Volatility-managed portfolios on French's factors: the replication, the real-time critique, and the 136 months after the sample

Carlo Hofer. November 2026. Code, notebooks and the tables every number comes from: github.com/ochofer/volatility-managed-portfolios, results/202609.

Abstract

Moreira and Muir (2017) divide a factor by last month's realised variance and find that this volatility-managed factor has a positive alpha against the original factor for five of the six factors studied here. Cederburg et al. (2020) find that, for these six, combining the two factors in real time gains little outside momentum. I first replicate both papers on Kenneth French's data, together with Barroso and Santa-Clara's (2015) scaled momentum, which holds a momentum portfolio at a constant volatility. The replications hold except on RMW, the profitability factor, most likely because French has revised it since. I then compare each volatility-managed factor's Sharpe ratio with the original's on Moreira and Muir's sample, with Ledoit and Wolf's (2008) test. The volatility-managed Sharpe ratio is higher by a statistically significant margin for momentum tested on its own, and for no factor once the six tests are read together. Finally, I run the volatility-managed factor and the real-time combination on the 136 months after Moreira and Muir's sample. In those months no Sharpe ratio difference is statistically significant, for the volatility-managed factor against the original or for the real-time combination against the original held at its real-time weight. For the volatility-managed factors, each difference's confidence interval contains the difference measured on Moreira and Muir's sample, so these months cannot show whether the signal did less or more than before.

Terms

Factor: one of the six long-short return series Kenneth French publishes in his data library, i.e., the market, SMB (size), HML (value), momentum, RMW (profitability) and CMA (investment) of Fama and French (2015) and Carhart (1997). As published, it is the **original** factor. **Excess return**: a return minus the one-month Treasury bill rate. Every factor is one. **Sharpe ratio**: the mean excess return over its standard deviation, annualised. **Realised variance**: the variance of a factor computed from its daily returns within one month, each paper with its own definition (Section 2). **Signal**: last month's realised variance of a factor, and nothing else. **Volatility-managed factor**: Moreira and Muir's construction, the factor's return times c over last month's realised variance. **c** is the constant that gives it the original's standard deviation over the whole sample, so it is known only at the sample's end. **Position**: the multiple of the factor held in a month. **Mean-variance weights**: the holdings of an investor who trades a portfolio's mean against its variance, here with a risk aversion of 5. **S1, S2 and S3**: Cederburg et al.'s three strategies. *S1*, the real-time combination, holds the mean-variance weights on the volatility-managed and the original factor estimated each month from the months before it. *S2*, the real-time original, holds the original alone at its real-time weight. *S3* holds the two at the mean-variance weights estimated on the whole sample, which no investor could have known. **WML**: winners minus losers, the return of the top decile of French's ten portfolios sorted on prior returns minus the bottom decile's. **Scaled momentum**: Barroso and Santa-Clara's construction, WML held at 12 per cent a year over its realised volatility of the previous 126 trading days. **Maximum drawdown**: the largest fall from a peak of the compounded return. **Alpha**: the intercept of the regression of the volatility-managed factor on its original, in per cent a year. It measures what the volatility-managed factor adds to a holding of the original. The **appraisal ratio** is the alpha over the residual's standard deviation, annualised. **White and Newey-West standard errors**: White's (1980) allows the residual's variance to change from month to month, and Newey and West's (1987), at 12 lags, also allows a month's residual to be correlated with the 12 before it. **Jobson-Korkie test**: the test of Jobson and Korkie (1981), with Memmel's (2003) correction, that two return series over the same months have the same Sharpe ratio, as Cederburg et al. use it. It assumes normal and independent months. **Kernel**: the weights that turn a series' autocovariances into its long-run covariance, falling to zero at a lag that the bandwidth sets. The Parzen and the QS kernels are two such sets of weights, and prewhitening first filters out each series' autocorrelation, the correlation of its values with those a month earlier. **Ledoit and Wolf's test**: the test of a Sharpe ratio difference used here (Section 2). It resamples the months in blocks of consecutive months, which allows for fat tails, i.e., extreme months more frequent than under the normal distribution, and for volatile months that follow each other. Each resampled history is a **bootstrap sample**, and the **critical value** z* is the multiple of the difference's standard error that gives its interval, in the place of the normal distribution's 1.96. **Interval**: the 95 per cent interval of Ledoit and Wolf's test for the true difference between two Sharpe ratios, annualised. A difference is statistically significant at the 5 per cent level when zero lies outside its interval. The **half-width** is the distance from the difference to either end of its interval. **p-value**: the probability of an estimate at least as far from zero as the one computed, if the true value were zero. **Holm's correction**: for n tests read together, the smallest p-value is multiplied by n, the next by n − 1 and so on, each kept at least as large as the one before and none above one (Holm, 1979). **Size**: the share of samples in which a test rejects when the two true Sharpe ratios are equal. A test of 5 per cent should reject 5 per cent of them. **Student-t shocks**: random values from Student's t distribution with a given number of degrees of freedom, the parameter that sets how fat its tails are. With five, extreme values are far more frequent than under the normal distribution. **Wilson interval**: the 95 per cent interval for a share counted out of a total (Wilson, 1927). **Turnover**: the absolute change in the position from the month before. The **break-even cost** is the cost per unit of turnover at which the alpha falls to zero. **Basis point**: a hundredth of a percentage point. **Expectation**: one of six pass-or-fail criteria, E1 to E6, in src/sq/constants.py (Section 2). Its **verdict** is met or not met, and a count that no expectation judges has no verdict. **Band**: the distance from a published number within which a computed one counts as matching. **Vintage**: the version of French's files, named by the CRSP database they were built from.

1. Introduction

A volatility-managed portfolio holds a factor in inverse proportion to the variance the factor showed in the month before, less after a volatile month and more after a calm one. Moreira and Muir (2017) applied the rule to the market and to the standard equity factors from 1926 to April 2015 and found alphas against the original factors that are positive and statistically significant at the 5 per cent level for the market, momentum and RMW among the six factors studied here. The market's is 4.86 per cent a year, with a standard error of 1.56. The rule needs one number from the past, last month's realised variance, and one constant, c, which sets the position's average size and is chosen on the whole sample. I ask how much of that result an investor could have used, in three parts: whether it reproduces on French's data as they stand today, whether it survives estimating the combination of the volatility-managed and the original factor in real time, and what the signal did in the months after Moreira and Muir's sample ended.

The existing evidence answers the second part and leaves the third open. Cederburg, O'Doherty, Wang and Yan (2020) argue that an alpha against the original factor is the gain of an investor who already holds the original and adds the volatility-managed factor in the amount the regression implies, an amount known only at the sample's end, like c. Estimating that amount each month from the months before it, they find the combination's Sharpe ratio below the real-time original's for four of the six factors studied here, and the difference statistically significant at the 5 per cent level on the Jobson-Korkie test only for momentum, where it is higher. That test assumes normal and independent months, and Moreira and Muir report alphas, so neither paper tests a single factor's Sharpe ratio difference in a way that allows for fat tails and for volatile months that follow each other. Neither can say what happened after its sample, which ends in April 2015 for Moreira and Muir, where their counts of months place it, and in December 2016 for Cederburg et al.

I replicate both papers and the scaled momentum of Barroso and Santa-Clara (2015) on French's six factors, each paper on its own sample and with its own definition of the realised variance. I test the Sharpe ratio differences with Ledoit and Wolf's (2008) test, which I first check on simulated pairs whose Sharpe ratios are equal by construction, and I carry the volatility-managed factor and the real-time combination forward unchanged to the 136 months from May 2015 to August 2026, with c fixed on the months to April 2015. I charge the trades the signal needs 10 basis points per unit of turnover, with 20 as a sensitivity. Every return reported is a simulation on French's history, with its standard error or interval beside it, and I estimate no premium.

French's factors are well suited to test these questions, for two reasons. First, both papers built their results on them and French keeps them current, so the replication and the months after the sample run on one source, and his revisions of history become visible: a factor whose statistics differ from a paper's on the same months may have changed since the paper's download, and an archived copy of the file can show by how much. Second, the months after the sample hold the highest realised variance of the market in French's history, March 2020, which cut the position for April to 0.013, a month in which the market returned 13.6 per cent, and they are few enough that every difference must be read against its interval.

Empirically, the replications hold except for Cederburg et al.'s two results on RMW, the factor whose mean has moved most since the paper. Moreira and Muir's Table I reproduces for 6 of 6 factors, scaled momentum in 4 of 4 comparisons and Cederburg et al.'s results in 22 of 24 comparisons, the two misses on RMW, whose mean in French's current file is 3.12 per cent a year on their months against their 2.92. On Moreira and Muir's sample the test of the alpha and the test of the Sharpe ratio difference disagree. The alphas of the market, momentum and RMW are statistically significant at the 5 per cent level on their Newey-West t-statistics after Holm's correction for the six tests. Momentum's Sharpe ratio difference, +0.51 annualised (0.07 to 0.96), is statistically significant at that level tested on its own (p 0.027) and not after the correction (p 0.16). An identity shows why the two tests differ. In the 136 months after the sample none of the twelve differences, S1 minus S2 and the volatility-managed factor minus the original for each of the six factors, is statistically significant at the 5 per cent level. At that length the test rejects 2.0 per cent of simulated pairs with equal Sharpe ratios (1.3 to 3.1) where 5 is intended, so these months are weak evidence either way.

Section 2 sets out the data, the constructions, the tests and the expectations. Section 3 reports the replications, the two tests on Moreira and Muir's sample, the months after it, and the costs and positions, and Section 4 sets them against the two papers' conclusions.

2. Design

Data. Kenneth French's data library (French, 2026), the vintage built from the CRSP database of September 2026 (202609): the three-factor and five-factor files, the momentum factor and the ten portfolios sorted on prior returns, monthly and daily, which the notebooks download at run time. I commit none of the files, and the committed tables carry the monthly factor returns behind every result. The market, SMB and HML start in July 1926, momentum and WML in January 1927, RMW and CMA in July 1963, and the sample ends in August 2026. I compare each paper's results on its own months: Moreira and Muir's to April 2015, the end their published counts of months imply (1,065 for the market, 621 for RMW and CMA), Cederburg et al.'s to December 2016, and Barroso and Santa-Clara's from May 1927 to December 2011.

The constructions. Each construction uses its paper's own realised variance: Moreira and Muir's sum of squared deviations of the daily returns from the month's mean (their equation 2), Cederburg et al.'s 22 over the month's trading days times the sum of squared daily returns (their equation 4), and Barroso and Santa-Clara's 21 times the mean of the squared daily returns over the previous 126 trading days (their equation 5). The volatility-managed factor's position is c over last month's realised variance. S1 holds the mean-variance weights of an investor with a risk aversion of 5, estimated from all the months before it once 120 are available, with its total position capped at 5.

The regression. I regress each volatility-managed factor on its original, monthly, by least squares. I judge Moreira and Muir's Table I with White's standard errors, the paper's own, and report the alpha in Sections 3.3 and 3.5 with Newey-West standard errors. I reproduce Cederburg et al.'s Sharpe ratio differences with their test, the Jobson-Korkie test.

The test of a Sharpe ratio difference. Ledoit and Wolf (2008) write the difference between two Sharpe ratios as a function of four means, those of each series and of its square, and take its standard error by the delta method, i.e., from the function's slopes and the long-run covariance of the four means. I take the covariance from the prewhitened Parzen kernel with Andrews' (1991) bandwidth, which the paper reports as performing very similarly to its QS kernel. I take the interval from a circular block bootstrap with 2,000 bootstrap samples of blocks of 12 consecutive months, each sample's difference studentised, i.e., divided by its own standard error, and the 95th percentile of the absolute studentised differences, the critical value z*, takes the place of the normal 1.96. Before I apply the test to any factor, I run it on 1,000 simulated pairs of 1,065 months with the market's monthly mean and standard deviation, a correlation of 0.601, Student-t shocks with five degrees of freedom and equal Sharpe ratios by construction, to measure its size. Section 3.4 gives its size at 136 months.

Costs and positions. I subtract the cost, 10 basis points per unit of turnover and 20 as the sensitivity, from the volatility-managed return, and estimate the alpha again on the net return. The factor's own rebalancing is the same for the volatility-managed and the original factor, and I leave it out. The break-even cost is the gross alpha over 12 times the average turnover. I compare the position's percentiles with Moreira and Muir's Table V and Cederburg et al.'s Table 1, Panel D.

The expectations. I judge the results against six expectations, each a strict pass or fail: E1 to E3 on the three replications, E4 on the test's size and its agreement with the alpha, E5 on the months after the sample, and E6 on the costs and the positions. Table 10 states each with what I found. I wrote every expectation into src/sq/constants.py before the code that tests it ran.

3. Results

3.1 The replications on each paper's sample

Table 1 reports Moreira and Muir's Table I beside the computed regression, and the replication holds for all six factors (E1 met). Every alpha has the published sign, is statistically significant at the 5 per cent level on its White standard error for the same three factors as in Table I, i.e., the market, momentum and RMW, lies within one published standard error of the published alpha, and has an appraisal ratio within 0.05 of the one the published figures give. The market's alpha is 4.93 per cent a year with a White standard error of 1.57, against 4.86 and 1.56, and its volatility-managed Sharpe ratio is 0.52, as in the paper's Table V. The largest gap is RMW's, 2.80 against 2.44 per cent a year, 0.44 of a published standard error, which Section 3.2 traces to a revision of French's data. The Newey-West standard errors are larger than White's for five factors, by up to 43 per cent for SMB, because the residuals of neighbouring months are correlated.

Table 1. Moreira and Muir's Table I on their sample, published beside computed: the alpha of each volatility-managed factor on its original, in per cent a year

		Alpha		*Standard error*			*Appraisal ratio*		
Factor	*Months*	*Published*	*Computed*	*Published*	*White*	*Newey-West*	*From the published alpha and RMSE*	*Computed*	*E1 holds*
Market	1065	4.86	4.93	1.56	1.57	1.76	0.33	0.33	yes
SMB	1065	−0.58	−0.59	0.91	0.91	1.31	−0.07	−0.07	yes
HML	1065	1.97	1.73	1.02	1.05	1.27	0.20	0.17	yes
Momentum	1060	12.51	12.83	1.71	1.72	2.05	0.86	0.87	yes
RMW	621	2.44	2.80	0.83	0.89	1.00	0.42	0.44	yes
CMA	621	0.38	0.36	0.67	0.67	0.55	0.08	0.07	yes

Each factor from its first month to April 2015: August 1926 for the market, SMB and HML, January 1927 for momentum, August 1963 for RMW and CMA. The published standard errors are White's, as are the computed ones the verdict uses, and the Newey-West standard errors are those of Sections 3.3 and 3.5. The appraisal ratio from the published figures is the published alpha over the published residual standard deviation (RMSE). The paper prints its own appraisal ratios for four of the six, i.e., the market 0.34, momentum 0.875, RMW 0.41 and HML 0.20 (p. 1620), and the computed ratios lie within 0.05 of these too. Source: results/202609/e1_comparison_202609.csv, notebook 03.

Table 2 reports Barroso and Santa-Clara's Table 3 for WML and scaled momentum, with the maximum drawdowns their text gives on p. 112, and their two headline results reproduce (E3 met). Scaling WML to 12 per cent a year raises its Sharpe ratio from 0.54 to 1.00, against the paper's 0.53 and 0.97, and cuts its maximum drawdown from 95.6 to 43.8 per cent, against 96.7 and 45.2. Scaled momentum's lowest monthly return is a fall of 24.2 per cent against WML's 77.7, and its excess kurtosis 2.0 against 18.0. On the 1,192 months from May 1927 to August 2026, the Sharpe ratios are 0.49 and 0.92 and the drawdowns the same.

Table 2. Barroso and Santa-Clara's Table 3 and maximum drawdowns on May 1927 to December 2011, published beside computed: WML and scaled momentum

	WML		*Scaled momentum*	
Statistic	*Published*	*Computed*	*Published*	*Computed*
Mean, % a year	14.46	14.86	16.50	16.69
Standard deviation, % a year	27.53	27.42	16.95	16.77
Sharpe ratio	0.53	0.54	0.97	1.00
Highest monthly return, %	26.18	26.24	21.95	21.71
Lowest monthly return, %	−78.96	−77.66	−28.40	−24.16
Skewness	−2.47	−2.39	−0.42	−0.36
Excess kurtosis	18.24	18.04	2.68	2.01
Maximum drawdown, %	−96.69	−95.56	−45.20	−43.83

1,016 months. The maximum drawdowns are from the paper's text, p. 112, and the other rows from its Table 3, p. 116. The paper's scaled momentum starts in March 1927, and French's daily file of the ten momentum portfolios starts on 3 November 1926, so scaled momentum here starts in May 1927. E3 judges the two Sharpe ratios and the two maximum drawdowns. Source: results/202609/bsc_table_3_computed_202609.csv and e3_comparison_202609.csv, notebook 05.

3.2 The real-time critique, and a revised factor

Tables 3 and 4 report Cederburg et al.'s Tables 1 and 5 beside the computed ones, and 22 of the 24 comparisons hold, so E2 is not met. Both misses are RMW's: its volatility-managed Sharpe ratio is 0.566 against their Table 1's 0.51, 0.006 outside the band of 0.05, and S1's is 0.510 against their Table 5's 0.44, 0.020 outside. The other comparisons hold. S1 minus S2 has the published sign for six of six factors, momentum's +0.41 is the only difference that is statistically significant at the 5 per cent level on the Jobson-Korkie test, as published, and S1 falls short of S3 by 0.08 to 0.12, against the published 0.07 to 0.14.

Table 3. Cederburg, O'Doherty, Wang and Yan's Table 1 on their sample, published beside computed: annualised Sharpe ratios of the original and the volatility-managed factor

	Original		*Volatility-managed*		*Difference*		*p-value*	
Factor	*Published*	*Computed*	*Published*	*Computed*	*Published*	*Computed*	*Published*	*Computed*
Market	0.42	0.420	0.51	0.513	+0.09	+0.092	0.30	0.31
SMB	0.23	0.226	0.08	0.076	−0.15	−0.150	0.09	0.09
HML	0.40	0.401	0.38	0.367	−0.02	−0.033	0.86	0.74
Momentum	0.48	0.475	0.99	0.962	+0.50	+0.487	0.00	0.00
RMW	0.38	0.406	0.51	0.566	+0.13	+0.161	0.29	0.20
CMA	0.53	0.534	0.40	0.392	−0.13	−0.142	0.23	0.19

Each factor from its first month to December 2016: 1,085 months for the market, SMB and HML, 1,080 for momentum, 641 for RMW and CMA. The p-value is that of the Jobson-Korkie test with Memmel's correction. E2's first clause, both Sharpe ratios within 0.05 of their Table 1, fails for RMW alone. Source: results/202609/cowy_table_1_computed_202609.csv, notebook 04.

Table 4. Cederburg, O'Doherty, Wang and Yan's Table 5, Panel A, on their sample, published beside computed: the real-time combination S1, the real-time original S2 and their differences

	S1		*S2*		*S1 − S2*		*p-value*		*S1 − S3*	
Factor	*Published*	*Computed*	*Published*	*Computed*	*Published*	*Computed*	*Published*	*Computed*	*Published*	*Computed*
Market	0.42	0.415	0.46	0.458	−0.04	−0.043	0.64	0.62	−0.11	−0.113
SMB	0.14	0.128	0.19	0.187	−0.06	−0.059	0.37	0.36	−0.12	−0.122
HML	0.38	0.375	0.43	0.442	−0.06	−0.067	0.41	0.33	−0.12	−0.122
Momentum	0.92	0.889	0.49	0.477	+0.44	+0.412	0.00	0.00	−0.07	−0.078
RMW	0.44	0.510	0.31	0.328	+0.13	+0.182	0.53	0.38	−0.14	−0.123
CMA	0.52	0.526	0.56	0.561	−0.03	−0.036	0.20	0.14	−0.11	−0.121

Months as in Table 3, each strategy from 120 months after the factor's first month. E2's second clause, S1 and S2 within 0.05 of their Table 5, fails for RMW alone, in S1, and its third and fourth clauses, on the sign of S1 minus S2 and on S1 minus S3, hold. Source: results/202609/cowy_table_5_computed_202609.csv, notebook 04.

The likeliest cause of the miss is French's revision of RMW since the paper's download. On the 641 months of their Table 1, RMW's mean in the current file is 3.12 per cent a year against their 2.92, with the same standard deviation, 7.70 against 7.71, while the other five factors' means lie within 0.12 of theirs. Two diagnostics, outside the verdicts, point the same way. Shifting RMW's monthly returns by one constant to the mean in their Table 1 closes the volatility-managed Sharpe ratio's gap from 0.056 to 0.022 and S1's from 0.070 to 0.027. The Internet Archive's copy of French's five-factor file of 9 July 2017 gives RMW a mean of 2.89 and a standard deviation of 7.74 on the same months, the mean within 0.03 of theirs, although I do not know the date of the paper's own download. About 40 per cent of each gap remains after the shift of the mean, and the data held here do not show where it comes from. RMW is not the only factor French has revised, since on the same months SMB's mean is 2.49 against their Table 1's 2.57 and momentum's 7.83 against 7.94, but RMW's Sharpe ratio moves the most, 0.027 above the one their Table 1's mean and standard deviation give, and RMW is the one factor that fails a band. The miss therefore most likely reflects a change in French's data, with about 40 per cent of each gap unexplained, and the replication of the construction holds on the other five factors.

3.3 The alpha and the Sharpe ratio difference

Ledoit and Wolf's test passes its size check at 1,065 months, the length of Moreira and Muir's sample: on the 1,000 simulated pairs it rejects 3.7 per cent at the 5 per cent level, inside the range of 3.5 to 6.5 expected. Table 5 reports it on Moreira and Muir's sample beside the alpha. Momentum's volatility-managed Sharpe ratio is 0.51 above the original's, annualised (0.07 to 0.96), which is statistically significant at the 5 per cent level when the factor is tested on its own (p 0.027). With Holm's correction for the six tests, momentum's p-value is 0.16 and every other factor's 0.94, so after the correction no factor's difference is statistically significant at that level. For the other five factors no difference is statistically significant even on its own, and with intervals that lie within −0.36 to 0.49, the data rule out a difference larger than 0.49 in either direction.

The tests of the alphas give a different result. Those of momentum, the market and RMW stay statistically significant at the 5 per cent level on their Newey-West t-statistics after the same correction, with adjusted p-values below 0.001, of 0.02 and of 0.02. For the market the Sharpe ratio difference, +0.10 (−0.11 to 0.31), is not statistically significant, while

the alpha, with a t-statistic of 2.80, is. For RMW the difference is +0.19 (−0.11 to 0.49), beside an alpha with a t-statistic of 2.81. E4 expected the two tests to agree in sign and significance for at least five factors, and they agree for two, SMB and momentum, so E4 is not met.

Table 5. The alpha and the Sharpe ratio difference on Moreira and Muir's sample: the volatility-managed factor against its original

Factor	*Alpha, % a year*	*t (Newey-West)*	*Sharpe ratio, volatility-managed*	*Sharpe ratio, original*	*Difference*	*95% interval*	*p-value*	*p-value, Holm*	*Correlation ρ*	*α/σ*	*(1 − ρ) × Sharpe ratio, original*
Market	4.93	2.80	0.515	0.419	+0.097	−0.112 to 0.305	0.33	0.94	0.601	+0.264	0.167
SMB	−0.59	−0.45	0.084	0.227	−0.143	−0.362 to 0.076	0.19	0.94	0.605	−0.053	0.090
HML	1.73	1.36	0.354	0.395	−0.041	−0.270 to 0.189	0.72	0.94	0.539	+0.141	0.182
Momentum	12.83	6.24	0.993	0.480	+0.514	0.072 to 0.955	0.03	0.16	0.450	+0.777	0.264
RMW	2.80	2.81	0.598	0.407	+0.192	−0.108 to 0.491	0.21	0.94	0.583	+0.361	0.170
CMA	0.36	0.66	0.414	0.538	−0.124	−0.337 to 0.090	0.24	0.94	0.672	+0.052	0.176

Months as in Table 1. The difference, its interval and its p-value are those of Ledoit and Wolf's test, with blocks of 12 months and 2,000 bootstrap samples, annualised. The Holm p-value carries Holm's correction for the six tests. The alphas' p-values are two-sided, from the normal distribution at each Newey-West t-statistic, and the source table holds them before and after the same correction. The difference equals α/σ minus the last column for every factor, to four decimals. Source: results/202609/sharpe_tests_mm_sample_202609.csv and alpha_and_sharpe_difference_202609.csv, notebook 06.

An identity shows why the two tests differ. With the volatility-managed factor scaled to the original's standard deviation σ, the regression's slope equals the correlation ρ between the two, and the volatility-managed factor's mean is α + ρμ, so its Sharpe ratio minus the original's equals α/σ minus (1 − ρ) times the original's Sharpe ratio. Table 5 reports both sides, equal to four decimals for every factor. The alpha therefore measures what the volatility-managed factor adds to a holding of the original, and the difference measures the volatility-managed factor held in the original's place, which gives up the share 1 − ρ of the original's Sharpe ratio. With correlations of 0.45 to 0.67, the alpha per unit of risk has to exceed a third to a half of the original's Sharpe ratio before the difference turns positive. For the market α/σ is 0.264 and (1 − ρ) times the Sharpe ratio 0.167, so the difference is +0.097, and for HML they are 0.141 and 0.182, so a positive alpha of 1.73 per cent a year comes with a difference of −0.041. Cederburg et al.'s Table 1 shows the same separation for the market, a difference with a p-value of 0.30 beside Moreira and Muir's t-statistic of 3.1.

3.4 The 136 months after the sample

Table 6 reports the Sharpe ratios of the original factor, the volatility-managed factor, S1 and S2 on May 2015 to August 2026, and the two differences, S1 minus S2 and the volatility-managed factor minus the original, and E5 is met. No factor's S1 minus S2 is positive by a margin that is statistically significant at the 5 per cent level, where at most two were allowed, and the median difference, +0.007, annualised, lies within the median half-width, 0.213. The six differences run from −0.090 for HML, with an interval of −0.234 to 0.053, to +0.198 for momentum, with an interval of −0.342 to 0.738. S1 and S2 are close where S1's position moves little: for SMB, HML and CMA their returns have correlations of 0.98 or more and the half-widths are 0.022 to 0.144, while S1's market position ran from 0.18 to 1.97 against S2's 0.44 to 0.49, a correlation of 0.798 and a half-width of 0.639.

Table 6. The 136 months after Moreira and Muir's sample, May 2015 to August 2026: Sharpe ratios and their differences with Ledoit and Wolf's intervals, annualised

	Sharpe ratio				*S1 − S2*			*Volatility-managed − original*		
Factor	*Original*	*Volatility-managed*	*S1*	*S2*	*Difference*	*95% interval*	*Half-width*	*Difference*	*95% interval*	*Half-width*
Market	0.784	0.554	0.693	0.775	−0.083	−0.722 to 0.556	0.639	−0.231	−1.120 to 0.659	0.889
SMB	−0.199	−0.301	−0.180	−0.211	+0.031	−0.092 to 0.154	0.123	−0.103	−0.690 to 0.485	0.587
HML	−0.019	−0.333	−0.141	−0.051	−0.090	−0.234 to 0.053	0.144	−0.315	−1.032 to 0.402	0.717
Momentum	0.207	0.383	0.394	0.197	+0.198	−0.342 to 0.738	0.540	+0.177	−0.449 to 0.803	0.626
RMW	0.270	0.410	0.311	0.258	+0.052	−0.229 to 0.334	0.282	+0.140	−0.394 to 0.675	0.535
CMA	−0.061	−0.317	−0.093	−0.075	−0.018	−0.040 to 0.005	0.022	−0.257	−1.046 to 0.532	0.789

c is fixed on the months to April 2015, and S1 and S2 are carried forward from their expanding windows. Each test uses 2,000 bootstrap samples of 11 blocks of 12 months. E5 judges the S1 − S2 columns, and no expectation judges the volatility-managed columns. Source: results/202609/after_sample_sharpe_202609.csv, e5_real_time_tests_202609.csv and e5_volatility_managed_tests_202609.csv, notebook 07.

The volatility-managed factor with c fixed on the months to April 2015 gives the same result: on no factor is its difference from the original positive by a margin that is statistically significant at the 5 per cent level, and the median difference of −0.167 lies within the median half-width of 0.672. Its Sharpe ratio is below the original's for four factors and above it for momentum and RMW. For the market it is 0.554 against the original's 0.784, a difference of −0.231 with an interval of −1.120 to 0.659. The three months with the largest gaps between the two returns show how the position works. In October 2018 the position was 3.56, after a calm September, and the market fell 7.65 per cent while the volatility-managed market fell 27.3 per cent. In April 2020 the position was 0.013, after March 2020, and the market rose 13.6 per cent while the volatility-managed market rose 0.18 per cent. In January 2018 the position was 3.38, and the market rose 5.58 per cent while the volatility-managed market rose 18.9 per cent.

The intervals are wide for two reasons. First, on 136 months against 621 to 1,065, the standard errors of the volatility-managed comparisons are 1.36 to 2.96 times those on Moreira and Muir's sample. Second, the critical value is 2.52 to 3.28 here against 2.05 to 2.66 there, because on 136 months each bootstrap sample holds only 11 blocks of 12, 132 months long. With 22 blocks of 6 it is 2.31 to 2.65, and with 45 blocks of 3 it is 2.09 to 2.57 (Table 7). Together the two widen the half-widths 1.42 to 4.27 times, to 0.53 to 0.89 annualised, against 0.21 to 0.44. An interval of −1.12 to 0.66 for the market is therefore consistent with a negative difference and with a positive one of the size estimated on Moreira and Muir's sample.

At 136 months in 11 blocks the test also rejects too rarely. On 1,000 simulated pairs of 136 months with equal Sharpe ratios, from the generator of Section 2, it rejects 2.0 per cent with blocks of 12, with a 95 per cent Wilson interval of 1.3 to 3.1, against the 5 per cent intended, so, for series like the simulated ones, the intervals of Table 6 are wider than 95 per cent intervals need be. With blocks of 6 it rejects 3.6 per cent (2.6 to 4.9), and with blocks of 3, 5.9 per cent (4.6 to 7.5), an interval that does not rule out 5. With blocks of 6 and of 3, none of the twelve differences is statistically significant at the 5 per cent level. The difference nearest its interval's edge is CMA's S1 minus S2, −0.018, with an interval of −0.040 to +0.005 at blocks of 12 and of −0.036 to +0.001 at blocks of 3. I judge E5 at blocks of 12, and since a test that rejects too rarely makes E5 easier to meet, the tests at blocks of 6 and 3 check that verdict.

Table 7. The test on 136 months by block length: its size on simulated pairs with equal Sharpe ratios, and the critical values of the twelve tests of Table 6

	A bootstrap sample		*Size, 1,000 simulated pairs*			*Critical value z*, lowest to highest of six*		
Block length, months	*Blocks*	*Months*	*Rejected*	*Rate*	*Wilson 95% interval*	*S1 − S2*	*Volatility-managed − original*	*Statistically significant at 5 per cent, of 12*
12	11	132	20	2.0%	1.3% to 3.1%	2.32 to 3.13	2.52 to 3.28	0
6	22	132	36	3.6%	2.6% to 4.9%	2.06 to 2.70	2.31 to 2.65	0
3	45	135	59	5.9%	4.6% to 7.5%	2.02 to 2.76	2.09 to 2.57	0

No expectation judges this table, and I judge E5 at blocks of 12 alone. Each simulated pair has the market's monthly mean and standard deviation and the correlation of its two series on Moreira and Muir's sample, with Student-t shocks of five degrees of freedom. At the six factors' correlations of S1 and S2 after the sample, 0.73 to 0.9996, the test rejects 1.7 to 1.9 per cent with blocks of 12 (95 per cent Wilson intervals of 1.1 to 2.7 at 17 rejections in 1,000 and 1.2 to 2.9 at 19), 2.7 to 3.5 with blocks of 6 and 5.4 to 5.7 with blocks of 3, so the low rate at blocks of 12 holds at those correlations too. Each test uses 2,000 bootstrap samples. Source: results/202609/size_check_136_months_202609.csv, size_check_after_sample_correlations_202609.csv and e5_tests_by_block_202609.csv, notebook 07.

3.5 What trading the signal costs, and how large the position gets

Table 8 reports the turnover and the alpha gross and net of costs. The market's alpha net of 10 basis points is statistically significant at the 5 per cent level on its Newey-West t-statistic, as E6 expects. The volatility-managed market's position changes by 0.752 of the factor a month on average, against Moreira and Muir's 0.73. At 10 basis points per unit of turnover its alpha is 4.03 per cent a year with a standard error of 1.74 (t 2.31), against the paper's 3.98. At 20 basis points it is 3.13 with a standard error of 1.73 (t 1.81), not statistically significant at the 5 per cent level, and the data rule out an alpha above 6.5 per cent a year at that cost. Each basis point of cost takes 12 times the average turnover, 9.0 basis points a year, from the market's alpha, so the break-even cost is 55 basis points, against the paper's 56. Momentum's alpha is 11.70 per cent a year net of 10 basis points (t 5.82) and 10.56 net of 20 (t 5.37), both statistically significant at that level, with a break-even of 112 basis points. RMW's is 2.06 net of 10 (t 2.09), statistically significant, and 1.32 net of 20 (t 1.35), not statistically significant, and the data rule out an alpha above 3.2 per cent a year at that cost. The alphas of SMB, HML and CMA are not statistically significant gross or net, and the data rule out any of them larger than 4.6 per cent a year in either direction.

Table 8. Turnover and the alpha gross and net of costs on Moreira and Muir's sample: the volatility-managed factor against its original

		Gross		*Net of 10 bp*		*Net of 20 bp*		
Factor	*Turnover, a month*	*Alpha, % a year*	*t*	*Alpha*	*t*	*Alpha*	*t*	*Break-even, bp*
Market	0.75	4.93	2.80	4.03	2.31	3.13	1.81	55
SMB	0.55	−0.59	−0.45	−1.25	−0.94	−1.92	−1.42	none
HML	0.69	1.73	1.36	0.90	0.70	0.07	0.06	21
Momentum	0.95	12.83	6.24	11.70	5.82	10.56	5.37	112
RMW	0.62	2.80	2.81	2.06	2.09	1.32	1.35	38
CMA	0.57	0.36	0.66	−0.32	−0.60	−1.01	−1.86	5

Months as in Table 1. The cost is charged on each unit of turnover, and the factor's own rebalancing, the same for the volatility-managed and the original factor, is left out. t-statistics with Newey-West standard errors. SMB's gross alpha is negative and has no break-even cost. Moreira and Muir's Table IV gives the market a turnover of 0.73, an alpha of 3.98 per cent net of 10 basis points and a break-even of 56. Source: results/202609/costs_mm_sample_202609.csv, notebook 08.

Table 9 reports the size of the position, and every percentile holds its band, so E6 is met in all 17 of its comparisons, i.e., the market's net alpha and 16 percentiles. The market's median position on Moreira and Muir's sample is 0.94 and its 99th percentile 7.27, against 0.93 and 6.39. On Cederburg et al.'s sample every factor's median lies within 0.03 of their Panel D and every 99th percentile within 0.48. The largest position, 16.8 for the market in March 1964, followed the market's calmest month of the sample, February 1964. For the market, a position of 7 holds seven times the capital in stocks, six of them borrowed at the bill rate, and a mandate that capped the position at 2 would have bound in 183 of the 1,065 months. The capped strategy is a different one, with its own alpha, which I do not estimate.

Table 9, Panel A. The size of the volatility-managed position, published beside computed: percentiles of c over last month's realised variance

	Median		*75th / 90th percentile*		*99th percentile*		
Factor and sample	*Published*	*Computed*	*Published*	*Computed*	*Published*	*Computed*	*Largest*
Market, Moreira and Muir's sample	0.93	0.94	1.59 / 2.64	1.62 / 2.71	6.39	7.27	16.8
Market, Cederburg et al.'s sample	0.96	0.96			6.47	6.00	13.1
SMB, Cederburg et al.'s sample	0.81	0.79			5.07	5.08	8.7
HML, Cederburg et al.'s sample	1.02	1.00			5.89	5.60	8.9
Momentum, Cederburg et al.'s sample	1.01	1.01			8.64	8.32	16.2
RMW, Cederburg et al.'s sample	1.11	1.09			5.02	5.27	8.7
CMA, Cederburg et al.'s sample	0.97	0.94			4.56	4.23	8.7

Moreira and Muir's Table V, Panel A, for the market, with their definition of the realised variance, and Cederburg et al.'s Table 1, Panel D, for the six factors, with theirs. E6 expects the median, 75th and 90th percentiles within 0.15 and the 99th within 1.0, and all 16 hold. Source: results/202609/position_market_mm_sample_202609.csv and position_cowy_sample_202609.csv, notebook 08.

The real-time combination's position is smaller and moves less (Table 9, Panel B). From its first month to August 2026, S1's median position runs from 0.43 for SMB to 1.57 for CMA and its 95th percentile from 1.45 to 4.82. It reaches the cap of 5 in at most 4.6 per cent of the months, for momentum, and it is below zero, i.e., a short holding of the factor, in 13.8 per cent of RMW's months and 6.4 per cent of SMB's. For the market S2 has a median position of 0.40, against S1's 0.56.

Table 9, Panel B. The real-time combination's position, its first month to August 2026: percentiles of S1's total position

			Percentile				*Share of months, %*		
Factor	*First month*	*Months*	*5th*	*Median*	*95th*	*Largest*	*Below 0*	*At the cap of 5*	*S2, median*
Market	August 1936	1081	0.19	0.56	1.78	5.00	0.8	0.3	0.40
SMB	August 1936	1081	−0.11	0.43	1.45	4.32	6.4	0.0	0.41
HML	August 1936	1081	0.19	0.75	2.08	5.00	2.6	0.1	0.58
Momentum	January 1937	1076	0.17	1.07	4.82	5.00	0.5	4.6	0.56
RMW	August 1973	637	−0.57	0.89	3.65	5.00	13.8	2.0	1.05
CMA	August 1973	637	1.01	1.57	1.99	2.47	0.6	0.0	1.57

S1's position is its total holding of the factor, i.e., the weight on the volatility-managed factor times its position plus the weight on the original, capped at 5 in either direction. Its first month is 120 months after the factor's first month. No expectation judges this panel. Source: results/202609/position_real_time_202609.csv, notebook 02.

Table 10. The six expectations and what the run found

No.	*Expected (src/sq/constants.py)*	*Found, vintage 202609*	*Verdict*
E1	On Moreira and Muir's sample, for at least five of the six factors: the alpha's sign and 5 per cent significance as in Table I, the alpha within one published standard error, the appraisal ratio within 0.05; for the market, the volatility-managed Sharpe ratio within 0.05 of 0.52	6 of 6 (Table 1)	met
E2	On Cederburg et al.'s sample, for each factor: (i) the original and volatility-managed Sharpe ratios within 0.05 of their Table 1; (ii) S1 and S2 within 0.05 of their Table 5; (iii) S1 − S2 with the published sign on at least five of six, momentum's the only significant difference; (iv) S1 − S3 within 0.05	22 of 24: (i) and (ii) fail for RMW alone, most likely because French has revised it since the paper; (iii) and (iv) hold (Tables 3 and 4)	not met
E3	On May 1927 to December 2011, the Sharpe ratios of WML and scaled momentum within 0.05 of 0.53 and 0.97, their maximum drawdowns within five points of 96.69 and 45.20 per cent	4 of 4 (Table 2)	met
E4	The size check rejects 3.5 to 6.5 per cent of 1,000 simulated pairs with equal Sharpe ratios; on Moreira and Muir's sample the bootstrap agrees with the alpha's Newey-West t-statistic in sign and significance for at least five of six factors	3.7 per cent; 2 of 6, the alpha and the Sharpe ratio difference being different tests; 4 of 6 by category, with no verdict (Table 5)	not met
E5	On May 2015 to August 2026, S1's Sharpe ratio exceeds S2's by more than the interval allows on at most two factors, and the median difference lies within the median half-width	0 of 6; +0.007 within 0.213; the volatility-managed factor, with no verdict, 0 of 6 and −0.167 within 0.672 (Table 6)	met
E6	The market's alpha net of 10 basis points positive and more than two Newey-West standard errors from zero; the market's position percentiles within 0.15 of Table V (99th within 1.0); each factor's median within 0.15 and 99th percentile within 1.0 of Cederburg et al.'s Panel D	4.03 per cent a year, t 2.31, and 3.13, t 1.81, at 20 basis points, which E6 does not judge; 17 of 17 (Tables 8 and 9)	met

Source: results/202609/e1_verdict_202609.csv to e6_verdict_202609.csv, notebooks 03 to 08.

4. Discussion

The natural comparison is with the two papers' own conclusions. On Moreira and Muir's sample the alphas reproduce for every factor, and the market's stays statistically significant at the 5 per cent level net of 10 basis points per unit of turnover (t 2.31). The paper reports the net alpha, 3.98 per cent, without a standard error. Cederburg et al.'s conclusion reproduces as well: an investor who estimated the combination in real time had a Sharpe ratio below the real-time original's for four of six factors, by 0.04 to 0.07, above it for momentum by 0.41, the one difference that is statistically significant at the 5 per cent level on the Jobson-Korkie test, and above it for RMW by 0.18. Ledoit and Wolf's test then compares the volatility-managed factor with the original directly. Held in the original's place, the volatility-managed factor has a Sharpe ratio higher than the original's by a margin that is statistically significant at the 5 per cent level for momentum alone, tested on its own, and for no factor after Holm's correction. A statistically significant alpha is therefore a statement about adding the volatility-managed factor to the original, at weights fixed on the whole sample, and the evidence for holding it alone rests on momentum's one test, read on its own, on this sample.

What the months after the sample imply is narrow. Neither construction shows a Sharpe ratio difference that is statistically significant at the 5 per cent level on any factor, and for the volatility-managed factors the intervals are too wide to rule out the estimates on Moreira and Muir's sample: the market's volatility-managed difference there, +0.097, lies inside its after-sample interval of −1.120 to 0.659, as does momentum's, +0.514, inside −0.449 to 0.803. These months are consistent with the signal doing after 2015 what it did before, with it doing less and with it doing more.

At the same time, the design has limits, each with its cause. First, I charge the costs on the change in a factor's position. The factors are long-short portfolios, and I leave out their own trading in the stocks that make them up, because it is the same for the volatility-managed and the original factor. A book that holds the stocks pays more, and the break-even costs bound what it could pay. Second, French revises history, so a rerun on a later vintage need not reproduce every digit, as RMW shows. Every table names its vintage, and the archived copy shows the size of one revision. Third, I fix the block length at 12 months and use the Parzen kernel, where Ledoit and Wolf choose the block length from the data and propose the QS kernel, and at 136 months that block length leaves 11 blocks in a bootstrap sample and a test that rejects 2.0 per cent of simulated pairs with equal Sharpe ratios where 5 is intended (Table 7). Fourth, I fix c for the

months after the sample on the months to April 2015, the paper's own constant, so the volatility-managed comparison after the sample measures what the paper's construction did next, and the real-time combination what an investor estimating it would have done. Finally, as in the paper, I never cap the volatility-managed factor's position, so its largest positions, up to 16.8, are positions no mandate would allow. Settling the question for the months after the sample would therefore require more of them.

Against this evidence, the claim is a narrow one, about the signal on French's six factors over these samples. The volatility-managed market, momentum and RMW earn alphas against their originals that are statistically significant at the 5 per cent level on Moreira and Muir's sample, the market's net of 10 basis points too (t 2.31) and not net of 20 (t 1.81). Held in the original's place, the volatility-managed factor has a Sharpe ratio higher than the original's by a margin that is statistically significant at that level for momentum alone, tested on its own, and for none after Holm's correction. Estimated in real time, the combination's Sharpe ratio is below the real-time original's for four of six factors on Cederburg et al.'s sample. And in the 136 months since the sample none of the twelve differences is statistically significant, with the intervals for the six volatility-managed factors too wide to rule out the differences found on Moreira and Muir's sample.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3), 817-858. https://doi.org/10.2307/2938229

Barroso, P. and Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, 116(1), 111-120. https://doi.org/10.1016/j.jfineco.2014.11.010

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1), 57-82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Cederburg, S., O'Doherty, M. S., Wang, F. and Yan, X. S. (2020). On the performance of volatility-managed portfolios. *Journal of Financial Economics*, 138(1), 95-117. https://doi.org/10.1016/j.jfineco.2020.04.015

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

French, K. R. (2026). Data library. mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, files built from the 202609 CRSP database.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65-70. https://www.jstor.org/stable/4615733

Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, 36(4), 889-908. https://doi.org/10.1111/j.1540-6261.1981.tb04891.x

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15(5), 850-859. https://doi.org/10.1016/j.jempfin.2008.03.002

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters*, 1(1), 21-23.

Moreira, A. and Muir, T. (2017). Volatility-managed portfolios. *Journal of Finance*, 72(4), 1611-1644. https://doi.org/10.1111/jofi.12513

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703-708. https://doi.org/10.2307/1913610

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817-838. https://doi.org/10.2307/1912934

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158), 209-212. https://www.jstor.org/stable/2276774