basis portfolios to construct jackknifed optimal test portfolios. The 4 panels correspond to tests in which the basis portfolios are the 8, 16, 24, or 32 basis portfolios associated with the Fama–French (1993) 3-factor model (FF3), the Carhart (1997) 4-factor model (Carhart), the Fama–French (2015) 5-factor model (FF5), and a 6-factor model where the Fama–French (2015) factors are augmented with a momentum factor (FF6). The 4 columns in the left matrix correspond to statistics for the traditional 3-, 4-, 5-, and 6-factor models. The 4 columns in the right matrix correspond to statistics for their corresponding jackknife mean–variance efficiency (MVE) benchmarks, denoted "MVE(*N*)," where *N* denotes the number of basis portfolios with nonzero weights in the jackknife benchmark. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from July 1973 to Dec. 2014.

	Jackknife				Jackknife			
Test Portfolios	FF3	Carhart	FF5	FF6	MVE(8)	MVE(16)	MVE(24)	MVE(32)
8 Basis Portfolios and Anomalies								
Difference in Sharpe ratio	0.2574***	0.1809***	0.1098**	0.0842*	0.0848**	0.0652*	0.0233	0.0037
p-value	[0.0002]	[0.0012]	[0.0234]	[0.0500]	[0.0466]	[0.0928]	[0.4215]	[0.8816]
16 Basis Portfolios and Anomalies								
Difference in Sharpe ratio	0.3557***	0.2866***	0.2030***	0.1861***	0.1860***	0.0652*	0.1370***	0.0037
p-value	[0.0000]	[0.0002]	[0.0004]	[0.0009]	[0.0014]	[0.0928]	[0.0008]	[0.8816]
24 Basis Portfolios and Anomalies								
Difference in Sharpe ratio	0.2669***	0.2104***	0.1327**	0.1105**	0.1126**	0.1055**	0.0233	0.0037
p-value	[0.0002]	[0.0008]	[0.0188]	[0.0440]	[0.0266]	[0.0358]	[0.4215]	[0.8816]
32 Basis Portfolios and Anomalies								
Difference in Sharpe ratio	0.3993***	0.3272***	0.2412***	0.2212***	0.2263***	0.1055**	0.1370***	0.0037
p-value	[0.0002]	[0.0002]	[0.0002]	[0.0003]	[0.0002]	[0.0358]	[0.0008]	[0.8816]

models. Both of these models price their respective basis assets plus the anomaly portfolios, as the Sharpe ratio differences between these 2 MVE benchmarks and their respective test assets are close to zero.

These significant Sharpe ratio differences also reveal an interesting fact: Neither the traditional factor portfolios nor their MVE alternatives can price basis portfolios that are not used to form the MVE portfolio. The jackknife MVE benchmarks still represent progress in that the Sharpe ratio difference between the 4 test portfolios and the MVE portfolios are smaller than those with traditional factor models. Moreover, looking above the diagonals for the 2 arrays indicates that only the MVE alternatives price subgroups of their basis assets plus anomaly portfolios. The one highly significant Sharpe ratio difference above the diagonal actually buttresses our argument. The failing of the MVE alternative to the 5-factor model (MVE(24)) to price the 16 basis portfolios formed from size, book-to-market, and momentum (plus the 20 anomaly portfolios) indicates only that the 24 portfolios used to form MVE(24) are missing the extreme momentum portfolios used in the Carhart (1997) model, MVE(16), and MVE(32).

Although we have no similar table for the pseudo factors or the traditional factors without the market, the conclusions from such a table would be similar to those reported in the left array of Table 5. Thus, once again, it is the unwarranted restriction to equal weights on basis portfolios that hinders the ability of traditional factor models to adequately price assets.

https://doi.org/10.1017/S0022109018000376 Published online by Cambridge University Press

III. Conclusion

When factors are constructed by combining a set of characteristic-sorted portfolios, prespecified simple weights are likely to produce nonzero estimates of alpha. This is true theoretically even when the underlying basis portfolios span the MVE portfolio because parsimony prevents the rank-deficient factor portfolios themselves from spanning that efficient portfolio.

We propose an alternative to the traditional factor construction method that corrects for spurious nonzero alphas, and we test the pricing acumen of these alternative factors against those constructed with the traditional approach. The alternative is to use the ex post efficient combination of a small subset of extreme basis portfolios. Tested with an out-of-sample jackknife approach to their construction, the MVE 1-factor alternatives we propose vastly outperform traditional factor models as predictors of the cross section of expected returns.

The MVE alternative prices not only its basis assets, but often anomaly portfolios as well. The MVE alternative also often prices anomaly portfolios plus the basis portfolios used to form the MVE portfolio. The jackknife MVE alternative to traditional factor models thus represents a large improvement over traditional factor portfolios. In the only fair comparison, where the optimal combinations of traditional factor models are jackknifed as well, even the Sharpe ratio of the MVE(8) portfolio (the proposed alternative to the traditional 3-factor model) is about the same as the Sharpe ratio of the optimal combination of the factor portfolios from the traditional 6-factor model, and it exceeds the Sharpe ratios of the optimal combinations of all the other sets of traditional factor portfolios. Thus, even when basis assets consists of only size and book-to-market portfolios, the mere relaxation of the equal weight restriction improves the pricing of assets associated with other characteristics besides size and value.

The jackknife MVE alternative studied here uses rudimentary estimation techniques. Sample averages and sample covariance matrices are used in the jackknife as inputs to portfolio optimization. However, a large body of statistical research suggests that the out-of-sample performance could be improved with more sophisticated estimation techniques. One useful extension of this article would be improve the estimation of the jackknife covariance matrix with shrinkage models from statistics. A recent paper by Engle, Ledoit, and Wolf (2018) offers a promising (albeit computationally intensive) way of estimating dynamic covariance matrices that works well out of sample. Another useful extension of this article would be to apply our jackknife MVE factors to other arenas such as mutual and hedge fund performance. Both extensions are currently work in progress, but are beyond the scope of this article.

Appendix. Alpha Estimation with Rank-Deficient Factor-Mimicking Portfolios

Without loss of generality, we represent a K-factor securities market with covariances determined by exposures to both a single mean-zero priced factor $\tilde{f}_0$ and $K-1$ mean-zero unpriced factors $f_1, \ldots, f_{K-1}$ that are orthogonal to the priced factor and to each other. To simplify algebra, we scale the factors to have equal variance. We ignore firm-specific risk because the asset pricing tests we discuss are all based on sets of well-diversified assets.

https://doi.org/10.1017/S0022109018000376 Published online by Cambridge University Press

In this case, Ross's (1976) no-arbitrage condition implies that the excess return-generating process for N diversified assets is:

$$\text{(A-1)} \qquad \tilde{\mathbf{R}} = \lambda \mathbf{B}_0 + \mathbf{B}\tilde{\mathbf{f}},$$

where $\tilde{\mathbf{R}}$ is the N-vector of excess returns, λ is the risk premium attached to unit exposure to the priced factor, $\mathbf{B}_0$ is the N-vector of loadings on the priced factor (also the first column of the $N \times K$ factor loading matrix $\mathbf{B}$), and $\tilde{\mathbf{f}}$ is the K-vector of priced and unpriced factors. We use boldface to refer to vectors and matrices.

In the absence of arbitrage, one prices the N assets by identifying K well-diversified assets or K portfolios of those assets with a full rank submatrix of factor loadings (i.e., linearly independent beta vectors). These factor portfolios, when combined with a risk-free asset, span the same space as the factors. However, empirical asset pricing research generally selects a smaller set of J factor-mimicking portfolios to price assets. The necessary and sufficient criterion that allows this smaller subset of portfolios to price assets is that the rows of the $J \times (K-1)$ matrix, $\mathbf{b^u}$, of factor loadings on the $K-1$ unpriced factors be linearly dependent. Any portfolio weight J-vector, $\mathbf{w}$, that makes $\mathbf{w^T b^u} = \mathbf{0^T}$ is an optimal portfolio provided that this portfolio return has a nonzero loading on the priced factor. If $J = K$ and the factor loading matrix, $\mathbf{b}$, of the factor portfolios is of full rank (i.e., no factor portfolios are redundant), a K-vector of portfolio weights can simultaneously set the $K-1$ loadings on the unpriced factors to zero while preserving a nonzero loading on the priced factor. However, if $J < K$, there is no guarantee that the optimal portfolio, and hence the priced factor, is captured in the span of the J factor-mimicking portfolios. Alphas are nonzero whenever the Sharpe-ratio-maximizing combination of the J factor-mimicking portfolios contains any unpriced factor risk. That is, the alphas of assets measured against the J factor-mimicking portfolios will differ from the zero true alphas they actually possess when measured against a K-factor model. The bias in measured alpha from benchmarking against factor-mimicking portfolios that cannot be combined to eliminate unpriced factor risk is easy to see.

Regressing the N well-diversified excess returns on J well-diversified factor portfolios with excess return vector $\tilde{\mathbf{F}}$ and $J \times K$ factor loading matrix $\mathbf{b}$ (with first column $\mathbf{b_0}$) yields,

$$\tilde{\mathbf{R}} = \boldsymbol{\alpha} + \boldsymbol{\beta}^T \tilde{\mathbf{F}} + \tilde{\boldsymbol{\epsilon}},$$

with the $J \times N$ matrix of slope coefficients and the N-vector of alphas given by

$$\boldsymbol{\beta} = (\mathbf{b}\mathbf{b}^\mathbf{T})^{-1}\mathbf{b}\mathbf{B}^\mathbf{T},$$

$$\boldsymbol{\alpha} = \lambda \mathbf{B_0} - \lambda \mathbf{B}\mathbf{b}^\mathbf{T}(\mathbf{b}\mathbf{b}^\mathbf{T})^{-1}\mathbf{b_0}.$$

The slope coefficients are derived from substituting the right side of equation (A-1) for the regression's dependent variable and applying the formula for projections; the intercepts use the slope coefficients and no-arbitrage mean excess returns in the standard regression formula for intercepts.

$\boldsymbol{\alpha}$ is the zero vector only when

$$\text{(A-2)} \qquad \mathbf{B}_0 = \mathbf{B}\mathbf{b}^\mathbf{T}(\mathbf{b}\mathbf{b}^\mathbf{T})^{-1}\mathbf{b_0}.$$

Equation (A-2) holds if $\mathbf{b}$ is of rank J and some J-vector $\mathbf{w}$ can make $\mathbf{w^T b} = \mathbf{e_1^T}$, the first unit vector. This is the case where the J factor portfolios span the optimal portfolio in the K-factor environment. If such a $\mathbf{w}$ exists, it is proportional to $(\mathbf{b}\mathbf{b}^\mathbf{T})^{-1}\mathbf{b_0}$ in the absence of arbitrage, as this is the familiar first-order condition for Sharpe ratio maximization applied to the case of one priced factor. Such a $\mathbf{w}$ always exist when $J = K$ and the factor portfolios have a $\mathbf{b}$ matrix of rank K. In this case of perfect regression fit, $\mathbf{b}^{\mathrm{T}}(\mathbf{b}\mathbf{b}^{\mathrm{T}})^{-1}\mathbf{b}$ is the identity matrix because $(\mathbf{b}\mathbf{b}^{\mathrm{T}})^{-1} = (\mathbf{b}^{\mathrm{T}})^{-1}\mathbf{b}^{-1}$, making

$$\mathbf{B} \quad = \quad [\mathbf{B}\mathbf{b}^{\mathrm{T}}(\mathbf{b}\mathbf{b}^{\mathrm{T}})^{-1}]\mathbf{b}.$$

This condition relates each column of $\mathbf{B}$ on the equation's left side to the corresponding column of the outside-the-bracket $\mathbf{b}$ on the right side, including the leftmost columns $\mathbf{B}_0$ and $\mathbf{b}_0$, thus yielding equation (A-2).

References

Barillas, F., and J. Shanken. "Which Alpha?" *Review of Financial Studies*, 30 (2017), 1316–1338.

Barillas, F., and J. Shanken. "Comparing Asset Pricing Models." *Journal of Finance*, 73 (2018), 715–754.

Basak, G. K.; R. Jagannathan; and T. Ma. "Jackknife Estimator for Tracking Error Variance of Optimal Portfolios." *Management Science*, 55 (2009), 990–1002.

Black, F.; M. Jensen; and M. Scholes. "The Capital Asset Pricing Model: Some Empirical Tests." In *Studies in the Theory of Capital Markets*, M. C. Jensen, ed. New York, NY: Praeger (1972), 79–121.

Carhart, M. M. "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52 (1997), 57–82.

Connor, G. "A Unified Beta Pricing Theory." *Journal of Economic Theory*, 34 (1984), 13–31.

Cremers, M.; A. Petajisto; and E. Zitzewitz. "Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation." *Critical Finance Review*, 2 (2013), 1–48.

Engle, R. F.; O. Ledoit; and M. Wolf. "Large Dynamic Covariance Matrices." *Journal of Business & Economic Statistics*, forthcoming (2018).

Fama, E. F., and K. R. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, 33 (1993), 3–56.

Fama, E. F., and K. R. French. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics*, 116 (2015), 1–22.

Fama, E. F., and J. D. MacBeth. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy*, 81 (1973), 607–636.

Gerakos, J., and J. T. Linnainmaa. "Dissecting Factors." Working Paper, University of Chicago Booth School of Business (2014).

Gibbons, M. R.; S. A. Ross; and J. Shanken. "A Test of the Efficiency of a Given Portfolio." *Econometrica*, 57 (1989), 1121–1152.

Huberman, G., and S. Kandel. "A Size Based Stock Returns Model." Working Paper, University of Chicago (1985).

Kan, R., and D. R. Smith. "The Distribution of the Sample Minimum-Variance Frontier." *Management Science*, 54 (2008), 1364–1380.

Lahiri, S. N. "Second Order Optimality of Stationary Bootstrap." *Statistics and Probability Letters*, 11 (1991), 335–341.

Ledoit, O., and M. Wolf. "Robust Performance Hypothesis Testing with the Sharpe Ratio." *Journal of Empirical Finance*, 15 (2008), 850–859.

Lintner, J. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics*, 47 (1965), 13–37.

Newey, W. K., and K. D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55 (1987), 703–708.

Novy-Marx, R. "The Other Side of Value: The Gross Profitability Premium." *Journal of Financial Economics*, 108 (2013), 1–28.

Politis, D. N., and J. P. Romano. "A Circular Block-Resampling Procedure for Stationary Data." In *Exploring the Limits of Bootstrap*, R. LePage and L. Billard, eds. New York, NY: Wiley (1992), 263–270.

Ross, S. A. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory*, 13 (1976), 341–360.

Sharpe, W. F. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance*, 19 (1964), 425–442.

Welch, I. "The Link between Fama-French Time-Series Tests and Fama-MacBeth Cross-Sectional Tests." Working Paper, University of California, Los Angeles (2008).