Factor Time

Boone Bowles, Adam V. Reed,

Matthew C. Ringgenberg, and Jacob R. Thornock*

August 2026

ABSTRACT

Factor models conventionally form portfolios using stale accounting data released months earlier. We construct "fresh" factors each month using the most recent data. Fresh factors modestly improve pricing, but augmenting existing models with the incremental fresh component reduces pricing errors by over 50%, because the two components carry different pricing content. This incremental component also reallocates measured abnormal returns, reverses the interpretation of event-study CARs for thousands of announcements, and changes performance classifications for many mutual funds. These findings point to a fundamental structure in returns: a slow-moving component reflecting persistent fundamental risk, and a fast-moving component capturing the continuous arrival of new information.

Keywords: Asset Pricing Tests, Factor Selection, Factor Models, Slow-Moving Capital, Stale Information

JEL Classification Numbers: G12, G14

*Mays Business School, Texas A&M University (boone.bowles@tamu.edu), Kenan-Flagler Business School, University of North Carolina (adam_reed@unc.edu), David Eccles School of Business, University of Utah (matthew.ringgenberg@eccles.utah.edu), Marriott School of Business, Brigham Young University (thornock@byu.edu). The authors thank Mike Cooper and conference and seminar participants at Texas A&M University. All errors are our own. ©2026 Boone Bowles, Adam V. Reed, Matthew C. Ringgenberg, and Jacob R. Thornock.

I. Introduction

Every major factor model in the literature, from Fama-French and Carhart to the q-factor model,[1] is premised on an assumption about *when* returns are earned. The traditional construction methodology forms factor portfolios annually, in June, to ensure that all data was publicly available at the time. But this construction implicitly assumes that returns do not exhibit different properties in the period immediately following information releases. Yet a recent literature has shown evidence of precisely that: returns exhibit different properties around macroeconomic news releases, firm-specific news article releases, and the release of financial statements.[2] If returns are driven by information arrival, why do standard factor construction methodologies ignore the period immediately following the release of financial statements? In this paper, we ask whether this timing assumption matters: we examine whether constructing factors from current, rather than stale, information affects the performance of standard factor models. We find that it does.

We refresh portfolio assignments by incorporating newly released accounting information as soon as it becomes available. Rather than reconstructing portfolios annually, in June, we reconstruct them at the end of each month using each firm's most recently released annual financial report. This lowers the average age of the information underlying the factor portfolios from approximately 350 days under the standard Fama-French methodology to approximately 200 days under ours – close to the theoretical minimum of 365/2=182.5 days achievable given that annual reports are released only once per year.[3] This is not an immaterial, cosmetic change to the inputs: refreshing the accounting-based sorting variables and the resulting portfolio weights changes the portfolios behind the factors, and those changes translate into economically significant differences in the factors themselves.

---

[1] Fama and French (1993, 2015, 2018); Carhart (1997); Hou et al. (2021).

[2] e.g., Savor and Wilson (2014) show that returns are different around macroeconomic news releases, Engelberg et al. (2012) show the relation between short sales and future returns is twice as large on news release days, and Engelberg et al. (2018) show that anomaly returns are 50% larger on news release days. Most recently Bowles et al. (2024) show that anomaly returns are concentrated in the first month after a firm's information release date and decay rapidly thereafter. And Bowles et al. (2026) extend the timeline backward, showing that returns drift predictably in the quarter before financial statements are released.

[3] As the calendar date moves further from June, the Fama-French methodology becomes increasingly stale. In May 2023, average information age falls from more than 500 days under the Fama and French methodology to approximately 110 days under our methodology.

In our main analyses, we examine the Fama and French (2018) factors: SMB (size), HML (book-to-market), RMW (profitability), and CMA (investment).[4] The largest changes appear in HML and CMA for straightforward economic reasons. HML is exposed to stale assignments because its sorting variable, book-to-market, relies on fast-moving market equity in the denominator which becomes significantly stale under standard construction methods.[5] CMA changes for a related reason: investment is less persistent than size or profitability, so stale classifications are more likely to miss meaningful shifts in firms' current investment states.[6] That economic logic shows up clearly in the data. The correlation between stale and fresh is only 0.813 for HML and 0.865 for CMA, versus 0.962 and 0.947 for SMB and RMW. Importantly, these differences survive controls for momentum and short-term reversal, indicating that the difference between stale and fresh is not simply a repackaging of familiar dynamic trading factors.

These differences suggest a natural hypothesis: because fresh factors more accurately measure firm information at the time of portfolio assignment, they should price assets better than stale ones. But refreshing does more than produce a fresh factor – it also isolates the traded payoff to revising stale portfolio assignments. The stale factor (S_t) captures the payoff to old assignments, the fresh factor (F_t) captures the payoff to refreshed assignments, and the difference between them captures the payoff to the revision margin itself. We call this difference the *update return*: $\Delta_t = F_t - S_t$. The question, then, is how this traded return should enter asset-pricing tests: should the update return be folded into a single, better-measured factor, or should the stale factor and the update return both enter as separate traded payoffs?

The first possibility treats the stale benchmark as an imperfect proxy for the true factor payoff, with the update return serving as a correction to that proxy. Put differently, replacing the stale factor with the fresh factor is equivalent to adding the full update return to the stale factor. But a correction need not be exact: the update return may contain noise, miss part of the relevant revision, or overstate it. If so, the best single factor is neither the stale nor the fresh factor, but

[4]We do not construct a fresh version of the market factor, since Mkt-RF is price-based rather than accounting-based and so is not subject to the same staleness problem.

[5]This point is also discussed in Asness and Frazzini (2013).

[6]The persistence of firm characteristics underlying anomaly portfolios emerges in Bowles et al. (2026).

a filtered combination that places the empirically appropriate weight on the stale component and the update return. This motivates the second possibility.

The second possibility holds that using only the stale or only the fresh factor is too restrictive, because the stale factor and the update return are distinct traded payoffs – one the return on the old assignment, the other the return generated when that assignment is revised. The fresh factor bundles these two payoffs into a single return and assigns each asset one exposure to the bundle. But assets need not covary with the stale factor and the update return in the same way, and if they do not, pricing should improve when the two enter separately. In this view, refreshing factors reveals a separate traded component embedded within the benchmark factor, rather than simply sharpening its measurement.

These two possibilities organize and motivate our empirical tests. We examine (i) replacing stale factors with fresh factors, (ii) filtering the factors one at a time, and (iii) specifications that let the stale factor and the update return enter separately. If refreshing factors mainly improves measurement, fresh or filtered factors should capture the relevant pricing information. But if refreshing factors reveals a distinct traded component, the update return should retain pricing content even after the individual factors are accounted for. We evaluate these possibilities using standard empirical asset-pricing tools: portfolio test assets, stock-level pricing exercises, cross-sectional R^2, RMSE, and model-comparison tests in the spirit of Barillas and Shanken (2017, 2018).

We find that replacing stale factors with fresh factors helps selectively, not broadly. Fresh factors improve pricing relative to stale factors using anomaly-based test assets, but in stock-level exercises and some portfolio families the gains are negligible. At first glance this is puzzling: given how much HML and CMA move when timing is refreshed, it seems reasonable to expect fresh factors to be significantly better. Yet they are not.

The results become dramatically stronger, however, when the update return is used alongside the stale factor. Rather than simply replacing the stale factor (S_t) with a fresh factor (F_t), we estimate pricing models of the form

$$R_{i,t+1} = \alpha_i + b_i S_t + c_i \Delta_t + \varepsilon_{i,t+1} \tag{1}$$

Although this specification is mechanically linked to fresh replacement, because $F_t = S_t + \Delta_t$, it is economically more flexible: it allows returns to load separately on the stale factor and the update return embedded in the fresh factor. This matters if the two components carry different pricing content, or if their combination provides a better-filtered proxy for the latent factor than either piece alone. Consistent with this, the specification in equation (1) performs significantly stronger across a wide variety of tests. Using a broad set of anomaly-based test assets, we find that adding the update return to the stale factor raises cross-sectional R^2 by approximately 8.5 percentage points, more than twice the 3.9 percentage-point gain from fresh-only replacement. The corresponding RMSE reduction is approximately 26.8 basis points, compared with 12.6 basis points under fresh-only replacement. The same pattern appears in stock-level exercises, where fresh-only replacement delivers no improvement while stale-plus-update is clearly better. Overall, pricing gains do not come from replacing stale with fresh. They come from allowing the update return to enter pricing directly.

Why does the update return help so much? One possibility is a single-best-proxy mechanism. Because fresh factors are still proxy portfolios, a filtered combination of the stale factor and the update return, $S_t + \kappa\Delta_t$, may be a better single traded proxy for the underlying latent factor than the fresh factor alone. A second possibility is a pricing-basis mechanism. Because $F_t = S_t + \Delta_t$, using only the fresh factor is equivalent to imposing a restriction that returns load on the two components (S_t and Δ_t) in the exact same way.[7] By contrast, allowing both S_t and Δ_t to enter separately[8] permits returns to load differently on the stale factor and the update return.[9] These two mechanisms have distinct empirical implications. The first says that the gains from refreshing should be absorbed by a better single filtered factor; the second says that pricing improves precisely because the stale factor and the update return enter as separate basis elements. We examine several tests to distinguish between these possibilities.

The first test targets the single-best-proxy mechanism. If the gains from refreshing come only from building a better single traded proxy for the underlying current factor, then a filtered factor

[7] $R_{i,t+1} = \alpha_i + b_i F_t + \varepsilon_{i,t+1} = \alpha_i + b_i S_t + b_i \Delta_t + \varepsilon_{i,t+1}$

[8] $R_{i,t+1} = \alpha_i + b_i S_t + c_i \Delta_t + \varepsilon_{i,t+1}$

[9] Economically, exposure to Δ_t means that an asset covaries not only with the stale factor's return, but with a different dimension of the factor – the payoff to revising stale portfolio assignments as new information arrives. Section II formalizes this interpretation.

of the form $S_t + \kappa\Delta_t$ should absorb the pricing gains from the update return. We therefore estimate the best κ and evaluate pricing models of the form $R_{i,t+1} = \alpha_i + \beta_i (S_t + \kappa\Delta_t) + \varepsilon_{i,t+1}$. This is the restricted version of the stale-plus-update specification: it allows the update return to matter, but only through a single loading on the filtered factor. We find that the best filtered factor yields only a modest improvement relative to fresh-only replacement, but it remains well short of specifications that allow stale and update to enter separately. The main gains from the update return therefore do not arise from finding a better single proxy.

That leaves the second mechanism: stale and update may matter not because they form a better single proxy, but because they form a better pricing basis. If the filtered-factor view were sufficient, returns should load on stale and update only through the single combination $S_t + \kappa\Delta_t$. By contrast, if stale and update are separate basis elements, the relevant pricing model is equation $R_{i,t+1} = \alpha_i + b_i S_t + c_i \Delta_t + \varepsilon_{i,t+1}$ and the restrictions implied by the filtered-factor case are too strong. The evidence supports this second view. Estimated loadings on stale and update differ meaningfully across assets, and allowing those loadings to vary delivers the strongest pricing performance. The data therefore point not to a single better proxy, but to a better pricing basis in which the stale factor and the update return matter separately.

Finally, we explore the implications of this. If updating changes the factor returns, then it could also change measured abnormal returns, event-study CARs, and mutual-fund performance classifications and rankings. We find that it does. Specifically, updating the factors reallocates a substantial amount of measured abnormal performance. The typical stock-month abnormal return changes by more than one percentage point, and the typical 22-day event-study CAR around earnings announcements and firm-specific news changes by roughly one-half to three-quarters of a percentage point. Perhaps more strikingly, we find the measured sign of the abnormal return flips and changes by more than 50bps in up to 3.4 percent of cases – economically significant given the thousands of earnings announcements and hundreds of thousands of news days in our sample. We find that *average* mutual-fund performance is unaffected: aggregate net and expense-added alpha are statistically indistinguishable whether we use the stale factor alone, the fresh factor, or the stale factor together with the update return. Yet this masks important heterogeneity across funds. We find that fund-level classifications are different: this choice of benchmark changes the sign of 6 to 9 percent of five-year fund alphas and replaces between one in seven and one in nine

top-decile funds. In other words, while average fund performance is unaffected, the measurement of which funds perform best and worst is significantly different. Together, these results show that factor timing changes not only which returns are labeled abnormal, but which events are judged informative and which funds are labeled skilled.

These results change our understanding of factor models. Using stale information to construct factors does not merely make benchmark factors noisier, it changes the portfolio behind the factor return. Moreover, the return generated when that portfolio is revised carries pricing content of its own. Put differently, the findings suggest there is a fundamental structure in returns: a slow-moving component reflecting persistent fundamental risk, and a fast-moving component capturing the continuous arrival of new information. Because Fama-French factors are used throughout empirical finance for risk adjustment, performance evaluation, event studies, and model comparison, timing is not a secondary implementation detail. It shapes measured alphas, estimated factor exposures, and the benchmark itself.

This paper adds to literature on factor construction and implementation. A natural starting point is Fama and French (1993), whose deliberately conservative timing convention remains embedded in the standard factors, and Fama and French (2015), which carries that convention forward to profitability and investment factors. Subsequent work makes clear that these construction choices are not innocuous: Asness and Frazzini (2013) show that using more timely prices materially changes HML, Fama and French (2018) show that model rankings and apparent factor redundancy can turn on seemingly modest design choices such as how profitability is measured or whether factors are defined as spreads or excess returns, and Akey et al. (2026) show that factor histories themselves are methodology-sensitive and can change retroactively. Against that background, we isolate timing as a first-order implementation choice and show that the economically relevant object is not simply a fresher factor, but the return to refreshing stale portfolio assignments.

More fundamentally, our paper belongs to the literature that treats observed factors as imperfect stand-ins for the underlying pricing forces they are meant to capture. Shanken (1987) provides the foundational insight: because the true pricing object is unobserved, empirical asset pricing must rely on observable factor returns that only approximate it. In different ways, Kelly et al. (2019) and Lettau and Pelger (2020) likewise study how returns and characteristics can be used to recover better empirical representations of latent pricing structure. Our approach is different. Rather than

recovering latent factor structure statistically, we isolate a specific and economically interpretable source of proxy distortion – the timing convention embedded in standard factor construction.

Another related literature studies whether the pricing information in firm characteristics reflects slow-moving components or more transitory movements. A broader tradition on characteristics and horizon-based return predictability includes Daniel and Titman (1997), Keloharju et al. (2021), and Moskowitz and Stambaugh (2021). Most directly related to our paper, Baba-Yara et al. (2024) show that persistent and transitory components can command different compensation across characteristics. The same logic carries naturally into our setting. The stale factor is the slow-moving component, while the update return is the revision margin created when stale assignments are refreshed. What we show is that this distinction has first-order consequences in traded factor models: fresh-only replacement helps only selectively, whereas allowing assets to load separately on the stale factor and the update return produces much larger pricing gains and a better traded pricing basis. The way we test these ideas fits squarely within the literature on comparing traded asset-pricing models. Gibbons et al. (1989) provide the classic absolute test of whether a traded model prices returns. For comparing competing traded models, the papers most closely connected to our design are Barillas and Shanken (2017, 2018), who show that model comparison turns on whether omitted factors retain alpha and on whether one model spans the relevant traded factor space. Closely related, Fama and French (2018) rank competing factor sets using maximum squared Sharpe ratios, and Barillas et al. (2020) develop formal model-comparison tests based on Sharpe-ratio improvements. Our paper brings that comparison logic to different implementations of the same factor dimension. Rather than compare unrelated factor families, we compare stale, fresh, and stale-plus-update representations, and show that stale-plus-update improves the traded pricing basis more clearly than fresh-only replacement.

Finally, the paper matters because Fama-French factors serve as benchmark inputs throughout empirical finance. They are used in cost-of-capital estimation, performance evaluation, and the evaluation of anomalies in work such as Fama and French (1997), Carhart (1997), McLean and Pontiff (2016), and Hou et al. (2020). Akey et al. (2026) underscore how sensitive such applications can be to the underlying factor series. What our paper adds is to show that, even holding fixed the broad factor family, timing conventions inside factor construction materially affect measured

betas, alphas, and model fit. Timing is therefore not a secondary implementation choice. It helps determine the content of the benchmark itself, and thus the conclusions researchers draw from it.

II. The Economics of Factor Timing

A. Stale and Fresh Factor Portfolios

The Fama–French convention is deliberately conservative (Fama and French, 1993, 2015). Firms are assigned to accounting-based portfolios at the end of June using information from the preceding fiscal year, and those portfolio assignments are maintained for the next twelve months. This rule gives financial statements ample time to become public and protects against look-ahead bias, but at a cost of mechanical staleness. Once the portfolios are formed, financial information released after June does not directly affect assignments until the following year. Standard factors therefore earn returns on portfolios whose underlying information is many months old. Accordingly, we use the term *Stale* to refer to factors constructed using this annual implementation.

We construct *Fresh* factors that preserve the definitions and two-by-three portfolio architecture of SMB, HML, RMW, and CMA. In other words, we follow the original construction methodology, only we change *when* information is eligible to enter portfolio formation. Instead of forming portfolios annually, in June, we form portfolios at each month-end using the most recent annual financial statement available. For market equity, we measure market equity at that formation date. The Fresh implementation does not use future information, revise realized returns, or introduce new characteristics. It simply uses more current information to determine the factor portfolios.

Monthly refreshing should not affect all factors equally. We expect the largest changes for HML and CMA. HML combines slow-moving book equity with a market-equity denominator measured at formation, so newly available accounting information and intervening price movements can move firms across value and growth portfolios. CMA should also be sensitive because investment rankings are less persistent, so a newly released balance sheet is likely to change a firm's investment assignment. By comparison, existing evidence shows that profitability rankings are more persistent, which should limit changes in RMW.[10] SMB is different: it averages size spreads across the

[10]See Bowles et al. (2026), Table IV, for evidence on the persistence of accounting-based anomaly portfolio assignments.

book-to-market, profitability, and investment sorts, and refreshed characteristics often leave a firm on the same side of the size breakpoint. We therefore expect the clearest differences between Stale and Fresh for HML and CMA.

This construction leads to a natural conjecture about replacing Stale factors with Fresh: Fresh may be a better benchmark than Stale. If using more timely public information produces a more useful implementation of the same factor, Fresh should improve cross-sectional fit and the traded opportunity set, and Stale should add little once Fresh is included. Whether these predictions hold is an empirical question.

B. The Update Return

Let r_{t+1} denote the vector of stock returns following formation date t, and let $w^S_{j,t}$ and $w^F_{j,t}$ denote the signed portfolio weights for factor j under the Stale and Fresh implementations. Their returns are

$$S_{j,t+1} = (w^S_{j,t})' r_{t+1}, \qquad \text{and} \qquad F_{j,t+1} = (w^F_{j,t})' r_{t+1}. \tag{2}$$

Both are contemporaneous traded returns at $t+1$. Their distinction is not between an old return and a new return, but comes from the portfolios earning those returns: one's portfolio weights reflect stale information, while the other's reflect fresher information. We define their difference as the *Update return*,

$$U_{j,t+1} = F_{j,t+1} - S_{j,t+1} = (w^F_{j,t} - w^S_{j,t})' r_{t+1}, \tag{3}$$

which is the observable return generated by the difference between the Fresh and Stale portfolio weights. For HML, the Update return captures the return generated when newly available accounting information and current market equity move firms toward or away from the value and growth portfolios. For CMA, the Update return captures the analogous effect when newly available balance-sheet information revises investment assignments.

Because Update is defined as the difference between Fresh and Stale, Fresh combines the payoff on the annually assigned portfolio with the incremental payoff generated when that portfolio is refreshed. Stale therefore captures a slower-moving maintained assignment, while Update captures the faster-moving change in portfolio assignments and weights. We use "slow" and "fast" only in

this implementation sense: Update is not assumed to be a primitive risk factor, a pure news shock, or a return realized only on filing dates.

C. One Better Factor or Separate Loadings?

The relation between Stale, Fresh, and Update raises two distinct possibilities. The first is *replacement*: monthly updating may help construct one better common factor, potentially better than either Stale or Fresh. The second is *expansion*: retaining Stale and allowing Update to enter the factor model separately may provide distinct return directions on which assets load differently.

To make the replacement view precise, suppose the observed portfolios are intended to proxy for an unobserved current factor payoff f_j^* (Huberman et al., 1987; Shanken, 1987). Stale may be an imperfect proxy, in part because its annual assignments do not incorporate recently available information. Fresh addresses this limitation by using newly available information to revise portfolios each month. Because $F = S + U$, treating Fresh as the preferred replacement for Stale amounts to adding the entire Update return to the proxy—that is, fixing its coefficient at one. But if Update only imperfectly captures the part of the factor payoff that Stale misses, a unit coefficient need not produce the best proxy for f_j^*.

This possibility motivates a broader class of one-factor proxies:

$$G_{j,t}(\kappa_j) = S_{j,t} + \kappa_j U_{j,t}. \tag{4}$$

This class contains both observed implementations. Stale is the endpoint $\kappa_j = 0$, while Fresh is the endpoint $\kappa_j = 1$. Other values shrink or expand the weight on Update. Thus, if Fresh does not outperform Stale in the pricing exercise, that result would not necessarily mean that Update is uninformative. It could instead mean that Update contains useful information, but that a unit weight on it does not provide the best one-factor proxy.

The true factor payoff f_j^* is unobserved, so we cannot directly determine which value of κ_j produces its best proxy. We instead choose κ_j using an observable pricing objective—for example, the combination of Stale and Update that best prices a given set of test assets. We therefore interpret $G_j(\kappa_j)$ as the best common factor for that pricing exercise, not as an estimate of the true factor itself.

The word *common* is important. Choosing κ_j creates one factor return, $G_j(\kappa_j)$, that is used to price every asset. Assets remain free to have different overall loadings on that factor, just as in a standard factor model. Once κ_j is selected, however, every asset must use the same relative combination of Stale and Update. To see this restriction, suppose asset i loads on the common factor with coefficient $\beta^G_{i,j}$:

$$R^e_{i,t} = \alpha_i + \beta^G_{i,j} G_{j,t}(\kappa_j) + \gamma_i' Z_t + \varepsilon_{i,t} \tag{5}$$

$$= \alpha_i + \beta^G_{i,j} S_{j,t} + \kappa_j \beta^G_{i,j} U_{j,t} + \gamma_i' Z_t + \varepsilon_{i,t}, \tag{6}$$

where Z_t contains the remaining factors. The implied component loadings are $\beta^S_{i,j} = \beta^G_{i,j}$ and $\beta^U_{i,j} = \kappa_j \beta^G_{i,j}$, so the common-factor representation imposes $\beta^U_{i,j} = \kappa_j \beta^S_{i,j}$. This restriction is factor-specific. HML and CMA may have different values of κ_j, and assets may have different overall loadings on each factor. But within a given factor, every asset must have the same relative exposure to Update and Stale. Fresh imposes $\beta^U_{i,j} = \beta^S_{i,j}$, while Stale imposes $\beta^U_{i,j} = 0$.

The expansion view begins from a different possibility. Stale and Update are separately observable traded payoffs, and assets need not covary with them in the same proportion. Two assets may, for example, have similar exposure to the Stale HML return but very different exposure to the HML Update return. A common HML factor cannot accommodate that difference because it requires every asset to use the same relative combination of Stale and Update.

We refer to the more flexible representation as *Separate*:

$$R^e_{i,t} = \alpha_i + \beta^S_{i,j} S_{j,t} + \beta^U_{i,j} U_{j,t} + \gamma_i' Z_t + \varepsilon_{i,t}. \tag{7}$$

Separate relaxes the common-loading restriction from the replacement view, allowing the relative exposure to Stale and Update to vary across assets: $\beta^U_{i,j}$ need not equal $\kappa_j \beta^S_{i,j}$. If this flexibility matters, the pricing benefit of Update does not come from finding one better version of HML or CMA. It comes from allowing assets to load differently on the slower-moving maintained portfolio and the faster-moving change in that portfolio.

D. Empirical Implications

These ideas organize the empirical tests. We first ask whether monthly updating materially changes the factor portfolios and returns, whether those changes are largest for HML and CMA, and what portfolio revisions generate the Update returns.

We then turn to the pricing questions: does Update matter because it helps construct one better common factor, or because assets benefit from separate exposures to Stale and Update? Under the replacement view, does Fresh improve on Stale, and can a different common combination in equation (4) provide a better one-factor proxy than either? Under the expansion view, do relative Stale and Update exposures differ across assets, and does Update add a traded return direction beyond Stale?

Finally, we ask whether factor timing matters for empirical inference. If updating changes the benchmark, it can also change measured rolling abnormal returns, event-study CARs, and mutual-fund performance classifications and rankings.

III. Data, Factor Construction, and Validation

We construct Stale and Fresh to preserve the familiar Fama–French factor definitions while changing when eligible information enters the portfolios. This section describes the data and construction, validates our implementation of Stale against the published Fama–French factors, and measures how much monthly updating reduces information age.

A. Data and Factor Construction

Factor construction combines CRSP stock returns and market equity with annual Compustat fundamentals. We follow the standard Fama–French investable-universe screens, characteristic definitions, breakpoint conventions, and portfolio-weighting rules (Fama and French, 1993, 2015). For Fresh, we assign each annual accounting observation an availability date using its matched 10-K filing date. An observation can first enter at a month-end strictly after its availability date, and the resulting assignment applies only to later returns.[11] The full sample runs from July 1972

[11] Filing dates come from Compustat Company's `CO_FILEDATE`. A missing filing date, or one more than 18 months after fiscal year-end, receives a flagged fallback date of June 30 in the year following the fiscal-year end. Before 1995, the capital-weighted fallback share is approximately 26 percent under both Stale and Fresh. From 1995 through 2025,

through December 2025. We also report January 1995 through December 2025 because observed filing dates are more complete in that period and public filings became more readily observable following the introduction of EDGAR. The full history is useful for long-sample return tests, while the later period provides the cleaner setting for interpreting exact information age.

We construct Stale and Fresh versions of the four nonmarket factors in the Fama–French five-factor model—SMB, HML, RMW, and CMA—using the standard two-by-three sorts. Stale follows the annual convention. At the end of June in year t, firms are assigned using accounting information associated with fiscal year $t-1$. HML pairs book equity with market equity from the preceding December, while RMW and CMA use operating profitability and investment from the corresponding annual financial statements. Assignments remain in place from July of year t through June of year $t+1$.

Fresh re-forms the factor portfolios at the end of every month. At each formation date, it uses the latest annual observation eligible under the availability rule described above and market equity measured at that formation date. HML therefore pairs the latest eligible book equity with current market equity. For RMW, Fresh uses operating profitability from the latest eligible annual statement. For CMA, it calculates investment from total assets in that statement and the firm's preceding annual observation.[12]

B. Replication of the Fama–French Factors

The tests that follow use our implementation of the annual factors, so we first compare it with the corresponding factors from the Kenneth French Data Library. For this validation, we refer to our annual-assignment series as *Reconstruction* and the published series as *Official*.

Table I summarizes the comparison across factors, return frequencies, and samples. In monthly data, Reconstruction–Official correlations range from 0.984 to 0.998 over the full history and from

---

the corresponding shares are 4.0 percent for Stale and 3.4 percent for Fresh. The Internet Appendix provides the complete data and filing-date procedures.

[12]Fresh does not introduce a new characteristic, use quarterly accounting data, change the factor formulas, or revise realized returns.

0.979 to 0.998 from 1995 through 2025. Full-history monthly variation retained is at least 96.7 percent.[13] Daily returns show similarly close agreement, and mean return differences are small.

[Table 1 about here.]

Figure 1 complements these summary statistics with an observation-level comparison. Each panel plots Reconstruction against Official for one factor, making individual deviations visible. For all four factors, the observations cluster around the 45-degree line and the fitted slopes are close to one in both sample periods. Together, Table I and Figure 1 show that our annual construction closely tracks the published factors.[14] The Stale–Fresh comparison then holds the source data and common construction procedures fixed, so its differences reflect the timing rule and the associated changes in eligible inputs, assignments, and weights rather than unrelated implementation choices. We use *Reconstruction* only for this validation and refer to our annual-assignment factors as *Stale* throughout the remaining analysis.

[Figure 1 about here.]

C. Information Age

Having validated the Stale benchmark, we next measure the age of the information underlying the portfolio assignments. For each accounting signal, information age is the number of days from the observation's availability date to the middle of a return month. Within each signal and month, we weight firms by market equity, then average equally across book-to-market, profitability, and investment and across months. Table II reports the average difference between Stale and Fresh. From 1995 through 2025, mean information age is 347 days under Stale and 203 days under Fresh—a reduction of 144 days, or almost five months. Under Stale, 43.2 percent of market capitalization is assigned using a financial statement more than one year old, compared with just 5.3 percent under Fresh.

[Table 2 about here.]

[13] Variation retained is $100 \times [1 - (\text{RMSE}/\text{SD}_{\text{Official}})^2]$. A value of 100 denotes exact equality; 96.7 means that the mean squared reconstruction error is 3.3 percent as large as the variance of the Official factor return.

[14] Exact equality is not expected because published factor histories can vary across data vintages and methodological revisions (Akey et al., 2026).

Figure 2 shows how this gap develops over the annual formation cycle. Under Stale, information becomes progressively older as the prior June formation date recedes into the past, exceeds 500 days in June, and then falls sharply when assignments refresh for July. Fresh uses younger information in every calendar month. Together, the exhibits show that the annual convention creates a large and predictable information lag that monthly formation substantially reduces.

[Figure 2 about here.]

D. Test Assets and Model Definitions

The recurring pricing tests use 49 Fama–French Industry portfolios, 153 U.S. long–short factor returns from Jensen et al. (2023), and a dynamic CRSP stock panel that preserves entry and exit. These asset families span diversified portfolios, characteristic-based strategies, and individual stocks.[15]

We compare three benchmark models. The *Stale* model contains Official MKT–RF, Official MOM, and Stale SMB, HML, RMW, and CMA. The *Fresh* model retains the same market and momentum controls and replaces the four Stale factors with their Fresh counterparts. The *Separate* model retains the Stale model and adds HML and CMA Update, each with its own loading. The factor-level evidence presented next motivates the focus on these two Update returns.

Before comparing the models, we first ask whether monthly updating materially changes the factor returns.

IV. What Monthly Updating Changes

Monthly updating makes the information in factor portfolios more current, but it need not materially change every factor return. We now ask which factors change most, when Stale and Fresh diverge most, and which portfolio revisions produce the Update return.

[15] Using diversified portfolios, anomaly strategies, and individual stocks also reduces the likelihood that model comparisons are driven by the selection or aggregation of test assets (Ang et al., 2020; Lewellen et al., 2010; Giglio et al., 2025).

A. HML and CMA Change Most

Table III shows that monthly updating changes HML and CMA substantially more than SMB and RMW, as predicted in Section II. From 1995 through 2025, the Stale–Fresh correlations are 0.813 for HML and 0.865 for CMA, compared with 0.962 for SMB and 0.947 for RMW. Monthly Update volatility is 245.4 basis points for HML and 123.9 basis points for CMA, versus 88.0 and 95.4 basis points for SMB and RMW. The same ordering holds over the full history.

[Table 3 about here.]

A large monthly difference, however, is not the same as a positive average payoff. HML Update averages only −1.84 basis points per month after 1994 and is not statistically distinguishable from zero. Fresh and Stale HML therefore have nearly identical average returns, even though they differ substantially from month to month.

The controlled results show what the unconditional mean hides. When HML Update is regressed on Official MKT–RF, Official MOM, and the four Stale characteristic factors, it has an alpha of 13.67 basis points and a MOM loading of −0.452. One source of this negative momentum exposure is mechanical: with book equity held fixed, recent winners become less value-like and recent losers become more value-like when Fresh uses formation-date market equity. Refreshing HML therefore tends to move against momentum. This exposure helps explain how a near-zero unconditional mean can coexist with a positive controlled intercept. The result echoes Asness and Frazzini (2013), who show that using timelier prices changes HML and that the performance of their timely-price value strategy becomes clearer once momentum is included in the benchmark.[16]

CMA is more straightforward. From 1995 through 2025, CMA Update averages 14.35 basis points per month and has a controlled alpha of 21.21 basis points. Unlike HML, CMA Update is positive both unconditionally and after controls. By comparison, neither SMB nor RMW has a large or statistically significant Update mean or controlled alpha in the later sample. Together, these results motivate our focus on HML and CMA and lead naturally to two questions: when are the return differences largest, and which portfolio revisions produce them?

[16] Asness and Frazzini (2013) focus on HML and use timelier price data while retaining the necessary lag for book equity. Our construction also allows newly public annual accounting information to enter, applies monthly formation to SMB, HML, RMW, and CMA, and isolates Update to distinguish replacement from expansion.

B. *The Return Difference Follows the Refresh Cycle*

The information-age results in Section III show that the information underlying Stale assignments becomes progressively older as the portfolio year advances. If that information gap drives the return difference, the absolute Update return should also grow as the Stale assignments age. Figure 3 plots calendar-month averages of $|U|$ for all four factors. In every factor and both samples, absolute Update is generally larger from January through June and falls when the Stale portfolios refresh for July.

[Figure 3 about here.]

Over the full history, the January–June minus July–December difference ranges from 23.7 basis points per month for SMB to 51.6 basis points for CMA. From 1995 through 2025, the corresponding range is 28.8 to 67.0 basis points. Because the figure plots absolute Update, the pattern describes how far Fresh and Stale move apart, not which implementation earns the higher return. The two implementations are farther apart, on average, as the Stale portfolios approach their annual refresh and move closer together when those portfolios are re-formed in July.

C. *What Produces the Update Return*

Monthly updating can change a factor portfolio in several ways. Updated inputs may move a firm across the factor's characteristic buckets—for example, from value toward growth or from conservative toward aggressive—thereby changing its contribution to the factor's long or short side. Even when the characteristic assignment is unchanged, Fresh may place the firm on the other side of the size breakpoint, assign it a different weight within the same portfolio, or include it when Stale does not, and vice versa.

Table IV separates characteristic reassignment—the first channel—from the remaining portfolio revisions. For each firm i, the Update portfolio weight is $w^F_{i,t} - w^S_{i,t}$. Panel A assigns this weight difference to one of the two groups and reports each group's contribution to the Update return and share of absolute Update exposure. The decomposition applies to HML, RMW, and CMA. SMB

is excluded because it averages size spreads across three characteristic sorts and therefore does not admit the same one-characteristic partition.[17]

[Table 4 about here.]

Over the full history, characteristic reassignment accounts for 69.3 percent of absolute HML Update exposure and 69.1 percent of CMA Update exposure. The corresponding RMW share is 47.4 percent. Thus, the factors that change most are also the factors for which moving firms across characteristic buckets accounts for most of the Update portfolio.

The signed return contributions show that HML and CMA arrive at their results differently. For HML, characteristic reassignment contributes −11.13 basis points per month to the mean Update return, while the other revisions contribute 11.94 basis points. These components nearly offset in the unconditional mean. After controls, however, characteristic reassignment contributes 21.37 of HML Update's total 28.40 basis-point alpha. For CMA, there is much less offset: characteristic reassignment produces 9.54 of its 10.95 basis-point mean and 16.37 of its 16.28 basis-point controlled alpha. RMW shows no comparable pattern.

Panel B reaches the same conclusion from the portfolio assignments themselves. Stale and Fresh place 27.0 percent of firms in different book-to-market buckets and 20.0 percent in different investment buckets, compared with 10.1 percent for profitability and 2.3 percent for size. The ordering remains under market-capitalization weights: the mismatch shares are 16.4 percent for HML and 17.5 percent for CMA, versus 8.2 percent for RMW and 1.3 percent for SMB. Monthly updating therefore changes HML and CMA mainly by changing characteristic assignments, not merely by reweighting otherwise unchanged portfolios. This evidence motivates our focus on HML and CMA Update in the pricing tests that follow.

V. Replacing Stale or Expanding the Pricing Basis?

Section IV shows that monthly updating materially changes HML and CMA. The next question is whether those changes improve the benchmark through replacement or expansion. Under the

[17] Characteristic reassignment includes a change in the factor's characteristic bucket, with or without a simultaneous size-bucket change. An eligibility change occurs when a firm receives a portfolio assignment under one implementation but not the other. Exposure is each group's share of total absolute Update weights, so the two groups sum to 100 percent within each factor.

replacement view, the natural first possibility is that Fresh simply displaces Stale one for one.[18] Under the expansion view, Stale and Update enter the model separately, allowing their relative loadings to vary across assets. Accordingly, we first compare Stale with Fresh—models of the same dimension that differ only in the implementation of the four characteristic factors—and then compare Stale with Separate, which retains Stale and adds HML and CMA Update with independent loadings. We evaluate these comparisons using rolling cross-sectional fit and traded-return tests that ask whether Fresh replaces Stale and whether Update contributes incremental mean–variance content beyond Stale.[19]

A. Rolling Exposures and Cross-Sectional Fit

Table V evaluates the three models using 49 Industry portfolios, 153 JKP factors, and a dynamic panel of CRSP stocks. We use a rolling two-pass procedure so that the exposures used to explain returns in each evaluation month are estimated only from earlier returns. In the first pass, we estimate model-specific exposures using the preceding 60 calendar months through $t-1$ and require at least 36 usable observations. For model m and asset i, we estimate

$$R^e_{i,s} = \alpha^m_{i,t-1} + \left(\beta^m_{i,t-1}\right)' f^m_s + \varepsilon^m_{i,s}, \qquad s = t-60, \ldots, t-1, \tag{8}$$

where $m \in \{\text{Stale}, \text{Fresh}, \text{Separate}\}$. In the second pass, following Fama and MacBeth (1973), we regress returns across assets in month t on those predetermined exposures:

$$R^e_{i,t} = \lambda^m_{0,t} + \left(\widehat{\beta}^m_{i,t-1}\right)' \lambda^m_t + e^m_{i,t}. \tag{9}$$

This second pass asks how well exposures estimated before month t line up with returns realized across assets during month t. Our focus is not on the individual estimates of λ^m_t, but on cross-

[18] The replacement view is broader than simple Fresh-for-Stale substitution. A common factor that combines Stale and Update using a freely chosen common weight may provide a better proxy than either observed implementation. Section II develops this possibility, and Section VI tests it directly.

[19] Kan et al. (2013) study cross-sectional R^2 as a measure of pricing-model performance and its use in model comparison. Following Barillas and Shanken (2017, 2018), we complement cross-sectional fit on selected test assets with comparisons of the traded factor returns themselves. The exact joint-alpha tests build on Gibbons et al. (1989).

sectional fit: a higher adjusted R^2 means that a model's predetermined exposures account for more of the differences in returns across assets.

[Table 5 about here.]

Panel A summarizes fit in the first-pass time-series regressions and shows that Fresh does not systematically improve this fit relative to Stale, whereas Separate has the highest adjusted R^2 for every asset family and sample period. We treat these averages as descriptive because adjacent estimates rely on heavily overlapping 60-month windows.

Panel B contains the main result. From 1995 through 2025, Fresh changes cross-sectional adjusted R^2 by only -0.05 to 0.12 percentage points across the three asset families, and none of the differences is statistically distinguishable from zero. The full history produces the same conclusion. This is the Fresh replacement puzzle: monthly updating materially changes HML and CMA, yet replacing Stale with Fresh does not make estimated exposures more useful for explaining cross-sectional returns.

Separate produces a different result. Retaining the Stale factors and adding independent loadings on the HML and CMA Updates raises adjusted R^2 by 0.47 to 3.33 percentage points after 1994, with all three gains significant at the 1 percent level. The full-history estimates are similar. Separate estimates two additional loadings for each asset, so added flexibility remains a natural concern. Four robustness checks indicate that the Panel B gains are difficult to explain as a purely mechanical consequence of those additional loadings, particularly for JKP factors and stocks.[20] The benefit of monthly updating therefore appears when Update enters alongside Stale, not when Fresh replaces Stale one for one.

A higher adjusted R^2 for a particular set of test assets does not by itself establish that Separate is a better pricing model. The Update returns can improve fit because assets covary with them

[20] First, adjusted R^2 already penalizes Separate for the two additional slopes it estimates. Holding raw fit fixed, the gains would become insignificant even at the 10 percent level only if the penalty were increased by the equivalent of 1.39 additional slopes for industries, 12.62 for JKP factors, and 6.67 for stocks, beyond the two slopes Separate actually adds. Second, all three gains remain significant at the 1 percent level with either 12 or 59 Newey–West lags (Newey and West, 1987). Third, when both Updates are retained and all exposures are re-estimated, removing any one Stale characteristic factor reduces adjusted fit by 1.36–2.00 percentage points for industries, 2.30–3.19 points for JKP factors, and 0.36–0.64 points for stocks—broadly the same order of magnitude as the corresponding gains from adding the Updates. Finally, we compare the fit gains from the actual Updates with those produced by 200 simulated pairs of generic Gaussian returns constructed to match the Updates' covariance structure with Stale and with each other. The industry gain is not unusual under this benchmark, whereas the simulated returns rarely match the JKP or stock gains. Detailed results are available from the authors upon request.

even if the Updates earn zero alpha relative to Stale and therefore add no new mean–variance opportunity. For traded-factor models, the natural comparison therefore uses the factor returns themselves. Table VI applies this approach.

B. Traded-Model Comparisons

Table VI asks whether Update contributes incremental mean–variance content beyond Stale in three ways. Panel A asks whether Fresh or Separate offers a larger mean–variance opportunity set than Stale after adjusting for finite-sample bias from model dimension (Fama and French, 2018; Barillas et al., 2020). Panel B asks whether returns included in one model but omitted from the other earn alpha relative to the benchmark model (Gibbons et al., 1989; Barillas and Shanken, 2017). Panel C raises the hurdle for concluding that the added returns are nonredundant by allowing the redundancy null to include small nonzero alphas (Barillas and Shanken, 2018).

[Table 6 about here.]

Panel A asks how much investment opportunity is contained in each model's traded returns. For a model with K factor returns observed over T months, let $\bar{f}$ and $\widehat{\Sigma}$ denote their sample mean and covariance matrix. The sample maximum squared Sharpe ratio and its bias-adjusted counterpart are

$$\widehat{\theta}^2 = \bar{f}'\widehat{\Sigma}^{-1}\bar{f}, \qquad \widehat{\theta}^2_{\text{adj}} = \frac{T-K-2}{T}\widehat{\theta}^2 - \frac{K}{T}.$$

The first quantity is the highest squared Sharpe ratio attainable in sample by combining the model's traded factors. It is upward biased because the portfolio weights are chosen to maximize Sharpe in the same sample, and that bias increases with the number of factors. Following Fama and French (2018), we therefore report the bias-adjusted monthly measure, denoted adjusted θ^2, together with the equivalent bias-adjusted annualized maximum Sharpe ratio, $\sqrt{12\,\widehat{\theta}^2_{\text{adj}}}$. Stale and Fresh each contain six traded returns, while Separate contains eight, so the adjustment explicitly penalizes Separate for its two additional return directions.

For inference, we ask whether the differences in these opportunity sets reflect incremental traded content rather than sampling error. Separate nests Stale, so the exact finite-sample test of Gibbons et al. (1989) asks whether the conditional alphas of HML and CMA Update are jointly zero relative

to Stale. Rejection means that the added returns expand the attainable mean–variance opportunity set. Fresh and Stale are nonnested—neither model contains the other—so we use the sequential maximum-Sharpe procedure of Barillas et al. (2020) to compare their bias-adjusted opportunity sets.

Panel A reports the opportunity-set levels as annualized maximum Sharpe and monthly adjusted θ^2, and reports changes relative to Stale in monthly adjusted θ^2 units. Over the full history, the annualized maximum Sharpe ratios are 1.11 for Stale, 1.27 for Fresh, and 1.48 for Separate. In monthly adjusted θ^2 units, Fresh exceeds Stale by 0.032, a difference significant at the 5 percent level. From 1995 through 2025, the increase is 0.017 and is not statistically significant. Separate raises adjusted θ^2 by 0.081 over the full history and 0.035 after 1994, with both differences significant at the 1 percent level. Thus, Update produces a larger and more consistently detectable improvement in the attainable risk–return tradeoff when it enters alongside Stale than when it is incorporated into Fresh and used to replace Stale.

Panel B asks whether the traded returns in one model contain mean–variance opportunities that a benchmark model fails to capture. Following the omitted-factor logic of Barillas and Shanken (2017), let $g_{j,t}$ denote one of the factor returns being tested and let f_t collect the returns in the benchmark model. For each tested return, we estimate

$$g_{j,t} = \alpha_j + \beta_j' f_t + \varepsilon_{j,t}.$$

The slope coefficients β_j capture how the tested return moves with the benchmark factors, while α_j measures the average return left unexplained by those exposures. If the benchmark fully captures the investment opportunities in the entire tested block, all of these alphas should be zero. The joint test of Gibbons et al. (1989) evaluates that restriction. Rejection means that some combination of the tested returns provides a mean–variance opportunity that is not available from the benchmark alone. Root mean square (RMS) alpha summarizes the typical magnitude of the unexplained average returns, while the p-value tests whether the alphas are jointly zero. Thus, Fresh beyond Stale asks whether Fresh contains something Stale does not, and Stale beyond Fresh asks whether Stale contains something Fresh does not.

Fresh adds beyond Stale in both samples, but Stale also adds beyond Fresh. Neither implementation therefore spans the other. Put plainly, Fresh captures some investment opportunities that Stale misses, while Stale captures others that Fresh misses.[21] To test the expansion view, the Separate beyond Stale row asks whether HML and CMA Update add anything further. The null that the HML Update and CMA Update alphas are jointly zero is rejected at the 1 percent level in both samples, with RMS alphas of 23.1 basis points per month over the full history and 17.8 basis points after 1994. These are the largest RMS alphas in Panel B, making the incremental traded content of Update clearest when it enters alongside Stale.

Panel B treats a tested return block as redundant—that is, as offering no additional mean–variance opportunity—only if all of its alphas are zero. With long samples, however, economically small but precisely measured alphas can reject that null. Panel C therefore raises the hurdle. Following Barillas and Shanken (2018), it asks whether the alphas are large enough to provide convincing evidence of an additional investment opportunity, rather than merely whether they differ from zero. To calibrate the scale of a small alpha, we consider annual standard deviations of 50 and 100 basis points.[22] The table then reports, for each calibration, the posterior probability that the tested returns jointly add to the benchmark after allowing their alphas to deviate modestly from zero. These probabilities are not p-values: 50 percent represents equal posterior odds, and values above 50 percent favor the additional-opportunity explanation. The 100-basis-point calibration allows larger alphas before concluding that the tested returns add to the benchmark and therefore provides the harder test.

This higher hurdle weakens the Fresh-for-Stale replacement evidence. Under the 100-basis-point null, the posterior probability that Fresh is nonredundant relative to Stale is 65.8 percent over the full history and 33.2 percent after 1994. The corresponding probabilities that Stale is nonredundant

[21] Because $U = F - S$, Separate beyond Fresh and Stale beyond Fresh generate the same incremental span, which is why the corresponding rows in Panel B are identical.

[22] The method compares two possibilities. In the first, the benchmark remains sufficient even though the tested returns may have modest alphas; the 50- and 100-basis-point settings describe the typical annual size of those alphas. In the second, the tested returns expand the opportunity set; following Barillas and Shanken (2018), this alternative is calibrated using a maximum Sharpe ratio 1.5 times that of the benchmark. The method begins with equal odds on the two possibilities and uses the data to update those odds. The 50- and 100-basis-point settings are not cutoffs applied to one alpha at a time: the calculation considers the alphas jointly, including their magnitudes, precision, and covariance.

relative to Fresh are both below equal posterior odds. The Panel B rejections therefore do not produce a robust case for replacing Stale with Fresh.

The expansion evidence for adding Update to Stale through the Separate model remains stronger. Under the 50-basis-point null, the posterior probability that HML and CMA Update are nonredundant relative to Stale is 100.0 percent over the full history and 84.8 percent after 1994. Under the harder 100-basis-point null, the corresponding probabilities are 91.4 and 52.4 percent. Expansion is therefore strongly supported over the full history. In the later sample, the evidence is substantial under the 50-basis-point calibration and, even under the harder 100-basis-point calibration, remains just above equal odds in favor of the additional-opportunity explanation.

Taken together, Tables V and VI favor expansion through Separate over one-for-one replacement of Stale with Fresh. Fresh materially changes the factor portfolios but does not systematically improve rolling cross-sectional fit, and replacing Stale with Fresh gives up some traded content that Stale retains. Separate retains Stale, improves fit, and provides stronger evidence that HML and CMA Update add traded content beyond the Stale model. Among the three models compared here, the evidence is strongest for Separate.

The broader replacement view nevertheless remains open. Replacing Stale with Fresh fixes the coefficient on Update at one, while another common mixture $S_j + \kappa_j U_j$ may use the information in Update more effectively without requiring separate asset-level exposures. Section II develops this possibility. We test that broader replacement view next.

VI. One Better Factor or Separate Loadings?

The remaining comparison turns on the exposures each model permits. Under the replacement view, any common HML or CMA factor requires every asset to use the same relative combination of Stale and Update. Under the expansion view, Separate allows that relative exposure to vary across assets. We ask whether choosing the common mixtures as favorably as possible is enough to match Separate's cross-sectional pricing fit.

Table VII conducts this comparison over 1995–2025. Every specification includes Official MKT–RF, Official MOM, Stale SMB, and Stale RMW. For HML and CMA, the common-factor specifications take the form in equation (4). Stale sets $\kappa_H = \kappa_C = 0$, while Fresh HML/CMA sets

$\kappa_H = \kappa_C = 1$. Fresh HML/CMA is therefore the Fresh endpoint for this focused comparison, not the fully Fresh model examined in Section V. Best Common jointly chooses κ_H and κ_C within each test-asset family to maximize average monthly cross-sectional adjusted R^2. The resulting HML and CMA mixtures are shared by all assets within that family, although each asset retains its own overall loading on each factor. Separate follows equation (7): it retains Stale HML and CMA and allows each asset to load independently on Stale and Update.[23]

[Table 7 about here.]

Panel A deliberately gives the common-factor explanation the benefit of hindsight. For each of the three test-asset families—Industries, JKP factors, and stocks—we search jointly over the HML and CMA weights and select the pair that produces the highest average monthly adjusted R^2 in the same 1995–2025 sample used to evaluate fit. Because the search includes both Stale and Fresh HML/CMA, Best Common cannot fit worse than either endpoint in this sample. The relevant questions are how much fit the optimized weights gain and whether they close the gap with Separate.[24]

Choosing the best common weights helps, but unevenly and not enough to match Separate's fit. Best Common raises adjusted R^2 relative to Fresh HML/CMA by 1.92 percentage points for Industry portfolios, 0.21 points for JKP factors, and 0.24 points for stocks. Relative to Stale, Best Common also improves fit for Industries and JKP factors, but essentially matches Stale for stocks. The more demanding comparison is Best Common against Separate. Separate exceeds Best Common by 1.29 percentage points for Industries, 2.42 points for JKP factors, and 1.12 points for stocks, with all three differences significant at the 1 percent level. Thus, Separate's fit advantage does not arise simply because Fresh fixes the HML and CMA Update weights at one. Even common

[23]Unlike Table V, Table VII does not estimate exposures using rolling windows that end before each evaluation month. It estimates each asset's exposures once using all observations in the period being evaluated. Every model uses the same asset-month observations and exposure estimator. This is therefore a contemporaneous fit comparison, and its adjusted-R^2 levels should not be compared directly with the rolling levels in Table V.

[24]For each test-asset family, we first compare 144 possible pairs of common factors: 12 Stale–Update mixtures for HML crossed with 12 mixtures for CMA. This initial grid includes Stale, Fresh, Update alone, and mixtures beyond the Stale–Fresh interval. We then start from the best pair on the grid, allow both coefficients to vary continuously, and select the pair that produces the highest average monthly adjusted R^2. Because we choose and evaluate the coefficients using the same sample, the reported standard errors do not account for the search; they allow for six months of serial dependence but treat the selected coefficients as given. We therefore emphasize the size and consistency of the fit differences across asset families.

factors tailored to each asset family with full-sample hindsight fit the test assets less well than does Separate.

Panel B asks whether the best common mixtures transfer through time and across test assets. Each row selects κ_H and κ_C using one source family from 1995 through 2009. Each column then evaluates that fixed pair using one target family from 2010 through 2025. For example, the upper-left cell takes the HML and CMA weights chosen using Industry portfolios in 1995–2009, $(\kappa_H, \kappa_C) = (2.35, -9.56)$, and evaluates those common factors using Industry portfolios in 2010–2025. The -1.78 entry means that this earlier Industry mixture produces an average adjusted R^2 1.78 percentage points lower than Fresh HML/CMA in the later Industry sample. The diagonal cells therefore test whether a mixture transfers through time within the same asset family, while the off-diagonal cells also test whether it transfers across asset families.[25]

The selected mixtures travel poorly. Eight of the nine transported mixtures fit worse than Fresh HML/CMA in the later period, and the only positive difference is 0.09 percentage points. The three diagonal entries are -1.78, -0.15, and -0.01 percentage points, so none of the earlier mixtures improves on Fresh HML/CMA even when it is later evaluated on the same asset family that selected it. The cross-family results are no better. A common mixture tailored to one family and period therefore does not reliably improve fit elsewhere.

One explanation for the weak transport is that the preferred common mixture changes over time. Under the replacement view, this means that the best-fitting balance between the slower-moving maintained portfolio and the faster-moving change in that portfolio is not stable. The coefficient pairs beneath the three column headings—$(0.26, 1.10)$ for Industries, $(0.58, 0.78)$ for JKP factors, and $(1.00, 0.08)$ for stocks—are the Best Common mixtures chosen anew in 2010–2025. They differ substantially from the corresponding 1995–2009 pairs shown in the rows, especially for Industries. This instability is itself a weakness for the replacement view, which seeks a common factor definition that remains useful across assets and through time.

[25]Only the two factor-construction weights, κ_H and κ_C, are carried from the earlier period into the later one. Each target asset's loadings on the resulting factors are re-estimated using its 2010–2025 returns; we do not carry the earlier asset betas forward. Each cell compares the transported mixture with Fresh HML/CMA using exactly the same target assets and months. Panel B therefore tests whether a factor definition selected earlier remains useful later, not whether earlier betas forecast future returns.

Even so, changing mixtures could explain why the earlier selections perform poorly later. The final two rows therefore give the common-factor explanation another favorable test by choosing Best Common anew using 2010–2025 and evaluating it on those same observations. This later-sample hindsight buys little. Best Common improves on Fresh HML/CMA by only 0.05 percentage points for Industries, 0.02 points for JKP factors, and 0.15 points for stocks; none of these gains is statistically significant. Separate exceeds the reoptimized common model by 1.29, 2.18, and 1.34 percentage points, with all three differences significant at the 1 percent level. The common weights are unstable, but allowing them to change does not close the gap with Separate.[26]

Taken together, Panels A and B clarify the results in Section V. Fresh's weak performance does not mean that Update lacks useful information: choosing better common weights sometimes helps. But the selected weights are not stable, and even common factors chosen with full-sample hindsight do not match the fit achieved by Separate. Among the models we examine, the pricing benefit of Update comes more clearly from allowing assets to load differently on Stale and Update than from constructing one better common factor.

VII. Consequences for Empirical Inference

Factor models determine which part of a realized return is labeled abnormal (Roll, 1978). We hold realized returns fixed and change only the benchmark from Stale to Fresh or Separate. Three applications—realized monthly abnormal returns, event-study CARs, and mutual-fund performance—ask how factor timing changes the conclusions drawn from the same data.

Importantly, some disagreement is inevitable. Even if Stale truly governs expected returns, Fresh and Separate can assign different abnormal performance because the models use different factors and estimate their exposures from finite samples. Zero disagreement is therefore not a useful benchmark. Instead, we separate the observed disagreement and its ordinary sampling uncertainty from the disagreement that would arise if returns followed the fitted Stale model. By *fitted Stale*, we mean the return predicted by the Stale regression using its estimated intercept and factor exposures from the relevant return history. We treat that fitted return as the expected return

[26]The corresponding full-history exercise produces the same pattern. Common mixtures chosen before 1995 do not reliably improve on Fresh HML/CMA after 1994, and common mixtures reoptimized after 1994 still fit less well than Separate.

in the simulation, add back the Stale residuals after randomly reversing their signs in calendar blocks, and then re-estimate Stale and each challenger—Fresh or Separate—and reconstruct the reported statistic. The resulting distribution shows how much disagreement can arise from residual variation and finite-sample estimation even when Stale governs expected returns, and whether the disagreement observed in the data is unusually large by comparison.[27]

A. Factor Timing and Realized Monthly Abnormal Returns

The first application asks how much the monthly abnormal return assigned to an asset changes with the benchmark. For each asset and evaluation month t, we estimate each model using the preceding 60 calendar months through $t-1$, require at least 36 observations, and then hold the estimated intercept and loadings fixed when calculating the month-t residual. In each evaluation month, we calculate the Fresh-minus-Stale and Separate-minus-Stale residual differences for every asset. We summarize each cross section using the median absolute difference and the RMS difference. Table VIII then averages each monthly statistic over time, giving every month equal weight. The median describes a typical asset, while RMS gives more weight to large disagreements. Because the same realized month-t return enters both residuals, their difference comes entirely from what each benchmark subtracts as normal performance.

[Table 8 about here.]

[27] Every observation sharing a calendar block receives the same randomly selected residual sign, either $+1$ or -1. Thus, each simulated return equals its fitted Stale return plus a sign-randomized Stale residual. This preserves the observed residual magnitudes, common shocks across assets, and the residual pattern within each block. Monthly abnormal returns and mutual funds use synchronized 12-month blocks; event CARs use synchronized 21-trading-day blocks. This is a restricted block-wild procedure in the tradition of Wu (1986) and Shao (2010). The monthly and CAR applications use $B = 4{,}999$ draws. Including the observed statistic as a possible rank produces 5,000 rank positions and a minimum probability of 1/5,000. The mutual-fund application uses $B = 10{,}000$ to obtain finer resolution for its nonlinear classification and ranking statistics. The difference in draw counts affects only the resolution of the simulated probabilities, not the definition or interpretation of the test. Let T_{obs} denote the observed disagreement and $T^{*(b)}$ its value in draw b. The one-sided simulation probability is

$$\widehat{p} = \frac{1 + \sum_{b=1}^{B} \mathbf{1}\{T^{*(b)} \geq T_{\text{obs}}\}}{B+1}.$$

The added one in the numerator and denominator prevents a reported probability of zero (Phipson and Smyth, 2010). A 95 percent null interval is the 2.5th through 97.5th percentiles of the simulated statistics under fitted Stale; it is not a confidence interval for the observed statistic. The exercise holds the observed factor returns, samples, and other design choices fixed. It therefore provides an approximate benchmark for disagreement around the estimated Stale model, not a test that Stale or either challenger is the true model.

The disagreement is largest for individual stocks. Over the full sample, the typical absolute stock-month difference is 116 basis points for Fresh and 119 basis points for Separate; both are about 130 basis points after 1994. Typical differences for Industry portfolios and JKP strategies range from 29 to 68 basis points, while stock-level RMS disagreement exceeds two percentage points under both alternatives in both samples. Thus, benchmark timing reallocates more than one percentage point of abnormal return for a typical stock-month, with economically meaningful changes even for diversified portfolios and long–short strategies.

The fitted-Stale comparison puts these large raw differences in perspective. Every Fresh RMS and typical difference lies below its fitted-Stale mean, and none is unusually large under the one-sided fitted-Stale benchmark. By contrast, every Separate difference exceeds its fitted-Stale mean, and all twelve comparisons are significant at the 1 percent level. Both alternatives materially change the abnormal return assigned to the same asset-month, but only Separate produces disagreement that consistently exceeds the amount generated under fitted Stale. This does not establish that Separate assigns the correct abnormal return, but it does show that factor timing can reallocate measured abnormal performance by economically meaningful amounts.

B. Event-Study CARs Around Earnings and News

Event studies aggregate benchmark-adjusted returns over a short window, and factor timing can change both the magnitude and the sign of the CAR assigned to the same event (Brown and Warner, 1985). Table IX examines CARs from trading day -1 through $+20$ for fiscal-fourth-quarter earnings announcements over the full available history and, from January 2000 through November 2025, for earnings announcements and non-earnings RavenPack news days. Stock-specific intercepts and loadings are estimated over days -272 through -21, with at least 180 observations, and held fixed through the event window. All models include Official MKT–RF and MOM.

[Table 9 about here.]

Panel A shows economically large differences. The median absolute CAR difference ranges from 51 to 77 basis points, the 90th percentile ranges from 224 to 317 basis points, and RMS disagreement ranges from 174 to 237 basis points. Every RMS estimate rejects at the 1 percent level the one-sided null that population RMS disagreement is no greater than 100 basis points.

Within the common RavenPack period, the differences are larger around earnings announcements than around non-earnings news, and Fresh produces larger differences than Separate. Thus, the benchmark changes a typical 22-day CAR by roughly one-half to three-quarters of a percentage point, with upper-tail differences exceeding two percentage points.

Panel B asks whether these differences meaningfully reverse an event's interpretation. At the $\pm$25-basis-point hurdle, one model must assign a CAR of at least $+25$ basis points and the other a CAR of at most -25 basis points; the stricter $\pm$50-basis-point hurdle requires at least a 100-basis-point disagreement. Reversal rates range from 2.5 to 3.4 percent at the lower hurdle and from 1.5 to 2.2 percent at the higher hurdle. For scale, the 3.27 percent Fresh–Stale rate represents almost 1,800 of the 54,694 full-history earnings announcements, while the 3.38 percent Fresh–Stale non-earnings-news rate represents more than 100,000 stock-news days.

Figure 4 applies the fitted-Stale comparison to the twelve reversal rates in Panel B of Table IX. All twelve observed rates exceed their fitted-Stale benchmarks at the 5 percent level.[28]

[Figure 4 about here.]

Fresh produces more observed reversals, but the fitted-Stale simulations also generate relatively high Fresh–Stale reversal rates. Separate produces fewer observed reversals, yet its rates lie farther above what the simulations generate and are therefore harder to explain through finite-sample estimation and model flexibility alone. Taken together, the results show that factor timing materially changes both the magnitude and the economic interpretation of event-study CARs: for thousands of earnings announcements and news days, benchmark choice determines whether the same realized return path is classified as meaningfully positive or meaningfully negative.

[28]The common fitted-Stale procedure is described in footnote 27. The reversal-rate standard errors are two-way clustered by stock and event month (Petersen, 2009; Thompson, 2011). For each event, Stale is estimated over its estimation window. We use those estimates to simulate returns over both the estimation and event windows, then re-estimate both models and reconstruct their CARs. The reported news reversal rates use all of the more than three million qualifying news days. Because reconstructing CARs for every news day in every simulation would be computationally costly, the fitted-Stale distribution uses a random sample of 100,000 news events. Selecting alternative 100,000-event samples using different random seeds leaves every 5 percent conclusion and the interpretation unchanged. Panel A's stars use a different test: they ask whether population RMS disagreement exceeds 100 basis points.

C. *Mutual-Fund Alpha and Performance Classification*

The final application asks two questions: does factor timing change the conclusion about average active-fund performance, and does it change which fund products appear to be winners and losers? Following Carhart (1997) and Fama and French (2010), Table X studies active, diversified U.S. equity fund products under Stale, Fresh, and Separate.[29] Panel A reports alpha for equal- and TNA-weighted fund portfolios from July 2003 through December 2025 using CRSP net returns and expense-added returns.[30]

[Table 10 about here.]

Net alpha is negative under all three benchmarks for both equal- and TNA-weighted portfolios. After reported operating expenses are restored, average alpha is close to zero and statistically indistinguishable from zero under every model. Factor timing therefore leaves the familiar conclusion about average active-fund performance intact.

Fund-level classifications are less stable. At each December, five-year alpha is estimated using the same 60 monthly returns under every benchmark. Panel A reports expense-added alpha-sign changes, Panel B ranks all eligible funds together, and Panel C ranks funds within their Lipper classes.

Fresh changes the sign of 8.5 percent of expense-added alphas and replaces 15.1 percent of Stale's all-fund top decile. Separate changes 6.2 and 11.4 percent, respectively. Within style, top-decile replacement rises to 17.3 percent for Fresh and 14.7 percent for Separate. Thus, benchmark choice replaces roughly one in seven top-decile funds under Fresh and one in nine under Separate when funds are ranked together, with still larger changes when funds are compared with their natural peers.[31]

---

[29]The sensitivity of measured mutual fund performance to benchmark choice is longstanding. Lehmann and Modest (1987) compare performance and fund rankings across alternative benchmarks, Daniel et al. (1997) develop characteristic-based performance benchmarks, and Kothari and Warner (2001) show that standard performance test can have lower power when a fund's style differs from the benchmark.

[30]Expense-added returns restore one-twelfth of the reported annual operating expense ratio but remain net of trading and other costs embedded in NAV. Share classes are combined using prior-month TNA, and dead products remain through their observed histories.

[31]These classification changes can have economic consequences because investor flows respond disproportionately to top performance and because investors differ in the factor benchmarks they use to evaluate fund skill (Sirri and Tufano, 1998; Barber et al., 2016).

Figure 5 plots the fitted-Stale comparisons underlying Table X.[32]

[Figure 5 about here.]

Fresh generally changes more fund classifications than Separate, but its expense-added alpha-sign changes and all-fund top-decile replacement are not unusually large under fitted Stale, though its within-style top-decile replacement is. Separate changes fewer classifications, yet every Separate classification-change rate displayed in Figure 5 exceeds its fitted-Stale benchmark at the 1 percent level. The mutual-fund evidence therefore separates an aggregate result from a cross-sectional one. Average performance is stable, but the benchmark changes which products receive positive-alpha and top-performance labels.

Across the three applications, factor timing changes the abnormal return assigned to the same asset-month, the CAR assigned to the same event, and the performance classification assigned to the same mutual fund. The effects are economically large: monthly abnormal returns move by tens to hundreds of basis points, event CARs can reverse sign, and a meaningful share of top-ranked funds is replaced. These are not secondary presentational differences. Factor models supply the benchmark against which empirical finance defines abnormal performance. When factor timing changes the benchmark's returns and exposures, it changes the empirical conclusions researchers draw from the same realized data. Timing is therefore part of factor measurement itself.

VIII. Conclusion

Standard accounting-based factor portfolios embed a choice about when public information changes portfolio assignments. We isolate that choice by preserving the familiar Fama–French def-

[32]The common fitted-Stale procedure is described in footnote 27. In the mutual-fund application, each Stale regression includes an intercept, so the simulated return history preserves each fund's fitted Stale alpha. This is therefore not a zero-alpha or zero-skill benchmark. It asks how much sign and ranking movement factor-model choice produces when funds retain their estimated Stale performance. Table X reports two distinct kinds of inference. In Panel A's portfolio-alpha columns, parentheses contain six-lag Newey–West t-statistics and stars test whether alpha equals zero. In the alpha-sign-change column and Panels B and C, the first line is the observed classification-change rate. Parentheses contain sampling standard errors obtained by resampling the annual December formation statistics 10,000 times in circular five-year blocks (Politis and Romano, 1992), which allows for dependence created by overlapping five-year estimation windows. Square brackets report the mean classification-change rate generated by 10,000 fitted-Stale simulations. Stars compare the observed rate with the full fitted-Stale distribution; they do not use the parenthetical standard error. Thus, parentheses describe the sampling precision of the observed rate, while square brackets and stars show how that rate compares with the fitted-Stale benchmark. Bootstrap analysis of nonlinear mutual-fund classifications follows the broader approach in Kosowski et al. (2006) and Fama and French (2010).

initions while replacing the annual assignment schedule with monthly formation using the latest eligible annual statement and current market equity. Monthly updating substantially reduces information age and changes HML and CMA most, primarily by changing characteristic assignments. Yet Fresh is a weak one-for-one replacement for Stale: it does not reliably improve rolling cross-sectional fit, and replacing Stale with Fresh gives up traded content that Stale retains. The evidence is stronger for expansion through Separate: retaining Stale and allowing independent loadings on HML and CMA Update improves fit across Industry portfolios, JKP long–short strategies, and individual stocks and provides clearer evidence of incremental traded content beyond Stale.

The broader replacement view asks whether a common factor of the form $S_j + \kappa_j U_j$ can use Update more effectively than Fresh's unit weight. We give this view a favorable test by selecting the HML and CMA weights by test-asset family with full-sample hindsight. The selected mixtures sometimes improve on the Fresh HML/CMA version, but they fit less well than Separate and do not transfer reliably through time or across assets. Among the models we examine, Update's pricing benefit therefore comes more clearly from allowing assets to load differently on Stale and Update than from constructing one better common factor.

Factor timing also carries through to empirical inference. Holding realized returns fixed, changing the benchmark from Stale to Fresh or Separate changes a typical stock-month's abnormal return by more than one percentage point and a typical event-study CAR by roughly one-half to three-quarters of a percentage point. It also changes which mutual funds receive positive-alpha and top-performance labels, even though the aggregate conclusion about average active-fund performance is unchanged. The fitted-Stale comparisons show that some disagreement is mechanical and that the largest raw change need not be the most unusual.

Factor models are constructed objects. Their returns depend on the underlying data, information availability, portfolio assignments, and refresh schedule. Our results favor Separate among the models we examine, but the broader point is simpler: factor timing is a substantive choice in benchmark construction, not a secondary implementation detail. Researchers and practitioners should treat and report that choice as part of the model.

References

Akey, Pat, Adriana Z. Robertson, and Mikhail Simutin, 2026, Noisy Factors? The Retroactive Impact of Methodological Changes on the Fama–French Factors, *Review of Finance* rfag002.

Ang, Andrew, Jun Liu, and Krista Schwarz, 2020, Using stocks or portfolios in tests of factor models, *Journal of Financial and Quantitative Analysis* 55, 709–750.

Asness, Clifford S., and Andrea Frazzini, 2013, The devil in HML's details, *The Journal of Portfolio Management* 39, 49–68.

Baba-Yara, Fahiz, Martijn Boons, and Andrea Tamoni, 2024, Persistent and transitory components of firm characteristics: Implications for asset pricing, *Journal of Financial Economics* 154, 103808.

Barber, Brad M, Xing Huang, and Terrance Odean, 2016, Which factors matter to investors? evidence from mutual fund flows, *The Review of Financial Studies* 29, 2600–2642.

Barillas, Francisco, Raymond Kan, Cesare Robotti, and Jay Shanken, 2020, Model comparison with sharpe ratios, *Journal of Financial and Quantitative Analysis* 55, 1840–1874.

Barillas, Francisco, and Jay Shanken, 2017, Which alpha?, *The Review of Financial Studies* 30, 1316–1338.

Barillas, Francisco, and Jay Shanken, 2018, Comparing asset pricing models, *The Journal of Finance* 73, 715–754.

Bowles, Boone, Adam V. Reed, Matthew C. Ringgenberg, and Jacob R. Thornock, 2024, Anomaly Time, *The Journal of Finance* 79, 3543–3579.

Bowles, Boone, Adam V. Reed, Matthew C. Ringgenberg, and Jacob R. Thornock, 2026, Predicting Anomalies, *Journal of Financial Economics,* forthcoming.

Brown, Stephen J, and Jerold B Warner, 1985, Using daily stock returns: The case of event studies, *Journal of financial economics* 14, 3–31.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *The Journal of finance* 52, 1035–1058.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *The Journal of Finance* 52, 1–33.

Engelberg, Joseph E., R. David McLean, and Jeffrey Pontiff, 2018, Anomalies and news, *The Journal of Finance* 73, 1971–2001.

Engelberg, Joseph E., Adam V. Reed, and Matthew C. Ringgenberg, 2012, How are shorts informed?: Short sellers, news, and information processing, *Journal of Financial Economics* 105, 260–278.

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

Fama, Eugene F., and Kenneth R. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153–193.

Fama, Eugene F., and Kenneth R. French, 2010, Luck versus skill in the cross-section of mutual fund returns, *The Journal of Finance* 65, 1915–1947.

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

Fama, Eugene F., and Kenneth R. French, 2018, Choosing factors, *Journal of Financial Economics* 128, 234–252.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Giglio, Stefano, Dacheng Xiu, and Dake Zhang, 2025, Test assets and weak factors, *The Journal of Finance* 80, 259–319.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2021, An augmented q-factor model with expected growth, *Review of Finance* 25, 1–41.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *The Review of Financial Studies* 33, 2019–2133.

Huberman, Gur, Shmuel Kandel, and Robert F. Stambaugh, 1987, Mimicking portfolios and exact arbitrage pricing, *The Journal of Finance* 42, 1–9.

Jensen, Theis L., Bryan T. Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *The Journal of Finance* 78, 2465–2518.

Kan, Raymond, Cesare Robotti, and Jay Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *The Journal of Finance* 68, 2617–2649.

Kelly, Bryan T., Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

Keloharju, Matti, Juhani T. Linnainmaa, and Peter Nyberg, 2021, Long-term discount rates do not vary across firms, *Journal of Financial Economics* 141, 946–967.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White, 2006, Can mutual fund "stars" really pick stocks? new evidence from a bootstrap analysis, *The Journal of Finance* 61, 2551–2595.

Kothari, SP, and Jerold B Warner, 2001, Evaluating mutual fund performance, *The Journal of Finance* 56, 1985–2010.

Lehmann, Bruce N, and David M Modest, 1987, Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons, *The journal of finance* 42, 233–265.

Lettau, Martin, and Markus Pelger, 2020, Factors that fit the time series and cross-section of stock returns, *The Review of Financial Studies* 33, 2274–2325.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175–194.

McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The Journal of Finance* 71, 5–32.

Moskowitz, Tobias J., and Robert F. Stambaugh, 2021, Pricing without mispricing, Working Paper 29016, National Bureau of Economic Research.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Petersen, Mitchell A, 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *The Review of financial studies* 22, 435–480.

Phipson, Belinda, and Gordon K. Smyth, 2010, Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn, *Statistical Applications in Genetics and Molecular Biology* 9, Article 39.

Politis, Dimitris N., and Joseph P. Romano, 1992, A circular block-resampling procedure for stationary data, in Raoul LePage, and Lynne Billard, eds., *Exploring the Limits of Bootstrap*, 263–270 (John Wiley, New York).

Roll, Richard, 1978, Ambiguity when performance is measured by the securities market line, *The Journal of finance* 33, 1051–1069.

Savor, Pavel, and Mungo Wilson, 2014, Asset pricing: A tale of two days, *Journal of Financial Economics* 113, 171–201.

Shanken, Jay, 1987, Multivariate proxies and asset pricing relations: Living with the roll critique, *Journal of Financial Economics* 18, 91–110.

Shao, Xiaofeng, 2010, The dependent wild bootstrap, *Journal of the American Statistical Association* 105, 218–235.

Sirri, Erik R, and Peter Tufano, 1998, Costly search and mutual fund flows, *The journal of finance* 53, 1589–1622.

Thompson, Samuel B, 2011, Simple formulas for standard errors that cluster by both firm and time, *Journal of financial Economics* 99, 1–10.

Wu, C. F. J., 1986, Jackknife, bootstrap and other resampling methods in regression analysis, *The Annals of Statistics* 14, 1261–1295.

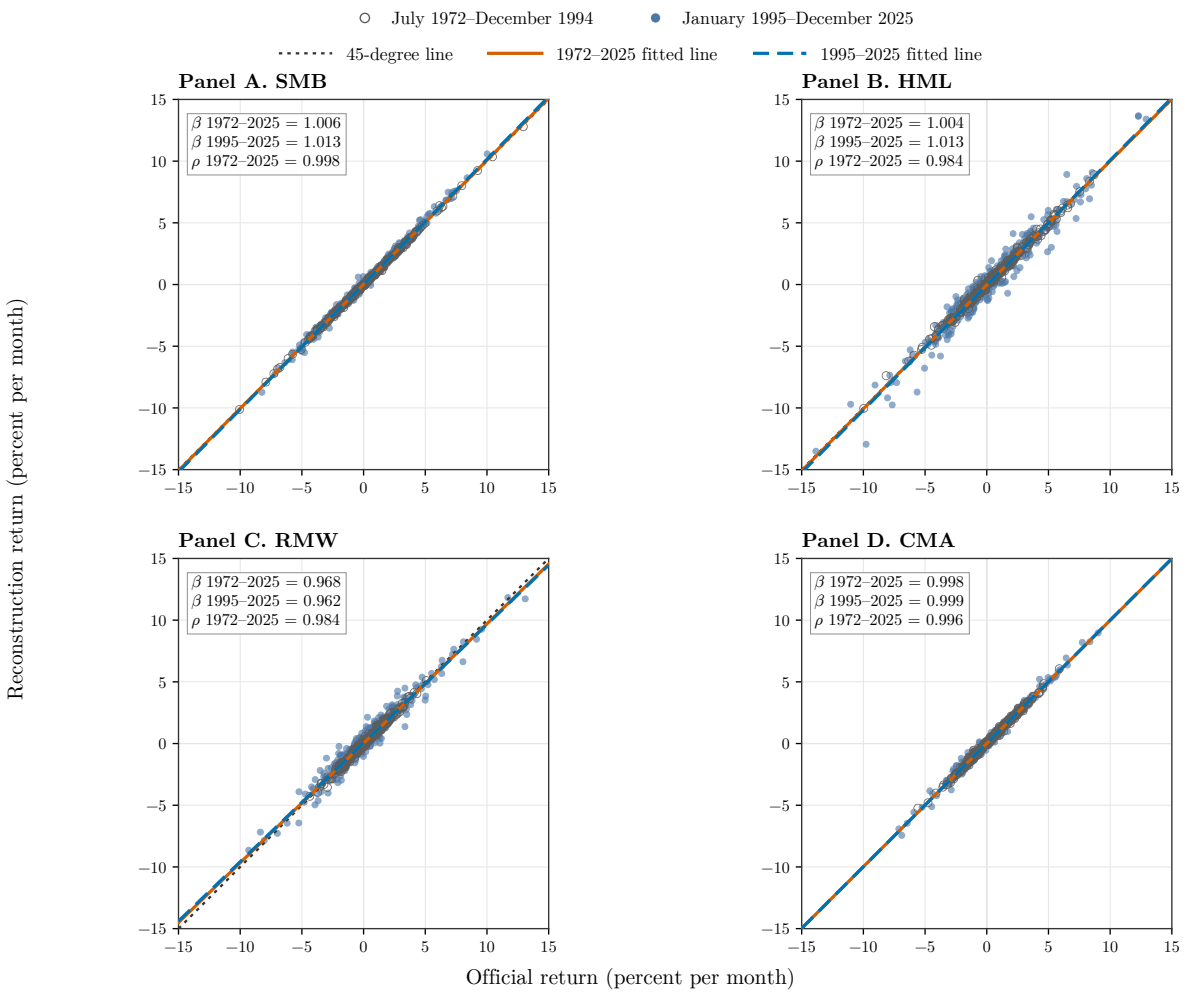


Figure 1. Reconstruction versus Official Fama–French Factor Returns

The figure compares monthly returns from our annual-assignment reconstructions (Reconstruction) with the corresponding published factor returns from the Kenneth French Data Library (Official). Panels A–D show SMB, HML, RMW, and CMA, respectively. Each circle represents one factor-month; the Official return is on the horizontal axis and the Reconstruction return is on the vertical axis, both in percent per month. Open gray circles denote July 1972–December 1994, while filled blue circles denote January 1995–December 2025. The dotted 45-degree line marks exact equality. The solid orange line is the OLS fit of Reconstruction on Official over July 1972–December 2025, and the dashed blue line is the corresponding fit over January 1995–December 2025. Each regression includes an intercept. In each annotation, β is the fitted slope for the indicated sample and ρ is the sample correlation over July 1972–December 2025. All panels use axes from -15 to 15 percent. Three observations lie outside the plotting range but remain in both regressions and the reported statistics: SMB in February and March 2000 and RMW in February 2000.

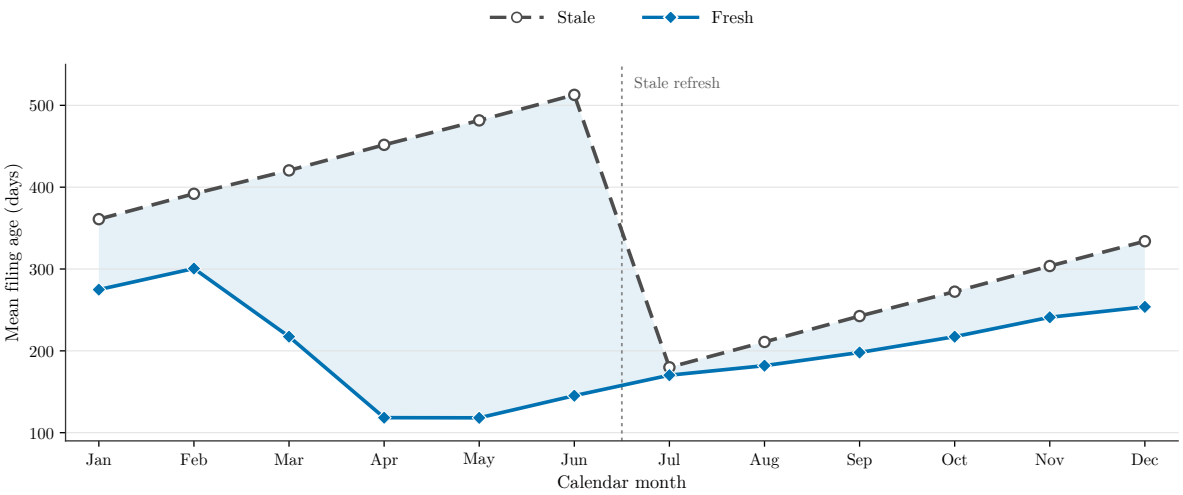


Figure 2. Information Age through the Calendar Year

The figure plots mean filing age by calendar month from January 1995 through December 2025 using the information-age measure and aggregation described in Table II. Each point averages the corresponding calendar month across 31 years, and the lines connect those monthly means. The horizontal axis runs from January through December, and the vertical axis reports mean filing age in days and ranges from 90 to 550 days. The dark gray dashed line with open circles denotes Stale, and the blue solid line with filled diamonds denotes Fresh. The vertical dotted line between June and July marks the June 30 Stale refresh, whose new assignments apply to July returns. Light shading is the month-by-month Stale-minus-Fresh age gap.

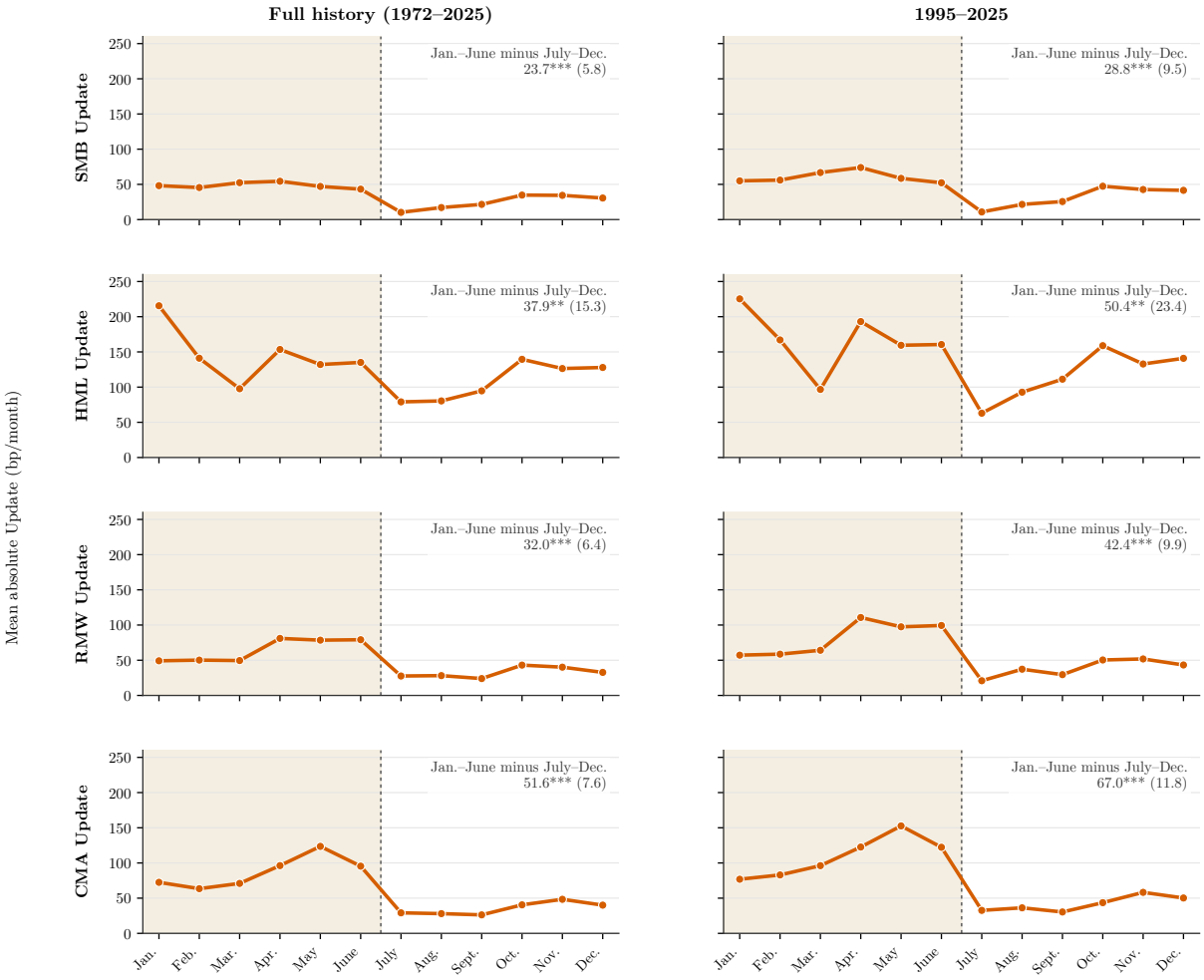


Figure 3. Update Return Magnitudes through the Calendar Year

The figure plots calendar-month means of the absolute SMB, HML, RMW, and CMA Update returns. Update is $U = F - S$, so its absolute value measures the magnitude of the Fresh-minus-Stale return difference without regard to sign. Rows identify factors, and columns cover July 1972–December 2025 (642 months) and January 1995–December 2025 (372 months), respectively. Within each panel, each filled orange circle averages absolute Update for the corresponding calendar month across all available years in the sample, and the solid orange line connects those 12 means. The horizontal axis runs from January through December, and the vertical axis reports mean absolute Update in basis points per calendar month; all eight panels use the same scale. Light shading marks January–June. The vertical dotted line between June and July marks the June 30 Stale refresh, whose new assignments apply to July returns. Each panel reports the January–June minus July–December difference in mean absolute Update, with its Newey–West standard error using 6 monthly lags in parentheses. Stars use two-sided normal-reference tests of whether the corresponding difference is zero; ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Because the figure uses absolute Update, it measures how far Fresh and Stale returns differ, not which return is higher or whether either series earns a seasonal premium.

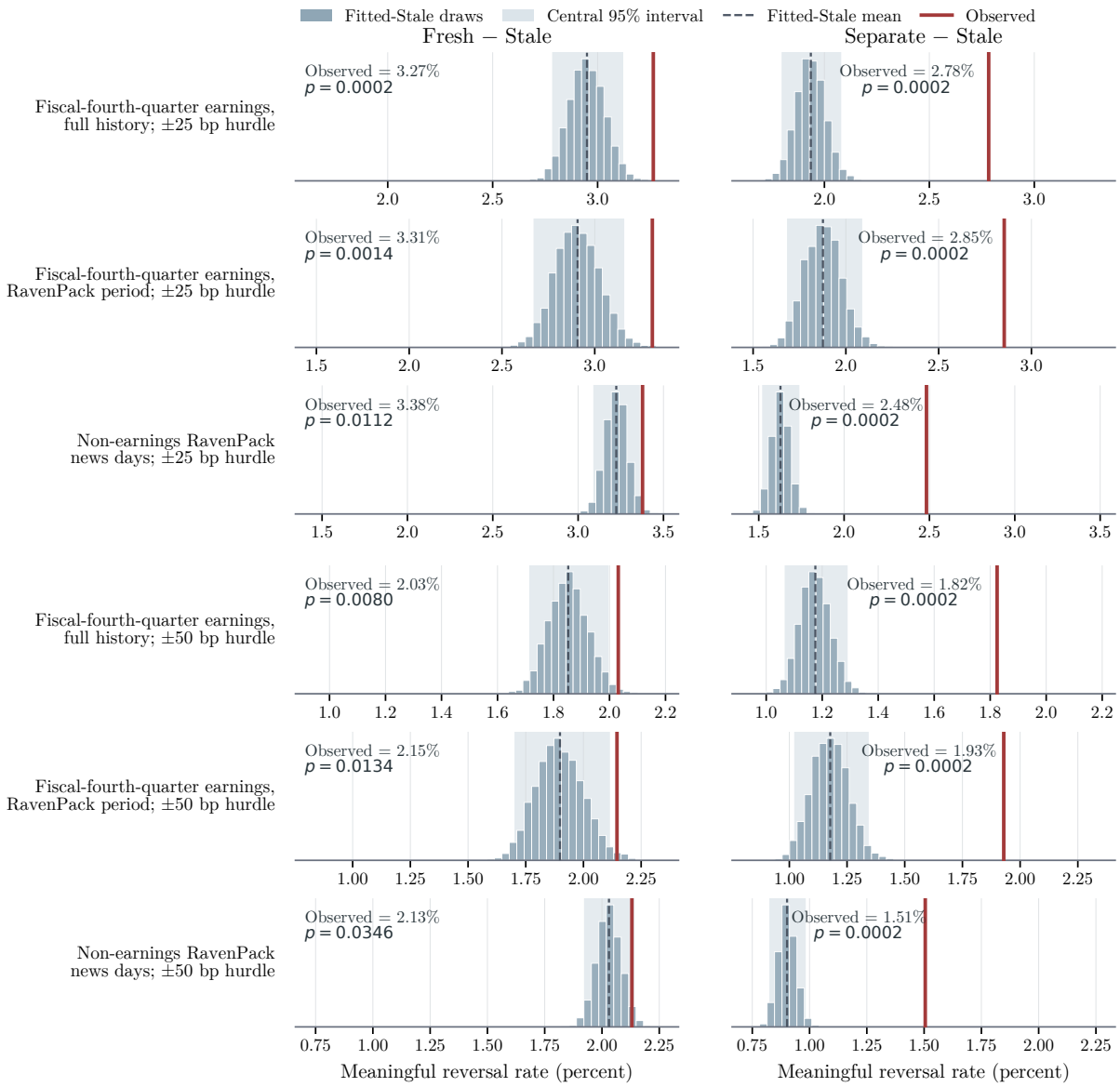


Figure 4. Meaningful CAR Reversal Rates and the Fitted-Stale Benchmark

The figure plots the fitted-Stale distributions for the twelve meaningful CAR reversal rates in Panel B of Table IX. A reversal occurs when one model assigns CAR$(-1,+20)$ at or above the positive hurdle and the other assigns it at or below the negative hurdle. Columns compare Fresh with Stale and Separate with Stale. Rows cover full-history fiscal-fourth-quarter earnings, RavenPack-period fiscal-fourth-quarter earnings, and non-earnings U.S.-company RavenPack news, first at the ±25-basis-point hurdle and then at the ±50-basis-point hurdle. Full-history earnings run from January 1974 through April 2025; the RavenPack samples run from January 2000 through November 2025. Each distribution uses 4,999 draws from the fitted-Stale procedure in footnote 27. All earnings events and a fixed sample of 100,000 news events enter the benchmark; observed rates use all qualifying events. Bars show the fitted-Stale draws; light shading marks their central 95% interval, dashed gray lines their mean, and solid red lines the observed rate. Each p-value is the one-sided probability of a fitted-Stale rate at least as large as observed. Horizontal scales are common within rows but vary across rows.

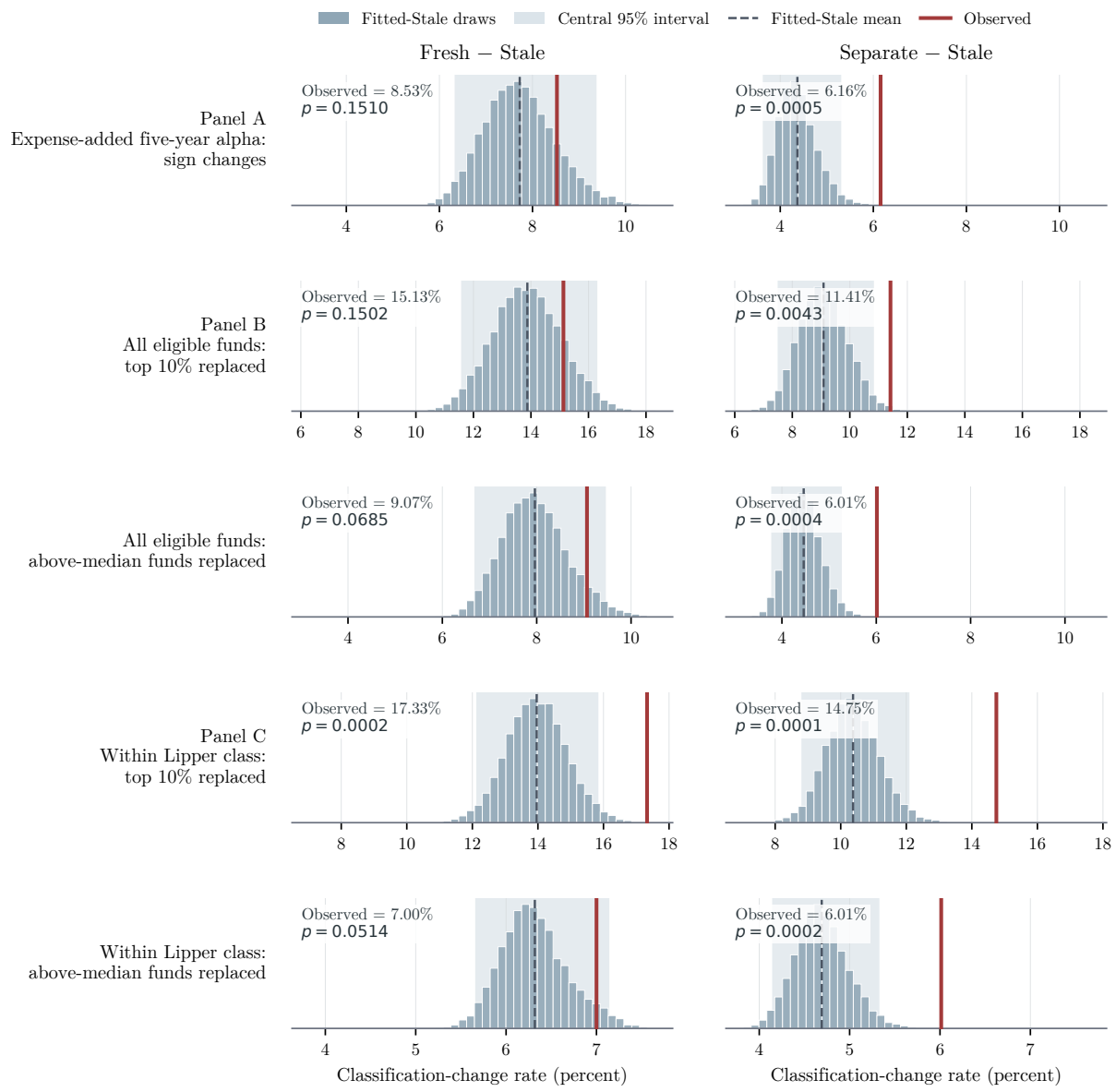


Figure 5. Mutual-Fund Classification Changes and the Fitted-Stale Benchmark

The figure plots the fitted-Stale distributions for the mutual-fund classification changes highlighted in Table X. Columns compare Fresh with Stale and Separate with Stale. Panel A reports the percentage of funds whose expense-added trailing five-year alpha changes sign, combining both directions across 17 December formations from 2008–2024. Panels B and C report top-decile and above-median replacement rates using alpha estimated from exactly 60 monthly net returns at 18 December formations from 2008–2025. Panel B ranks all eligible funds together; Panel C ranks funds within their Lipper class at formation. Each distribution uses 10,000 draws from the fitted-Stale procedure in footnote 27. Bars show the fitted-Stale draws; light shading marks their central 95% interval, dashed gray lines their mean, and solid red lines the observed rate. Each p-value is the one-sided probability of a fitted-Stale rate at least as large as observed. Horizontal scales are common within rows but vary across rows.

Table I. Replication of the Official Fama–French Factors

The table compares our annual-assignment reconstructions of SMB (small minus big), HML (high minus low), RMW (robust minus weak), and CMA (conservative minus aggressive) with the corresponding published factor returns from the Kenneth French Data Library. For this validation, we call our series Reconstruction and the published series Official. Panel A uses monthly returns and Panel B uses daily returns. Within each metric, the left and right columns cover July 1972–December 2025 (642 observations in Panel A and 13,487 in Panel B) and January 1995–December 2025 (372 and 7,802 observations), respectively. Correlation is the sample correlation between Reconstruction and Official. Variation retained is $100 \times [1 - (\text{RMSE}/\text{SD}_{\text{Official}})^2]$, where RMSE is the root mean squared Reconstruction-minus-Official return difference and $\text{SD}_{\text{Official}}$ is the standard deviation of the corresponding Official returns; 100 denotes exact equality, and higher values denote a closer match. Mean difference is the average Reconstruction-minus-Official return, in basis points per calendar month in Panel A and per trading day in Panel B. Parentheses beneath mean differences contain standard errors in the same units that allow for changing volatility and serial dependence in the return differences; the Newey–West adjustment uses 6 monthly lags in Panel A and 21 daily lags in Panel B. Stars test whether the mean difference is zero; ***, **, and * denote two-sided significance at the 1%, 5%, and 10% levels, respectively.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Monthly returns.						
	Correlation		Variation retained in %		Mean difference in bps	
Factor	1972–2025	1995–2025	1972–2025	1995–2025	1972–2025	1995–2025
SMB	0.998	0.998	99.7	99.6	1.16*	1.76*
					(0.67)	(1.06)
HML	0.984	0.979	96.7	95.5	-3.04	-5.27
					(2.09)	(3.46)
RMW	0.984	0.981	96.7	96.3	1.20	2.41
					(1.50)	(2.44)
CMA	0.996	0.995	99.2	99.1	0.01	0.62
					(0.68)	(1.02)
Panel B. Daily returns.						
	Correlation		Variation retained in %		Mean difference in bps	
Factor	1972–2025	1995–2025	1972–2025	1995–2025	1972–2025	1995–2025
SMB	0.998	0.997	99.5	99.4	0.06*	0.08
					(0.03)	(0.05)
HML	0.978	0.975	95.5	95.0	-0.17	-0.28
					(0.11)	(0.18)
RMW	0.975	0.974	95.1	94.8	0.05	0.10
					(0.08)	(0.13)
CMA	0.994	0.994	98.9	98.9	-0.01	0.04
					(0.03)	(0.05)

Table II. Information Age in Stale and Fresh Factor Portfolios

The table compares the age of accounting information used in Stale and Fresh factor portfolios. Stale assigns firms each June for the following July–June portfolio year. Fresh updates assignments monthly using the latest annual filing public before formation; updated assignments apply only to subsequent returns. Filing age is the number of days from an annual observation's availability date to the middle of the return month. Within each signal and month, filing age and the indicator that age exceeds 365 days are weighted by lagged market equity. Book-to-market, operating profitability, and investment are then averaged equally within month, and months are averaged equally within each sample. The full-history and 1972–1994 rows begin in July 1972; the other rows begin in January, and every row ends in December of its final year. Reduction in days is Stale minus Fresh, and Percent is $100 \times (\text{Stale} - \text{Fresh})/\text{Stale}$. These reductions are computed from unrounded ages and therefore need not equal differences of the displayed rounded means. The last two columns report the percentage of market capitalization assigned using a filing more than 365 days old.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean filing age in days		Reduction in filing age		Capital using a filing older than one year (%)	
Sample	Stale	Fresh	Days	Percent	Stale	Fresh
Full history (1972–2025)	319	204	114	35.8	35.8	5.8
1972–1994	280	206	73	26.2	25.7	6.6
1995–2025	347	203	144	41.5	43.2	5.3
2010–2025	368	200	168	45.6	48.4	4.3

Table III. Properties of Stale, Fresh, and Update Factor Returns

The table compares monthly Stale and Fresh returns for SMB, HML, RMW, and CMA and reports properties of Update, defined as $U = F - S$. Stale assigns firms each June for the following July–June portfolio year. Fresh updates assignments monthly using the latest annual filing public before formation; updated assignments apply only to subsequent returns. Panel A covers July 1972–December 2025 (642 months), and Panel B covers January 1995–December 2025 (372 months). Correlation is the sample correlation between Stale and Fresh returns. Update mean is the average Fresh-minus-Stale return, and Update volatility is its sample standard deviation; both are in basis points per calendar month. Alpha is the intercept, in basis points per calendar month, from a time-series regression of Update on Official MKT–RF, Official MOM, and Stale SMB, HML, RMW, and CMA. MOM loading is the fitted coefficient on Official MOM in the same regression and is unitless. Parentheses beneath Update means, alphas, and MOM loadings contain Newey–West standard errors in the same units using 6 monthly lags. Stars use individual two-sided normal-reference tests of whether the corresponding mean, alpha, or MOM loading is zero; ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Monthly returns, July 1972–December 2025.					
	Stale–Fresh correlation	Update return		FF5 plus Official MOM regression	
Factor		Mean (bp)	Volatility (bp)	Alpha (bp)	MOM loading
SMB	0.974	0.99 (2.25)	69.9	4.88** (2.48)	-0.083*** (0.014)
HML	0.828	0.81 (8.16)	213.2	28.40*** (5.44)	-0.430*** (0.024)
RMW	0.948	5.06* (2.83)	78.3	1.83 (2.95)	0.003 (0.018)
CMA	0.890	10.95** (4.44)	101.0	16.28*** (5.16)	-0.040 (0.025)
Panel B. Monthly returns, January 1995–December 2025.					
	Stale–Fresh correlation	Update return		FF5 plus Official MOM regression	
Factor		Mean (bp)	Volatility (bp)	Alpha (bp)	MOM loading
SMB	0.962	2.09 (3.68)	88.0	3.32 (3.18)	-0.102*** (0.015)
HML	0.813	-1.84 (12.34)	245.4	13.67* (7.13)	-0.452*** (0.028)
RMW	0.947	5.45 (4.55)	95.4	5.64 (4.40)	-0.004 (0.025)
CMA	0.865	14.35** (7.18)	123.9	21.21*** (7.16)	-0.052 (0.036)

Table IV. Sources of the Update Return

The table describes the portfolio revisions that produce monthly Update returns from July 1972 through December 2025 (642 months). Stale assigns firms each June for the following July–June portfolio year. Fresh updates assignments monthly using the latest annual filing public before formation; updated assignments apply only to subsequent returns. Update is $U = F - S$. Panel A partitions the HML, RMW, and CMA Update returns into two mutually exclusive portfolio-revision groups. Characteristic reassignment means that a firm's factor-specific characteristic bucket changes, whether or not its size bucket also changes. Other portfolio revisions combine size-only reassignment, changes in weights among firms that remain in the same portfolio cell, and eligibility changes in which a firm is assigned under only one design. Panel A excludes SMB because it combines size legs across the book-to-market, profitability, and investment sorts and does not admit the same partition. Mean contribution is the average signed contribution of each group to Update, in basis points per calendar month; positive values raise Fresh relative to Stale and negative values lower it. Alpha is the intercept, in basis points per calendar month, from a time-series regression of each contribution on Official MKT–RF, Official MOM, and Stale SMB, HML, RMW, and CMA. Component alphas describe contribution time series, not returns on separately traded portfolios. Exposure is computed each month as the group's share of total absolute stock-level Update weights and then averaged equally across months. Within each factor, the two component mean contributions and alphas sum to the Total Update return, and their exposure shares sum to 100 percent. Panel B compares Stale and Fresh characteristic assignments for all four factors. For each month, characteristic-rank correlation is the Spearman correlation among firms eligible under both designs. The last two columns report, among those same firms, the percentage assigned to different characteristic buckets, weighting firms equally or by market capitalization. Panel B averages each monthly statistic equally across the 642 months. The relevant characteristic is size for SMB, book-to-market for HML, operating profitability for RMW, and investment for CMA.

(1)	(2)	(3)	(4)	(5)
Panel A. Sources of HML, RMW, and CMA Update returns.				
Factor	Portfolio-revision group	Mean contribution (bp/mo)	Alpha (bp/mo)	Exposure (%)
HML	Characteristic reassignment	-11.13	21.37	69.3
	Other portfolio revisions	11.94	7.03	30.7
	Total Update return	0.81	28.40	100.0
RMW	Characteristic reassignment	-0.45	-2.26	47.4
	Other portfolio revisions	5.51	4.08	52.6
	Total Update return	5.06	1.83	100.0
CMA	Characteristic reassignment	9.54	16.37	69.1
	Other portfolio revisions	1.41	-0.09	30.9
	Total Update return	10.95	16.28	100.0

(1)	(2)	(3)	(4)
Panel B. Stale and Fresh characteristic assignments.			
Factor	Characteristic-rank correlation	Different characteristic bucket (% of firms)	Different characteristic bucket (% of market cap)
SMB	0.983	2.3	1.3
HML	0.828	27.0	16.4
RMW	0.925	10.1	8.2
CMA	0.748	20.0	17.5

Table V. Time-Series and Cross-Sectional Fit of Stale, Fresh, and Separate Factor Models

The table compares three monthly factor models. Stale includes Official MKT–RF, matched-frequency Official MOM, and Stale SMB, HML, RMW, and CMA. Fresh replaces all four Stale characteristic factors with their Fresh counterparts; Separate retains Stale and adds HML and CMA Update returns. For each evaluation month t, model-specific first-stage time-series regressions estimate each asset's factor exposures using the preceding 60 calendar months through $t-1$ and require at least 36 usable observations. Panel A reports the adjusted R^2 from those regressions, averaged first across assets within each month and then equally across months. Entries are percentages and describe time-series covariance fit, not cross-sectional pricing; no Panel A inference is displayed. Panel B regresses month-t excess returns cross-sectionally on the predetermined exposures. Columns 2–4 report time-series averages of monthly cross-sectional adjusted R^2, in percent. Columns 5–7 report time-series means of the paired monthly differences Fresh minus Stale, Separate minus Stale, and Separate minus Fresh, in percentage points. All adjusted R^2 measures account for model dimension, including Separate's two additional factor dimensions. Parentheses contain Newey–West standard errors in percentage points using 6 monthly lags. Stars test whether the corresponding mean difference is zero; ***, **, and * denote two-sided significance at the 1%, 5%, and 10% levels, respectively. Evaluation samples are July 1977–December 2025 (582 months) and January 1995–December 2025 (372 months); pre-1995 returns may enter the later sample's initial estimation windows. The stock panel is dynamic, but all three models use identical stocks within each month. Panel B uses exposures estimated before month t to explain returns realized in month t; it is not a next-month return forecast. JKP denotes Jensen, Kelly, and Pedersen.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. First-stage time-series fit.						
	July 1977–December 2025			January 1995–December 2025		
Test assets	Stale	Fresh	Separate	Stale	Fresh	Separate
49 Industry portfolios	65.1	64.9	65.6	60.7	60.5	61.2
153 JKP factors	56.9	58.3	59.5	56.8	58.2	59.2
Dynamic CRSP stocks	33.1	33.0	33.8	31.1	31.0	31.8
Panel B. Month-t cross-sectional fit.						
	Adjusted R^2 (%)			Difference (pp)		
Test assets and sample	Stale	Fresh	Separate	Fresh − Stale	Separate − Stale	Separate − Fresh
49 Industry portfolios						
1977–2025 sample	25.3	25.3	28.1	0.06	2.85***	2.79***
				(0.21)	(0.25)	(0.27)
1995–2025 sample	24.8	25.0	28.0	0.12	3.14***	3.02***
				(0.29)	(0.30)	(0.35)
153 JKP factors						
1977–2025 sample	65.2	65.3	68.4	0.11	3.21***	3.10***
				(0.19)	(0.18)	(0.21)
1995–2025 sample	65.8	65.9	69.2	0.07	3.33***	3.26***
				(0.26)	(0.24)	(0.27)
Dynamic CRSP stocks						
1977–2025 sample	7.6	7.5	8.0	-0.04	0.45***	0.49***
				(0.03)	(0.03)	(0.04)
1995–2025 sample	7.9	7.9	8.4	-0.05	0.47***	0.52***
				(0.05)	(0.04)	(0.05)

Table VI. Traded Factor-Model Comparisons

The table compares three models of monthly traded factor returns. Stale includes Official MKT–RF, matched-frequency Official MOM, and Stale SMB, HML, RMW, and CMA. Fresh replaces the four Stale characteristic factors with their Fresh counterparts. Separate retains Stale and adds HML and CMA Update returns, where Update is Fresh minus Stale. Stale and Fresh therefore contain six factors, and Separate contains eight. Panel A reports annualized maximum Sharpe and monthly adjusted θ^2. Monthly adjusted θ^2 is the bias-adjusted monthly maximum squared Sharpe ratio; annualized maximum Sharpe is its annualized equivalent, $(12\,\theta^2_{\text{adj}})^{1/2}$. The adjustment accounts for model dimension. The Change in adjusted θ^2 versus Stale columns report Fresh-minus-Stale and Separate-minus-Stale differences in monthly adjusted θ^2; positive values favor the model named in the row. Stars for Separate versus Stale use the exact finite-sample Gibbons–Ross–Shanken (GRS) test. Stars for the nonnested Fresh-versus-Stale comparison use the sequential Barillas–Kan–Robotti–Shanken test with 6 monthly lags in the Sharpe-difference standard error. Panel B reports exact joint GRS spanning tests. RMS alpha is the root mean square of the conditional alphas for the factor returns added by the first model in each row, in basis points per calendar month. A low p-value rejects the joint null that those returns add nothing beyond the conditioning model. Fresh beyond Stale and Stale beyond Fresh each test four characteristic-factor directions; Separate beyond Stale tests HML and CMA Update returns. Separate beyond Fresh and Stale beyond Fresh are algebraically equivalent. Panel C reports posterior probabilities, in percent, that the tested factor set is jointly nonredundant under annual-alpha prior standard deviations of 50 and 100 basis points. The 100-basis-point null is more permissive. The calculation uses equal prior odds and the Barillas–Shanken alternative prior calibrated to 1.5 times the benchmark maximum Sharpe ratio. These probabilities are not p-values; 50 percent denotes equal posterior odds, and higher values favor nonredundancy. The samples are July 1972–December 2025 (642 months) and January 1995–December 2025 (372 months). ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

(1)	(2)	(3)	(4)	(5)	(6)	(7)

Panel A. Dimension-adjusted opportunity sets and formal comparisons.

	July 1972–December 2025			January 1995–December 2025		
Model	Annualized maximum Sharpe	Monthly adjusted θ^2	Change in adjusted θ^2 vs. Stale	Annualized maximum Sharpe	Monthly adjusted θ^2	Change in adjusted θ^2 vs. Stale
Stale	1.11	0.102		1.10	0.100	
Fresh	1.27	0.134	0.032**	1.19	0.117	0.017
Separate	1.48	0.183	0.081***	1.27	0.135	0.035***

(1)	(2)	(3)	(4)	(5)

Panel B. Exact directional spanning.

	July 1972–December 2025		January 1995–December 2025	
Direction	RMS alpha (bp/mo)	p-value	RMS alpha (bp/mo)	p-value
Fresh adds beyond Stale	16.6	<0.0001	13.0	0.0016
Stale adds beyond Fresh	13.8	<0.0001	11.9	0.0179
Separate adds beyond Stale	23.1	<0.0001	17.8	0.0013
Separate adds beyond Fresh	13.8	<0.0001	11.9	0.0179

Panel C. Bayesian posterior probability of joint nonredundancy (%).

	July 1972–December 2025		January 1995–December 2025	
Direction	50 bp null	100 bp null	50 bp null	100 bp null
Fresh adds beyond Stale	99.8	65.8	69.6	33.2
Stale adds beyond Fresh	88.7	32.9	26.7	17.4
Separate adds beyond Stale	100.0	91.4	84.8	52.4

Table VII. Common HML and CMA Mixtures versus Separate Exposures

The table compares Best Common HML/CMA mixtures with Separate using the models and cross-sectional fit procedure described in the text. Fresh HML/CMA is the focused benchmark: it updates only HML and CMA while retaining Stale SMB and RMW, rather than replacing all four characteristic factors. Best Common selects one pair of HML and CMA Update weights for each test-asset family and applies it to all assets in that family; Separate permits asset-specific loadings on Stale and Update HML and CMA. Panel A uses January 1995–December 2025. Columns (2)–(5) report average monthly cross-sectional adjusted R^2 in percent for Stale, Fresh HML/CMA, Best Common, and Separate. Columns (6)–(7) report paired differences in percentage points. Panel B assesses transport. Its first three rows identify the source family used to select the common mixture during January 1995–December 2009 and report that mixture's adjusted-R^2 difference from Fresh HML/CMA for each target family during January 2010–December 2025. Only the selected mixture weights, (κ_H, κ_C), are transported; target-family factor exposures are re-estimated in the evaluation sample. The coefficient row gives mixtures reselected within the evaluation period for each target family. The final two rows compare those reselected mixtures with Fresh HML/CMA and Separate. Adjusted R^2 penalizes model dimension. Parentheses contain paired Newey–West standard errors using 6 monthly lags. Stars test whether each mean difference is zero; ***, **, and * denote two-sided significance at the 1%, 5%, and 10% levels. Because Best Common coefficients are selected and evaluated on the same observations in Panel A and in Panel B's final two rows, inference for those comparisons treats the coefficients as fixed and does not account for the search.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Best Common mix within each test-asset family, January 1995–December 2025.						
	Adjusted R^2 (%)				Difference (pp)	
Test assets	Stale	Fresh HML/CMA	Best Common	Separate	Best Common − Fresh HML/CMA	Separate − Best Common
49 Industry portfolios	23.2	23.8	25.7	27.0	1.92*** (0.36)	1.29*** (0.24)
153 JKP factors	67.0	67.0	67.2	69.6	0.21** (0.09)	2.42*** (0.18)
Dynamic CRSP stocks	11.7	11.5	11.7	12.8	0.24*** (0.07)	1.12*** (0.15)

(1)	(2)	(3)	(4)
Panel B. Common-mixture transport through time and across test assets.			
Mix selected using source assets January 1995–December 2009	Evaluation sample: January 2010–December 2025 First three rows: transported mixture − Fresh HML/CMA (pp)		
Source assets; selected (κ_H, κ_C)	49 Industry portfolios	153 JKP factors	Dynamic CRSP stocks
49 Industry portfolios			
Selected $(2.35, -9.56)$	-1.78*** (0.42)	-1.75*** (0.54)	-0.16 (0.23)
153 JKP factors			
Selected $(0.23, 0.16)$	-0.57*** (0.20)	-0.15 (0.17)	0.09 (0.11)
Dynamic CRSP stocks			
Selected $(-0.23, -0.35)$	-1.11*** (0.31)	-0.57* (0.29)	-0.01 (0.15)
Best Common coefficients selected in January 2010–December 2025	$(0.26, 1.10)$	$(0.58, 0.78)$	$(1.00, 0.08)$
Best Common − Fresh HML/CMA in evaluation sample	0.05 (0.09)	0.02 (0.04)	0.15 (0.12)
Separate − Best Common in evaluation sample	1.29*** (0.43)	2.18*** (0.22)	1.34*** (0.23)

Table VIII. Factor Timing and Differences in Realized Monthly Abnormal Returns

The table measures how changing the factor benchmark changes the realized abnormal return assigned to the same asset-month. For each asset and evaluation month t, each model is estimated over the preceding 60 calendar months through $t-1$, requiring at least 36 observations; the estimated intercept and loadings then determine the month-t abnormal return. Stale uses annually assigned SMB, HML, RMW, and CMA; Fresh replaces all four with their point-in-time versions; Separate retains Stale and adds HML and CMA Update returns. All models include Official MKT–RF and matched-frequency Official MOM and use identical asset-months. Panel A covers July 1977–December 2025, and Panel B covers January 1995–December 2025; pre-1995 observations may enter Panel B's initial estimation windows. Columns (2)–(3) report evaluation months and the average number of assets per month. Within each month, columns (4) and (6) compute the cross-sectional root mean squared and median absolute challenger-minus-Stale abnormal-return differences. The Observed entries average those monthly statistics equally over time and are in basis points per month. Parentheses contain Newey–West standard errors using 6 monthly lags. Columns (5) and (7) report the mean and 2.5th–97.5th percentile interval from the fitted-Stale benchmark described in footnote 27. In each of 4,999 draws, fitted Stale returns, including intercepts, are combined with synchronized nonoverlapping 12-month block-wild residual signs, and both models are re-estimated. Stars compare the Observed value with the full fitted-Stale distribution; they do not test against zero or use the parenthetical standard error. ***, **, and * denote one-sided significance at the 1%, 5%, and 10% levels, respectively. Dynamic CRSP stocks preserve entry and exit, with all models evaluated on identical monthly support.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Full evaluation sample: July 1977–December 2025.						
			RMS difference (bp/month)		Median absolute difference (bp/month)	
Test assets and model change	No. of months	Avg. assets per month	Observed	Fitted-Stale mean [95% null interval]	Observed	Fitted-Stale mean [95% null interval]
49 Industry portfolios						
Fresh − Stale	582	49	98.01	105.46	59.68	63.14
			(5.38)	[101.55, 109.62]	(3.58)	[60.89, 65.50]
Separate − Stale	582	49	93.74***	84.30	55.09***	48.20
			(5.52)	[80.50, 88.11]	(3.39)	[46.09, 50.37]
153 JKP factors						
Fresh − Stale	582	153	57.50	57.78	37.29	37.92
			(3.92)	[56.23, 59.32]	(2.45)	[36.89, 38.93]
Separate − Stale	582	153	44.60***	33.78	28.84***	21.52
			(2.68)	[31.09, 36.86]	(1.70)	[19.54, 23.70]
Dynamic CRSP stocks						
Fresh − Stale	582	1,466	203.54	220.96	116.07	125.46
			(11.45)	[215.26, 226.83]	(6.28)	[122.08, 129.08]
Separate − Stale	582	1,466	214.61***	196.69	118.90***	108.01
			(12.84)	[192.51, 201.56]	(7.19)	[105.54, 110.82]
Panel B. Later evaluation sample: January 1995–December 2025.						
			RMS difference (bp/month)		Median absolute difference (bp/month)	
Test assets and model change	No. of months	Avg. assets per month	Observed	Fitted-Stale mean [95% null interval]	Observed	Fitted-Stale mean [95% null interval]
49 Industry portfolios						
Fresh − Stale	372	49	111.47	120.03	68.16	71.52
			(7.63)	[114.41, 126.17]	(5.13)	[68.19, 74.96]
Separate − Stale	372	49	103.66***	93.03	59.47***	53.26
			(7.92)	[88.36, 98.08]	(4.93)	[50.77, 55.91]
153 JKP factors						
Fresh − Stale	372	153	65.76	66.82	42.39	43.59
			(5.74)	[64.72, 68.94]	(3.58)	[42.21, 45.03]
Separate − Stale	372	153	47.51***	37.86	30.40***	23.89
			(3.83)	[34.44, 41.90]	(2.39)	[21.36, 26.67]
Dynamic CRSP stocks						
Fresh − Stale	372	1,509	233.35	252.91	130.46	140.97
			(16.21)	[244.67, 261.58]	(9.03)	[136.05, 146.19]
Separate − Stale	372	1,509	240.72***	220.76	130.06***	118.56
			(18.54)	[215.10, 227.80]	(10.54)	[115.37, 122.31]

Table IX. Factor Timing and Differences in Event-Study CARs Around Earnings and News

The table reports how Fresh and Separate change the CAR assigned by Stale to the same stock-event. CAR$(-1,+20)$ sums abnormal returns over 22 trading days; day zero is the first factor-trading day strictly after the information date. Stock-specific intercepts and loadings are estimated over days -272 through -21, with at least 180 observations, then held fixed. Fresh replaces all four Stale characteristic factors; Separate adds HML and CMA Update returns to Stale. Every model includes Official MKT–RF and MOM. Full-history fiscal-fourth-quarter earnings announcements run from January 1974 through April 2025. RavenPack-period earnings and non-earnings U.S.-company news run from January 2000 through November 2025. Column (2) reports events, with distinct stocks in parentheses. Panel A reports median absolute, 90th-percentile absolute, and root mean squared challenger-minus-Stale CAR differences in basis points. Parentheses contain standard errors clustered by stock and event month. Stars in the root mean squared column test the one-sided null that population disagreement is no greater than 100 basis points. Panel B reports the percentage of events for which one model assigns a CAR at or above the positive hurdle and the other a CAR at or below the negative hurdle. Parentheses contain clustered standard errors in percentage points. Fitted-Stale columns report the mean and 2.5th–97.5th percentile interval from 4,999 draws using the procedure in footnote 27. The benchmark uses all earnings events and a fixed sample of 100,000 news events. Panel B stars compare the observed rate with the full fitted-Stale distribution, not the parenthetical standard error. ***, **, and * denote one-sided significance at the 1%, 5%, and 10% levels, respectively.

(1)	(2)	(3)	(4)	(5)
Panel A. Absolute CAR differences, CAR$(-1,+20)$.				
		Absolute difference in CAR (basis points)		
Event setting and model change	Events (stocks)	Median absolute CAR difference	90th pct. absolute CAR difference	RMS CAR difference
Fiscal-fourth-quarter earnings, full history				
Fresh − Stale	54,694	74.59	277.90	206.50***
	(2,755)	(3.69)	(21.84)	(18.51)
Separate − Stale	54,694	58.30	257.62	192.19***
	(2,755)	(3.68)	(20.69)	(15.97)
Fiscal-fourth-quarter earnings, RavenPack period				
Fresh − Stale	28,667	76.79	316.93	236.71***
	(1,742)	(6.79)	(42.74)	(27.16)
Separate − Stale	28,667	61.31	278.46	214.23***
	(1,742)	(5.50)	(36.17)	(22.78)
Non-earnings RavenPack news days				
Fresh − Stale	3,073,648	68.69	271.99	210.06***
	(2,115)	(2.26)	(11.61)	(9.94)
Separate − Stale	3,073,648	51.20	223.90	173.55***
	(2,115)	(1.84)	(9.19)	(8.39)

(1)	(2)	(3)	(4)	(5)	(6)
Panel B. Meaningful CAR sign reversals, CAR$(-1,+20)$.					
		CAR hurdle ±25 bp		CAR hurdle ±50 bp	
Event setting and model change	Events (stocks)	Observed	Fitted-Stale mean [95% null interval]	Observed	Fitted-Stale mean [95% null interval]
Fiscal-fourth-quarter earnings, full history					
Fresh − Stale	54,694	3.27***	2.95	2.03***	1.85
	(2,755)	(0.26)	[2.79, 3.12]	(0.22)	[1.72, 1.99]
Separate − Stale	54,694	2.78***	1.94	1.82***	1.18
	(2,755)	(0.24)	[1.80, 2.08]	(0.20)	[1.07, 1.29]
Fiscal-fourth-quarter earnings, RavenPack period					
Fresh − Stale	28,667	3.31***	2.91	2.15**	1.90
	(1,742)	(0.41)	[2.68, 3.15]	(0.36)	[1.71, 2.11]
Separate − Stale	28,667	2.85***	1.88	1.93***	1.18
	(1,742)	(0.35)	[1.69, 2.08]	(0.31)	[1.03, 1.34]
Non-earnings RavenPack news days					
Fresh − Stale	3,073,648	3.38**	3.22	2.13**	2.03
	(2,115)	(0.14)	[3.10, 3.35]	(0.12)	[1.93, 2.14]
Separate − Stale	3,073,648	2.48***	1.63	1.51***	0.90
	(2,115)	(0.11)	[1.53, 1.73]	(0.09)	[0.83, 0.98]

Table X. Factor Timing, Mutual-Fund Alpha, and Performance Classification

The table reports mutual-fund alpha and performance classifications. Panel A reports monthly equal- and TNA-weighted portfolio alpha in basis points from July 2003 through December 2025. Expense-added returns restore one-twelfth of the date-effective annual operating expense ratio but remain net of costs embedded in NAV. Parentheses in columns (2)–(5) contain Newey–West t-statistics using 6 monthly lags; stars use two-sided tests of whether alpha is zero. Column (6) reports the percentage of funds whose expense-added trailing five-year alpha changes sign relative to Stale, combining both directions across 17 December formations from 2008–2024. Panels B and C rank funds by alpha estimated from exactly 60 monthly net returns at 18 December formations from 2008–2025. Panel B ranks all funds together; Panel C ranks funds within their Lipper class at formation and weights class-specific replacement rates by selected-group size. Columns (2)–(4) report the percentage of each Stale-selected group replaced under Fresh or Separate. The final column records whether the highest-alpha fund changes in Panel B and averages that indicator across 13 classes in Panel C. Formation dates receive equal weight. In column (6) and Panels B–C, the first line is the observed rate, parentheses contain standard errors in percentage points from 10,000 circular block bootstraps of formation-year statistics using five-year blocks, and square brackets contain mean rates from 10,000 draws using the fitted-Stale procedure in footnote 27. Classification stars use unadjusted one-sided comparisons with the full fitted-Stale distribution, not the parenthetical standard errors. TNA denotes total net assets. ***, **, and * denote significance at the 1%, 5%, and 10% levels under the applicable test.

(1)	(2)	(3)	(4)	(5)	(6)

Panel A. Portfolio alpha and expense-added five-year alpha-sign changes.

	CRSP net returns		Expense-added returns		Expense-added five-year alpha
Model	Equal-weighted	TNA-weighted	Equal-weighted	TNA-weighted	Sign changes (%)
Stale	-13.25***	-7.82***	-3.23	-0.87	–
	(-3.50)	(-2.61)	(-0.84)	(-0.28)	
Fresh	-13.55***	-7.35**	-3.50	-0.36	8.5
	(-3.95)	(-2.45)	(-1.00)	(-0.12)	(0.91)
					[7.7]
Separate	-13.33***	-7.75***	-3.29	-0.71	6.2***
	(-3.60)	(-2.63)	(-0.87)	(-0.23)	(0.57)
					[4.4]

(1)	(2)	(3)	(4)	(5)

Panel B. All eligible funds ranked together.

	Stale-selected funds replaced (%)			
Model comparison	Top 5%	Top 10%	Above median	Highest alpha
Fresh − Stale	18.6**	15.1	9.1*	33.3
	(1.94)	(1.41)	(0.97)	(11.63)
	[15.4]	[13.9]	[8.0]	[24.9]
Separate − Stale	13.2***	11.4***	6.0***	27.8
	(1.34)	(0.98)	(0.46)	(7.85)
	[10.5]	[9.1]	[4.5]	[17.0]

(1)	(2)	(3)	(4)	(5)

Panel C. Funds ranked within their Lipper classes at formation.

	Stale-selected funds replaced (%)			
Model comparison	Top 5%	Top 10%	Above median	Highest alpha
Fresh − Stale	18.1**	17.3***	7.0*	24.8*
	(2.41)	(1.90)	(0.93)	(2.85)
	[15.8]	[14.0]	[6.3]	[20.2]
Separate − Stale	15.8***	14.7***	6.0***	22.2***
	(1.50)	(1.40)	(0.54)	(2.99)
	[12.4]	[10.4]	[4.7]	[15.4]

Internet Appendix for "Factor Time"

This Internet Appendix documents the paper's data, factor-construction, information-timing, and sample procedures.[1]

A. Data Sources and Factor Construction

Factor construction combines monthly and daily CRSP stock returns and market equity with annual Compustat fundamentals. We follow the familiar Fama–French investable-universe convention and retain U.S. common stocks listed on the NYSE, AMEX, or Nasdaq (share codes 10 and 11 and exchange codes 1, 2, and 3). We use CRSP holding-period returns and set missing returns to zero. Market equity is measured at the company level, and CRSP firms are linked to annual Compustat through date-valid CCM links.[2] Eligible NYSE firms determine the breakpoints applied to the full eligible universe. We use the NYSE median for size and the 30th and 70th percentiles for book-to-market, operating profitability, and investment.

We construct book equity, operating profitability, and investment from annual Compustat records using the definitions below. Book-to-market pairs book equity with the market-equity observation specified by the applicable formation rule. Stale and Fresh differ in when these measures enter portfolio formation, not in how the measures themselves are defined.

Accounting measure	Definition and eligibility
Book equity (BE)	Stockholders' equity plus deferred taxes and investment tax credit, less preferred stock. Stockholders' equity uses `SEQ`, otherwise `CEQ` plus preferred stock; deferred taxes use `TXDITC`, otherwise `ITCB`; preferred stock uses `PSTKRV`, then `PSTKL`, then `PSTK`. Nonpositive book equity is missing.
Operating profitability (OP)	Revenue less cost of goods sold, SG&A, and interest expense, divided by book equity, $\text{OP} = (\text{REVT} - \text{COGS} - \text{XSGA} - \text{XINT})/\text{BE}$. A missing expense is set to zero only when revenue and at least one of the three expense inputs are observed.
Investment (INV)	Annual asset growth relative to the prior observed annual total-assets record, $\text{INV} = (\text{AT}_y - \text{AT}_{\text{prior}})/\text{AT}_{\text{prior}}$. The prior observed total-assets value must be positive, and the two observations need not represent consecutive fiscal years.

Note: Definitions use standardized, domestic, consolidated annual industrial Compustat records. Signal eligibility requires at least two annual Compustat observations, available inputs, and the applicable positive denominator.

[1] Citation format: Bowles, Boone, Adam V. Reed, Matthew C. Ringgenberg, and Jacob R. Thornock, Internet Appendix to "Factor Time." Working Paper, 2026.

[2] Security market equity is absolute price times shares outstanding. We aggregate by PERMCO and use the largest security's return and PERMNO. We keep date-valid LU/LC CCM links with primary status P/C.

Filing dates come from Compustat's `CO_FILEDATE`. We retain 10-K records, match each filing to annual Compustat by GVKEY and fiscal-period end, and retain the latest filing date when multiple records match the same firm-period. An observation can first enter Fresh at the first formation trading date strictly after its observed or fallback availability date. A missing filing date, or a filing date more than 18 months after fiscal year-end, receives a flagged June 30 fallback in the following year.

Filing-date coverage improves markedly after 1994. Before 1995, approximately 26 percent of capital-weighted observations are associated with fallback dates under both implementations. From 1995 through 2025, the corresponding shares are 4.0 percent for Stale and 3.4 percent for Fresh. We use July 1972–December 2025 when long return histories are valuable and emphasize January 1995–December 2025 when interpreting exact information age.

A.1. Stale, Fresh, and Information Age

Stale and Fresh differ in the information used to assign firms, not in how returns are combined. Both use the same two-by-three architecture, characteristic definitions, breakpoint rules, and value-weighted factor formulas. Realized breakpoints and eligible firms can differ because the two implementations form portfolios on different dates using different accounting observations.

Let $R^X_{gc,t}$ denote the value-weighted return in period t on the portfolio formed at the intersection of size bucket $g \in \{S, B\}$ and characteristic bucket c in sort X. Here, S and B denote Small and Big. The characteristic buckets are $\mathcal{C}_{\text{BM}} = \{L, N, H\}$ (Low, Neutral, and High), $\mathcal{C}_{\text{OP}} = \{W, N, R\}$ (Weak, Neutral, and Robust), and $\mathcal{C}_{\text{INV}} = \{C, N, A\}$ (Conservative, Neutral, and Aggressive). The equations below apply separately to Stale and Fresh:

$$\text{SMB}^X_t = \frac{1}{3} \sum_{c \in \mathcal{C}_X} \left(R^X_{Sc,t} - R^X_{Bc,t} \right), \qquad X \in \{\text{BM}, \text{OP}, \text{INV}\}, \tag{A.1}$$

$$\text{SMB}_t = \frac{1}{3} \left(\text{SMB}^{\text{BM}}_t + \text{SMB}^{\text{OP}}_t + \text{SMB}^{\text{INV}}_t \right), \tag{A.2}$$

$$\text{HML}_t = \frac{1}{2} \left(R^{\text{BM}}_{SH,t} + R^{\text{BM}}_{BH,t} \right) - \frac{1}{2} \left(R^{\text{BM}}_{SL,t} + R^{\text{BM}}_{BL,t} \right), \tag{A.3}$$

$$\text{RMW}_t = \frac{1}{2} \left(R^{\text{OP}}_{SR,t} + R^{\text{OP}}_{BR,t} \right) - \frac{1}{2} \left(R^{\text{OP}}_{SW,t} + R^{\text{OP}}_{BW,t} \right), \tag{A.4}$$

$$\text{CMA}_t = \frac{1}{2} \left(R^{\text{INV}}_{SC,t} + R^{\text{INV}}_{BC,t} \right) - \frac{1}{2} \left(R^{\text{INV}}_{SA,t} + R^{\text{INV}}_{BA,t} \right). \tag{A.5}$$

Thus HML, RMW, and CMA average the relevant characteristic spread across the Small and Big portfolios. SMB first forms a Small-minus-Big spread within each of the three characteristic sorts and then averages those spreads. The Neutral portfolios enter SMB but not HML, RMW, or CMA.

At the end of June in year t, Stale uses the annual record whose fiscal period ends in calendar year $t-1$, without imposing Fresh's filing-date gate. HML pairs book equity with market equity from December of $t-1$. RMW and CMA use the corresponding annual profitability and investment observations. Assignments apply from July of t through June of $t+1$. July weights begin with June market equity and subsequently drift with cumulative ex-distribution returns. The assignment is maintained for one year, but its value weights are not frozen.

Fresh re-forms the portfolios monthly. Return month t uses the assignment associated with month-end $t-1$. Market equity is measured on the designated formation date—the trading day immediately preceding that month-end return date—and only annual information available before formation can enter. HML pairs eligible book equity with formation-date market equity; RMW uses eligible operating profitability; and CMA uses investment computed from eligible total assets and the firm's prior observed annual total-assets record.

Each characteristic carries forward separately. A missing characteristic in a newer filing does not erase the latest nonmissing annual value. The newer report's dates and the firm's annual history still update, and the characteristic remains subject to current eligibility requirements. Assignments affect only subsequent returns.

Stale preserves the conventional annual Fama–French match and is therefore not retrofitted with Fresh's filing-date screen. The comparison is between that annual implementation and an availability-timed monthly implementation. Formation dates, eligible records, breakpoints, assignments, and weights can therefore differ. For nonmarket factor j, the Update return is the identity

$$U_{j,t} = F_{j,t} - S_{j,t}, \tag{A.6}$$

where $S_{j,t}$ and $F_{j,t}$ denote the Stale and Fresh returns. Update is the traded return on the difference in portfolio assignments and weights, not an assumed primitive shock.

Daily factor returns do not re-form Fresh each day. Each trading day uses the Stale annual assignment or Fresh monthly assignment in force at the beginning of the holding month. Portfolio

weights then drift within the month with cumulative gross daily CRSP returns. Daily Update remains Fresh minus Stale.

Information age is the number of calendar days from the associated observed or fallback availability date to the trading day nearest calendar day 15 of the return month. Within each signal-month, Stale and Fresh use the same lag-one formation-market-equity weights. We form capital-weighted age and older-than-365-day shares for each signal, then average equally across the three signals within month and across months.

B. Test Assets and Application Samples

B.1. Industry Portfolios, JKP Strategies, and Dynamic CRSP Stocks

The recurring monthly tests use the 49 Industry portfolios from the Kenneth French Data Library and 153 published U.S. monthly value-weighted long–short strategies from Jensen et al. (2023). We convert the Industry portfolio returns to excess returns by subtracting the monthly risk-free rate reported in the Kenneth French Data Library (the Official monthly RF rate). Both panels are complete throughout each named analysis sample.

Dynamic-stock eligibility is reassessed each month. We retain CRSP common shares with share code 10 or 11 and exchange code 1, 2, or 3, a valid constructed return, and a positive lagged portfolio weight at or above the contemporaneous twentieth percentile of the NYSE weight distribution. Gross returns are winsorized within month at the 0.5th and 99.5th percentiles before Official monthly RF is subtracted. The construction preserves entry and exit and imposes no balanced-history or survival requirement.

For rolling tests in evaluation month t, each model is estimated over the preceding 60 calendar months through $t-1$, with at least 36 observations, and the stock must have a return in month t. Competing models use identical stock-month support. The July 1977–December 2025 rolling sample contains 853,208 stock-months from 6,576 stocks; the January 1995–December 2025 sample contains 561,502 stock-months from 5,111 stocks. Average adjacent-month retention is 98.80 and 98.73 percent, respectively.

The Best Common analysis in Table VII estimates each stock's factor exposures once within each sample. Stocks must have at least 120 monthly observations in that sample, but need not have

a balanced history. The January 1995–December 2025 sample contains 2,120 stocks. The earlier and later samples used in Panel B contain 1,313 and 1,304 stocks.

B.2. Event Applications

Annual earnings announcements are Compustat fiscal-fourth-quarter results announced from January through March. The full sample contains 54,694 announcements for 2,755 stocks from 1974 through 2025. The common January 2000–November 2025 information-date window contains 28,667 announcements from 1,742 stocks and 3,073,648 non-earnings RavenPack stock-news days from 2,115 stocks. Coverage is limited to *Barron's*, *MarketWatch*, the *Wall Street Journal*, and Dow Jones Newswires, not the full RavenPack universe. We retain U.S.-company items with relevance of at least 75, collapse multiple items to one stock-day, and exclude stock-days that coincide with a Compustat quarterly earnings announcement.

The information date is Compustat RDQ for earnings and the reconstructed publication date for news. Day zero is the first factor-trading day strictly after that date. An event must belong to the stock set for the applicable named sample, which requires at least 120 monthly observations within that sample, and satisfy the dynamic-stock screen in the event month. Stock-specific daily intercepts and loadings are estimated over days -272 through -21, with at least 180 observations, and held fixed. Each event is retained only if all three benchmark regressions can be estimated uniquely and the stock has complete returns from day -20 through day $+20$, which includes the reported $(-1, +20)$ CAR window. The counts incorporate these conditions.

B.3. Mutual Funds

The mutual-fund application uses active, diversified U.S. equity fund products from the CRSP Survivor-Bias-Free U.S. Mutual Fund Database from July 2003 through December 2025. Date-effective CRSP class groups define products. We combine share classes using prior-month TNA and retain dead products through their observed histories. A product enters once it is at least 12 months old and has crossed $5 million in lagged TNA during an age-eligible month. The unit is a fund product.

Net return is the prior-month-TNA-weighted return across eligible share classes. Following Fama and French (2010), expense-added return adds one-twelfth of the date-effective annual op-

erating expense ratio using the same weights and is observed only when reported fees cover 100 percent of eligible lagged class TNA. It remains net of trading and other costs embedded in NAV and therefore is not an all-cost gross return. Five-year classifications require one return in every month of the exact 60-calendar-month window ending in December, with identical observations across models. Net-return classifications use 18 December formations from 2008 through 2025. Expense-added alpha-sign classifications use 17 formations from 2008 through 2024 because complete fee support is unavailable for the December 2025 trailing window.