Noisy Factors

Pat Akey, Adriana Z. Robertson, and Mikhail Simutin*

October 27, 2021

ABSTRACT

The Fama-French factors are ubiquitous in empirical finance, industry, and law. We find that factor returns differ substantially depending on when the data were downloaded. The effects of these retroactive changes are large. Holding the sample period constant and varying only factor vintages, we show this in three contexts. First, in cross-sectional equity pricing, unconditional alphas of a third of long-short 'anomaly' portfolios lose statistical significance. Second, in mutual fund analyses, we show that annual alphas of almost half of individual funds and even portfolios of funds change by more than 1%. Third, in the context of testing asset pricing models, F-statistics from GRS tests of the three-factor model on standard test portfolios vary by up to 40% due only to changes in factor vintages. Our results do not suggest that any particular factor vintage is dominant but point to a source of latent noise that is ignored in conventional tests. We provide suggestions to empiricists on dealing with the noise. Our findings have significant implications for the replicability and robustness of finance research and have a direct bearing on a variety of empirical contexts.

*The authors are with the University of Toronto. Akey: pat.akey@rotman.utoronto.ca; Robertson: adriana.robertson@utoronto.ca; Simutin: mikhail.simutin@rotman.utoronto.ca. We thank Oliver Boguth, Alon Brav, Winston Dou, Cam Harvey, Marina Niessner, Raymond Kan, Jonathan Klick and seminar participants at Arizona State University, ESCP Paris, the Foundations of Law and Finance Seminar Series, HEC Paris, Université Laval, Université Paris-Dauphine, University of Alberta, University of Illinois, University of Pennsylvania Law/Wharton Finance seminar series, and University of Toronto for helpful comments.

In one of the most cited papers in financial economics, Fama and French (1993) propose that sensitivities to returns on three long-short portfolios—the excess return on the market, small minus large stocks (SMB), and value minus growth stocks (HML)—explain the cross-section of expected stock returns. This three-factor model has revolutionized finance research, becoming the go-to model for empirical researchers. In asset pricing, it is used to measure factor-adjusted returns on stocks, mutual funds, and other investments. In corporate finance, the model is widely used in event studies and cost of capital calculations. It is taught to PhD, MBA and undergraduate students, and is a part of the CFA curriculum. The model has also had a tremendous impact in practice, where it is used to evaluate real and financial investment decisions, as well as in legal settings to establish liability and to estimate damages.

To apply the model, researchers begin with factor returns. While they can construct their own, researchers commonly rely on factors produced by Kenneth French, which are available on his website and through data vendors like WRDS. These factors undergo periodic revisions chronicled on French's website, and after each update only the most recent factor vintage is available. In noting that the data change, WRDS explains that "Research Portfolios incorporate any revisions in the historical underlying data, and thus computations that use the most recent vintage... may differ from computations that use an earlier vintage. The revisions are typically very small and this set is most commonly used in academic studies."[1] While the last sentence may provide some comfort to researchers, we show in this paper that changes to factor returns are frequent, often substantial, and impact conclusions about first-order questions in finance.

We use archived versions of French's website to obtain factor vintages going back as far as 2005. Even between adjacent vintages, the differences in factor returns are substantial, and tend to increase with the length of time between vintages. While there are large changes in all the factors, the revisions are particularly pronounced for HML. For example, comparing monthly HML returns between the 2005 and 2006 vintages,

[1] https://wrds-www.wharton.upenn.edu/pages/support/manuals-and-overviews/fama-french/fama-french-research-portfolios-and-factors/

90% of the observations are different, with the average absolute difference exceeding 1% annualized. Means are also affected: the average HML return is 5 basis points per month higher in the 2021 vintage than the 2005 vintage, a difference that is both statistically significant and economically meaningful.[2]

We evaluate the effects of these changes in three widely studied settings: the cross-section of stock returns, the performance of actively managed equity mutual funds, and in the comparison of asset pricing models. We begin the cross-sectional returns analysis by showing that estimates of alphas and betas of individual stocks vary dramatically, both in level and significance, depending on the factor vintage used. Switching between 2005 and 2021 vintages causes a quarter of alphas estimated from three-factor regressions on five years of monthly data to change by more than 1% per year, and causes a tenth of significant alphas to lose significance. Estimated loadings on the three factors change by more than 0.1 for between a tenth and a quarter of observations, suggesting important implications for estimates of cost of capital that use betas as inputs. These effects are even more pronounced using shorter (3- or 1-year) estimation periods.

While the impact of noise on individual stock returns generally decline in diversified portfolios, here the noise resides in the factors. As a result, regression estimates obtained using different factor vintages continue to differ substantially even for diversified portfolios. These effects are also pervasive across stocks with different characteristics, affecting portfolios sorted on market size, book-to-market ratio, return runups, and other attributes. For example, alpha estimates of decile book-to-market portfolios change by as much as 4% per year due to nothing more than changing factor vintages. Even unconditional alphas, estimated with 80 years of data, are not immune: for some industry portfolios, they change by over 2% per year across vintages.

In the final set of tests focused on the cross-section of stock returns, we examine performance of "anomalies," or investment strategies that have been shown to generate significant factor-adjusted returns. We obtain returns of 187 high-low anomaly port-

[2] In making comparisons such as this, we hold the sample period fixed by basing the analysis on data common to both factor vintages. In this case, we compare means of HML returns in two vintages through 2005.

folios from Lu Zhang's website (Hou et al., 2020, 2021). We find that unconditional alphas of almost a third of portfolios lose statistical significance due to changes in factor vintages. Economically, the effects are also large, with many alpha estimates changing by more than 1% annually. Momentum-related anomalies tend to be the most sensitive to factor vintages; investment-related anomalies tend to be least affected.

Next, we investigate how the choice of factor vintage impacts inferences about mutual fund risk and performance. Annual alphas and betas of individual funds can vary dramatically across vintages, with almost half of alpha estimates changing by more than 1% annually. The effects are present across funds with different styles and are more pronounced for larger funds. Remarkably, the choice of factor vintage also affects estimates of the average alpha for the *overall* active equity mutual fund industry: In some years, switching vintages changes the average annual alpha by more than 1%.

In our final set of analyses, we study how the performance of the three-factor model is affected by revisions to factor data. Factors change across vintages either because of changes to the input data from CRSP and Compustat, or because of changes to the construction of the factors. To the extent that these changes reduce the noise in approximating the true unobservable factors, updated factor vintages should constitute improvements relative to their predecessors. To evaluate whether such improvements occur, we first ask how a factor's Sharpe ratio changes with each vintage. We find that the Sharpe ratio of HML has increased substantially—by approximately 10%—over the course of the 11 vintage updates in our sample. Over the same period, the Sharpe ratio of SMB has declined slightly, and that of the market factor has not changed.

We next perform a more formal analysis of model performance using GRS tests (Gibbons et al., 1989). These tests usually involve comparing two competing models, each with its own set of factors. Here, we keep the model fixed and instead vary factor vintages. We first use the 25 Fama-French portfolios sorted by size and book-to-market as test portfolios. Just as factor vintages change, returns of these portfolios get updated as well. We find that "model" performance is dramatically unstable, producing F-statistics that vary by over 40% across vintages. We find that vintage updates generally result in lower F-statistics, pointing to better model performance when using more recent vintages. However, when we use industry portfolios as test assets, we observe the opposite, leading us to conclude that GRS tests do not provide evidence to suggest that updates to factors lead to systematic improvements in model performance.

We also compare the performance of the factor vintages using the tests developed by Barillas et al. (2020). The results of these tests, which use factor data as the only input, are similarly inconclusive. While we find nothing to suggest that newer factors perform *worse* than older ones, we also do not find consistent evidence that they are improving. Collectively, these model performance tests suggest that no particular factor vintage dominates the others. Rather, they point to a source of latent noise in the factors that conventional empirical tests do not currently account for.

Taken together, our results suggest that a wide range of commonly studied quantities in finance are sensitive to changes in factors. These changes are substantial, and their impact is far-reaching: estimates of risk and factor-adjusted performance of stocks, characteristic-sorted portfolios, anomaly portfolios, and mutual funds can change significantly when vintages change.

Given how ubiquitous the use of the three factors is in finance and how pronounced the effects from vintage changes can be, our results have significant implications for the replicability and robustness of finance research. They suggest that some findings may fail to replicate not because an empiricist made a mistake but because factor data were updated. This observation leads to two recommendations. First, researchers can facilitate replication by disclosing which factor vintage they use. Second, when feasible, empiricists should evaluate robustness of their results to using different vintages.

For alphas and betas of individual stocks and portfolios, our results suggest that estimates from short-horizon (one year, or even five years) regressions may be unreliable. Longer-horizon estimates are less susceptible to factor noise but of course come with their own problems, particularly if betas are time-varying (e.g., Boguth et al., 2011).

For the same reason, mutual fund alphas and betas, which are often estimated using horizons of one to five years, are also sensitive to factor vintages. Using longer-horizon estimates may be impractical here: many mutual funds have only existed for a few years, and those that have been around for longer may not necessarily be pursuing the same strategy over time. Researchers seeking to overcome noise in the factors may factors based on returns of passive index funds and ETFs rather than the traditional factors, as suggested by Berk and Van Binsbergen (2015).

While our focus is on the three-factor model, the findings extend to other models that use the market, HML, and SMB factors as inputs, including the four-, five-, and six-factor models that add momentum, profitability, and investment (Carhart, 1997, Fama and French, 2015). The noise in these models can be even greater, because noise from additional factors plays an incremental role.

Given that 30% of firms use multifactor models such as the three-factor model to estimate their cost of capital (Graham and Harvey, 2001), our findings also have implications for capital allocation in the economy. Because that cost of capital estimates can vary significantly across factor vintages, noise in factors may contribute to misallocation of capital.

Our paper connects to several strands of literature. First, we contribute to the ongoing debate as to whether there is a "replication crisis" in empirical finance. Hou et al. (2020) and Linnainmaa and Roberts (2018) suggest that a large number of asset pricing anomalies fail to replicate or are due to data snooping. Other research indicates that p-hacking is pervasive in empirical financial economics (e.g., Harvey et al., 2016, Harvey, 2017, Chordia et al., 2020). In contrast, Chen (2020) argues that p-hacking alone cannot explain the large number of asset pricing anomalies that have been identified, while other authors find that many strategies do replicate, although there is evidence that alphas of these trading strategies decay over time (McLean and Pontiff, 2016, Pénasse, 2020, Chen and Zimmermann, 2021, Jensen et al., 2021). Our analyses suggest that the retroactive changes in the factors can have an important role in explaining the difficulty of replicating past studies.

While many papers highlight problems with commonly-used databases in financial economics, three are particularly close in spirit to the subject of our study. Ljungqvist et al. (2009), Patton et al. (2015), and Berg et al. (2020) provide evidence that retroactive changes to the I/B/E/S, hedge fund, and Refinitiv ESG databases, respectively, can change conclusions of research conducted on previous versions of the data.[3] In the same spirit, we show that the qualitative and quantitative conclusions of research questions in equity pricing, mutual funds, and corporate finance that rely on the Fama-French factors can change depending on when the data used for analysis were downloaded. To be clear, nothing in our results suggests that the factors are changed to improve model performance. Rather, the changes could be due to reported updates to definitions, or to retroactive changes in underlying CRSP and Compustat data.

Our study also contributes to the growing literature that evaluates the empirical practices in financial economics, law, and accounting. A number of recent papers summarize current empirical practices in the field (e.g., Bowen et al., 2017) or provide guidance on best practices (Atanasov and Black, 2016, 2021, Fisch et al., 2017, Fisch and Gelbach, 2021, Harvey et al., 2020, Harvey and Liu, 2021, Heath et al., 2020, Mitton, 2020a,b, Spamann, 2019). Several papers discuss the problem of measurement error in various empirical contexts (e.g., Erickson and Whited, 2000, Jennings et al., 2020). Yet other studies provide guidance on how to best account for unobserved heterogeneity or to calculate standard errors (e.g., Gormley and Matsa, 2014, Petersen, 2009). Our paper adds to this literature by identifying a previously unappreciated source of noise in the Fama-French factors and offers some suggestions for empiricists.

[3] The literature pointing out issues in commonly used financial data is vast. Some of it summarized on https://www.kellogg.northwestern.edu/rs/services/computationalconsulting/trainingandreference/database_biases_and_errors.aspx. See also Evans (2010), Aiken et al. (2013), Karpoff et al. (2014), Heider and Ljungqvist (2015), and Schwarz and Potter (2016).

I. The (Noisy) Factor Data

We obtain the current versions of the market, value, size, and momentum factors from Ken French's website (Fama and French, 1993, Carhart, 1997).[4] These data are widely used by researchers, who can also access them via WRDS. French's website also provides returns on the five factors of Fama and French (2015) and on a variety of characteristic-sorted and industry portfolios, which we also download.

To obtain historical vintages of the factor data, we use the Internet Archive, a non-profit digital library. One of its services, the Wayback Machine, allows users to access archived versions of over 580 billion web pages. For each year in which an archived version of the French's website is available, we keep a single data vintage. If more than one archive is available for a given year, we select the one that is closest to the midpoint of the year (i.e., the end of June). Following this procedure, we obtain vintages of the monthly market, value, and size factors from 12 years starting in 2005 and ending in 2021.[5] We follow a similar procedure to obtain daily factor data, historical vintages of the momentum, profitability, and investment factors, and returns of characteristic-sorted and industry portfolios.

When we compare vintages from different years, we restrict the analysis to the sample period that is common to considered vintages. For example, when comparing two vintages containing data through 2005 and 2021, the sample ends in 2005. In all comparisons, we hold everything other than the vintages constant.

A. Factor differences across vintages

We begin by exploring the extent to which factors vary across vintages in Figure 1, where each panel compares the earliest and the latest vintages of a particular factor. The solid black line shows the monthly difference in the realized return of the factor between the two vintages. The blue dash-dotted and red dashed lines plot the cumulative returns

[4] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[5] No archived versions of the three factors are available in 2008, 2009, 2011, 2013, and 2018.

of each vintage.[6] We also show means and standard deviations of the two vintages and their difference in the top left of each panel.

Panel A shows the results for the market factor. Since this factor should change across vintages only when the definition of what constitutes the market or the risk-free asset changes, or to the extent that historical stock returns are revised (perhaps to correct errors in the underlying data) it is reasonable to expect small differences across vintages. While the average difference in factor realizations in the 2005 and 2021 vintages is small (under 1 bp per month), the mean *absolute* difference is considerably larger (over 10 bps). The absolute value of the difference between the two vintages exceeds a quarter of a percent in 77 months.

We observe much larger differences across vintages for the HML (value) factor, presented in panel B. The average return in the 2021 vintage is one-eighth larger than in the 2005 vintage (45 vs 40 bps), a difference that is both statistically significant (t=2.51) and economically important, producing much larger cumulative returns over the sample. Monthly return differences across the two vintages frequently exceed 1% and are particularly substantial in the beginning (1920-40s) and near the end (1990s-2000s) of the sample. The volatility of differences is large: at 0.64%, the variation in the *difference* between HML factor vintages is about one sixth (0.64/3.58) of the magnitude of the *total* variation in the HML factor.

We also find non-trivial differences in average returns of the SMB (size) factor across vintages. Here, it is the 2005 vintage that generates higher cumulative returns, due mainly to its better performance in the last decade of the sample. Again, we observe substantial absolute differences (19 bps on average) over time, which are particularly large early and late in the sample. The standard deviation of the difference between the vintages is also large, representing over 11% of the standard deviation of either of the vintages.

Turning to the remaining factors, Panel D shows that differences in realizations of the UMD (momentum) factor are particularly large in the first half of the sample, frequently

[6]The results are qualitatively similar using daily data.

exceeding 100 bps per month. Panel E shows that the RMW (profitability) factor, whose first vintage dates to 2015, exhibits large differences in average and cumulative returns, particularly since the 1990s. Here, the variation in the differences between the vintages represents more than 18% of the variation in the factor. Finally, Panel F shows that the differences between the earliest (2015) and latest (2021) vintages of the CMA (investment) factor are very small.

In the remainder of the paper, we focus our analyses on the three Fama and French (1993) factors. We do not include the other factors because their vintages do not always correspond to the vintages of the three factors, and, in the case of RMW and CMA, we do not have a rich of a set of vintages.

We present the differences across all pairs of vintages in Table I. The upper triangular entries reflect the results using monthly data, while the lower triangular use daily data. Each pairwise comparison uses the data that is available in both vintages. As a result, the time series is longer when two later vintages are compared.[7]

Several features stand out from Table I. First, the differences in factor realizations across vintages are substantial, even when comparing two vintages that are close in time. For example, the average absolute difference in HML factor returns from the 2005 and 2006 vintages is 11 bps per month. Even between these adjacent vintages, almost a third of monthly return differences exceed 1% in annualized magnitude, and only 10 percent of the reported returns are identical. Second, the absolute magnitude of the differences tends to increase with the time between vintages. For example, when the 2005 vintage is compared with the 2021 vintage, 62 percent of HML observations differ by more than 1% annualized, and only 3 percent are the same. Third, the differences tend to be largest for the HML factor, although all three are affected. Fourth, differences are large in both monthly and daily data.[8]

[7]Monthly and daily factor data are provided in separate files on the French's Data Library. In some years, vintages from the Wayback Machine may be available only for daily or only for monthly data.

[8]While CRSP has periodically invested in verifying the historical data values, our pairwise comparisons suggest that specific updates to the database do not explain our results. One such example occurred in January 2015 when there was a major change of shares outstanding for observations prior to 1946. We observe large differences in factors even when no revisions to data are announced.

While the source of factor noise from retroactive data updates we identify is new, the important question is whether this noise meaningfully affects the results of analyses where factors are used in practice. We next turn to addressing this question.

II. Equities

In this section, we evaluate the effects of factor vintages on the measured performance of equities. We start by examining the extent to which alpha and beta estimates of individual stocks are sensitive to switching factor vintages. While single-stock estimates are known to be noisy, they are used in a variety of asset pricing, corporate finance and legal applications, including firm valuation and event studies. We then turn to diversified portfolios of equities. Even in this context, we find that switching factor vintages can substantially change alpha and beta estimates. Finally, we turn to long-short anomaly strategies, and show that the unconditional alphas of many of these strategies are highly sensitive to the choice of factor vintage.

A. Individual Stocks

For each common stock in CRSP listed on the NYSE, Amex, or Nasdaq, we estimate alphas and betas from three-factor regressions at the end of each calendar year using rolling five-year windows. We use monthly return data and require a minimum of 36 monthly observations for a stock $\times$ year to be included in the sample. We estimate the regressions using each factor vintage, so we obtain 12 sets of regression estimates for each stock $\times$ year in the sample. Alphas are winsorized cross-sectionally at the 1st and 99th percentiles.

To gauge the effects of the noisy factors on the estimates, for each stock $\times$ year, we compute the difference between the parameter estimates obtained using each pair of vintages. Figure 2 plots the histograms and kernel densities for the resulting differences in alpha and beta estimates obtained using the earliest (2005) and the most recent (2021) factor vintages. For convenience, the figure also shows summary statistics in the upper left of each plot.

Panel A shows that the choice of vintages has a large impact on estimated alphas. The average difference in alphas is 15 bps per year, which is both economically meaningful and statistically significant (t=2.3).[9] For more than a quarter of observations, the choice of factor vintage changes the estimated alpha by more than 100 bps, and switching vintages causes 10% of statistically significant alphas to lose significance. These results suggest that in settings where individual stock alphas are important—such as event studies—conclusions about risk-adjusted returns can change substantially just due to using different vintages of factors.

Betas of individual stocks are used in a variety of settings, including estimating the cost of capital for firm valuation, and are of interest in their own right. Panel B of Figure 2 shows that switching between the two vintages causes 11% of estimated market betas to change by more than 0.1. Assuming a market risk premium of about 5% per year, this implies that the choice of factor vintage can generate a difference of 50 basis points per year in the discount rate. The standard deviation of the difference—0.08—is large, amounting to about 16% of the standard deviation of betas estimates.

Consistent with the earlier evidence showing the substantial cross-vintage variation in the HML factor, Panel C of Figure 2 shows that switching between the 2005 and 2021 vintages causes about a quarter of HML loadings to change by more than 0.1, and the standard deviation of the difference corresponds to 17% of the of the standard deviation of the estimated loadings. The magnitudes for SMB, shown in Panel D, are similar to those for market betas.

The cross-vintage differences in alphas and betas in Figure 2 are particularly striking given the relatively long (five-year) estimation window. Table II shows that the effects of noisy factors are substantially amplified in shorter horizons, which are also commonly used in a variety of applications. For example, the last column of the table shows that in the full 1926-2004 sample, when alphas and betas are estimated using a one-year (three-year) window, 65% (34%) of alpha estimates change by more than 100 bps and

[9] The positive mean alpha (using either vintage) reflects the fact that smaller stocks are known to have higher average three-factor alphas. In the pooled mean, all stocks receive the same weight.

15% (10%) of alpha estimates lose significance.

Table II also shows that the effects are substantial throughout the sample period. Importantly, they are particularly large in the most recent subsample, 1986-2004. In this period, 41% of alphas estimated over five years—and 76% of alphas estimated over one year—differ by more than 100 bp. Given that the focus of empirical work tends to be on more recent data, these results suggest that factor noise can be acute in these studies.

In Table III, we show statistics for differences between alphas calculated using each pair of factor vintages. While the differences tend to increase with the time between vintages, that is not uniformly the case. Even for adjacent vintages, differences are large. For example, 17% of alphas change by more than 100 bps when switching between the 2016 and 2017 factor vintages. Even the most recent pair of vintages are not immune to factor noise: there are substantial differences between the alphas computed using the 2020 and 2021 vintages.

In the final set of analyses of individual stocks, we ask whether the effect of factor noise varies with stock characteristics. We sort stocks into quintiles by each four characteristics: market equity, book-to-market ratio, asset growth, and profitability. We summarize the effects of factor noise on alphas by the characteristic quintiles in Table IV. While the effects of microstructure noise are known to be greater in smaller stocks, Panel A shows that the effects of noisy factors are important in both small and large stocks: 21% of the stocks in the largest quintile have alphas that change by more than 100 bp, which is only slightly less than the 29% for stocks in the bottom quintile. The proportion of alphas of large stocks turning insignificant is slightly larger than the corresponding proportion for small stocks (13% versus 12%). These results provide a first indication that the effects of factor noise are likely to remain non-trivial even in diversified value-weighted portfolios, a topic we explore in more detail in the next two subsections.

Panels B, C, and D show the results for stocks in quintiles sorted by book-to-market ratio, asset growth, and profitability, respectively. As in Panel A, there is little variation across the quintiles in each panel. Across all groups and characteristics, the proportion of alphas changing by more than 100 bps is between 17% and 25%, whereas the share of statistically significant alphas that lose significance is between 9% and 14%. In sum, we find no evidence that the effect of noise in factor vintages is concentrated in a subset of stocks with particular characteristics. Based on our analysis, it appears to be ubiquitous.

B. Portfolios of Random Stocks

Many settings in empirical finance involve analyzing portfolios rather individual stocks. It is therefore important to understand how factor noise affects portfolios. To do so, we begin by constructing portfolios of randomly selected stocks. At the beginning of each calendar year t, we split the cross-section of stocks into portfolios containing N random stocks. We hold these portfolios for five years without rebalancing, and calculate alphas of each portfolio during the t through $t+4$ holding period by regressing monthly value-weighted portfolio returns in excess of the risk-free rate on the three factors from different vintages.

We consider portfolios of $N = 10$, 30, 50, and 100 stocks, and repeat the process of creating random portfolios 1, 3, 5, or 10 times, respectively, each year. That is, when forming portfolios of 100 stocks at the beginning of year t, we randomly split the cross-section into portfolios containing 100 stocks each and then repeat this process 10 times. In a year when the cross section contains 3,000 stocks this procedure would produce $(3,000/100) \times 10 = 300$ random portfolios.

Figure 3 shows histograms and kernel densities of the differences in alphas between the 2005 and 2021 factor vintages for the resulting portfolios. Noise in the factors continues to meaningfully affect alphas even in these diversified positions. While, as expected, the percentage of alphas that change by more than 100 bps falls relative to what we observed for individual stocks (26%), it remains substantial at 19%, 14%, 12%, and 9% in portfolios containing 10, 30, 50, and 100 stocks, respectively.

At the same time, the proportion of significant alphas that lose significance actually increases relative to the single-stock setting (11%). In the case of a 100-stock portfolio, it rises to almost 17%. The standard deviation of the difference in alphas as a proportion of the cross-sectional standard deviation also increases from 11% in the case of individual stocks and 10-stock portfolios to 13% for 100-stock portfolios. This set of results may seem surprising. In standard applications, alphas of diversified portfolios can generally be estimated more precisely (relatively to alphas of individual stocks) because noise in the constituent stocks washes out. Here, however, noise resides in factors rather than the constituent stocks. As a result, differences in factors across vintages continue to meaningfully affect alphas even in diversified portfolios. As a result, empirical studies that use portfolios are not immune to factor noise.

C. Characteristic-Sorted Portfolios

While the results in the prior subsection indicate that factor noise is not "diversified away," most studies analyze portfolios that share common characteristics, not portfolios of random stocks. We therefore turn to the question of how differences in factor vintages affect inferences about alphas of portfolios sorted on commonly studied characteristics. Specifically, we consider value-weighted decile portfolios from Ken French's website that are sorted on the following 10 attributes: market equity, book-to-market ratio, profitability, investment, accruals, net issuance, momentum, market beta, variance, and residual variance.

Using characteristic-sorted portfolios introduces a second dimension of vintages: not only do factor returns change across vintages, so do the returns of the portfolios themselves. To keep the analysis focused, we compare the earliest vintage of both the factors and the portfolios to the latest vintage of each. As before, we compute alphas at the end of every calendar year using five years of monthly data.

We present the results in Table V. While there is considerable variation across the ten sets of portfolios, vintages have a substantial effect on all ten sets of alphas. Portfolios sorted by book-to-market and profitability are affected the most, with approximately a third of alpha estimates changing by more than 100 bps and over a half of significant alphas losing significance. In untabulated results, we observe that some estimates change by as much as 4% per year. For the book-to-market and profitability portfolios, the standard deviation of the difference in alphas correspond to over 55% of the cross-sectional standard deviation of the respective alphas. These are dramatic effects in diversified portfolios for which empiricists typically compute a single estimate of alpha.

The effects are also economically meaningful in the other portfolios. For example, changes in vintages cause the alphas of over a tenth of the portfolios sorted by market equity, momentum, accruals, or residual variance to change by 100 bps. On average across the remaining characteristics, the standard deviation of the difference in alphas represents approximately 22% of the cross-sectional standard deviation. These results are all the more striking given that for some portfolios, the earliest available vintages are in 2015, meaning that some comparisons involve comparing vintages from 2015 and 2021.

We also consider unconditional alphas, estimated over the full sample period rather than in five-year windows. Rather than overwhelm the reader with another set of results from a broad set of portfolios, we estimate alphas for two sets commonly considered in asset pricing literature: 25 portfolios sorted on size and book-to-market, and 17 industry portfolios. For each portfolio, we estimate 132 alphas (12 factor vintages $\times$ 11 portfolio vintages), which we round to one tenth of one percent. Figure 4 shows alphas visually, with the size of the bubble representing the frequency with which a particular rounded alpha estimate occurs within the 132 estimates.

For some portfolios, such as S3V2 (corresponding roughly to mid-cap core stocks), oil, and utilities, alphas exhibit little variation across vintages. For many others, the variation is substantial. For example, alpha estimates for the S1V1 portfolio differ by as much as 1.5% per year (-10.4% to -8.9%). The estimates for the S2V5 portfolio differ by 1.3% from a low of -0.64% to a positive high of 0.67%. The effects are even more dramatic in some industry portfolios: alpha estimates for durables range between -3.5% and -0.2%, and the estimates for mines straddle zero, varying between -0.4% and 0.9%.

D. Anomaly Portfolios

Finally, we turn to "anomalies," or investment strategies that have been shown to generate significant factor-adjusted returns. The study of anomalies represents a large and active literature in cross-sectional asset pricing, with hundreds of apparent anomalies documented over the last three decades. We obtain returns on the 187 anomaly portfolios from Lu Zhang's Global-q Data Library.[10] These anomalies are constructed by Hou et al. (2020) and used in Hou et al. (2021) to test an augmented version of the q-factor model of Hou et al. (2015).

The anomalies are grouped into 6 categories: frictions, intangibles, investment, momentum, profitability, and value vs growth. For each of the 187 anomalies and each of the 12 factor vintages, we compute the unconditional three-factor alpha of the high-low portfolio using the sample period common to all vintages (i.e., through 08/2005). In Panel A of Figure 5, for each anomaly, we plot the average alpha across the vintages, and group the anomalies into the six categories. The alphas are represented by one of three symbols: squares denote the strategies that do not produce a statistically significant alpha in any of the 12 vintages; they account for 28% of the 187 anomalies. These "never significant" anomalies are mostly concentrated among intangibles, momentum, and profitability categories. Circles indicate anomalies that are statistically significant in all factor vintages, representing 51% of anomalies. Most of these "always significant" anomalies are in the investment category.

The most interesting group of anomalies is indicated by diamonds. These are strategies that produce statistically significant alphas using at least one factor vintage, but that *do not* produce statistically significant alphas using at least one other factor vintage. In other words, these are anomalies for which the (arbitrary) choice of factor vintages is sufficient to turn the anomaly from significant to insignificant, or vice versa.

[10] http://global-q.org/testingportfolios.html.

These "inconsistently significant" anomalies represent 21% of all the strategies. Remarkably, this indicates that conditional on the anomaly being significant using at least one vintage, switching factor vintages is enough to cause 29% (0.21/(0.21+0.51)) of these "significant" anomalies to lose significance. The inconsistent significance is most common among anomalies based on intangibles (where 50% anomalies lose significance), momentum (46%), and value vs growth (36%). None of investment anomalies are inconsistently significant, and profitability and frictions fall in between (23% and 22%, respectively).

Importantly, many anomalies lose significance not just because they are on the "cusp" of significance. Rather, Figure 5 shows that many of the "inconsistently significant" strategies earn substantial average alpha. For example, the average alpha of the "dividend yield" anomaly is -4.9% per year, but its t-statistic varies between -4.1 and -1.2. The "payout yield" anomaly generates annual alpha of -3.9% on average, yet its t-statistic ranges from -3.2 to 0.6. Panel B of the Figure 5 makes this point by plotting the difference between t-statistics from anomaly-level regressions using 2005 and 2021 factor vintages. The differences are large, exceeding 1 in absolute value for 54 of the 187 anomalies.

Overall, the results in this section demonstrate that changes in factor vintages can have substantial effects on inferences about the risk and return of equities in a variety of contexts. Individual stock alphas can vary dramatically, both in level and significance, depending on which factor vintage is used to estimate them. These effects are not diversified away: alphas of diversified portfolios also exhibit substantial sensitivity to the choice of vintage. Nor are these effects confined to stocks with particular characteristics; rather, they are present among stocks with different market size, book-to-market ratios, return runups, and other attributes. The choice of factor vintages also significantly impacts inferences about the alphas of anomaly portfolios, a key statistic of interest in asset pricing. For almost 30% of anomalies, statistical significance depends on the (arbitrary) choice of which factor vintage is used in the analysis.

III. Mutual Funds

In our next set of analyses, we turn to another empirical setting that relies heavily on multifactor models: mutual funds. Factor-adjusted fund performance is an important area of focus in both the academic literature and in practice.

A. Individual Funds

Our mutual fund return data are from the CRSP Survivirship Bias-Free Mutual Fund Database. We use all returns from actively managed domestic equity mutual funds from 1980 to 2020. We exclude index, sector, and target date funds and group share classes into funds using the MFLINK dataset. For each mutual fund in the sample, we estimate alphas and three-factor beta loadings annually at the end of every calendar year using each factor vintage. The structure of our mutual fund tests closely follows that in the stock-level analysis in the previous section, but we use one year of data (rather than five years) for our baseline analysis. We do this for four reasons: (i) microstructure noise is less of a concern in diversified mutual funds than it is in individual stocks; (ii) performance horizons as long as five years are not commonly analyzed in mutual fund settings; (iii) time variation in mutual fund betas can bias long-horizon estimates of performance; and (iv) running five-year regressions may introduce a survivorship bias in the mutual fund sample. For all these reasons, we focus on one-year regressions but show robustness to using three- and five-year windows. We winsorize alphas cross-sectionally at the 1st and 99th percentiles.

For each mutual fund, we calculate the difference between the estimated alphas and betas from different vintages. Figure 6 plots histograms and kernel densities of the differences in alphas and betas obtained using the earliest (2005) and the most recent (2021) vintages. Panel A shows that, as expected, average net-of-fees alphas are substantially negative. Their magnitude, however, is sensitive to the choice of factor vintage. Average underperformance is 9 bp per year greater when estimated using the 2021 vintage than the 2005 vintage. While this average difference is not statistically significant in the full sample (t=1.11), it exhibits significant variation over time. Figure 7 shows that the average difference exceeds 1% in some calendar years and is below -1% in others. Put differently, inferences about average yearly performance of the overall active equity mutual fund industry can change quite dramatically due to nothing more than switching factor vintages.

Returning to Panel A of Figure 6, almost half of estimated annual alphas change by more than 100 bps between the two factor vintages and 31% of alphas that are statistically significant using one vintage become insignificant using the other. These results further underscore the extent to which mutual fund performance evaluation is sensitive to factor vintage.

Mutual fund factor loadings are often used to assess the risks that funds are exposed to and to investigate the extent to which funds are following their stated strategies (e.g., Sensoy, 2009). Panel B shows the differences in market beta estimates obtained using the two vintages. While the average difference is zero, its standard deviation represents 21% of the cross-sectional standard deviation in market betas, and almost 8% of mutual funds have loadings on the market that change by more than 0.1.

Panel C shows that the variation in HML loadings is even larger, consistent with differences in HML returns across vintages being particularly pronounced (Figure 1). While the mean difference in HML loadings across vintages is small, the standard deviation of that difference is equivalent to about a quarter of the cross-sectional variation in the loadings, and 29% of loadings change by more than 0.1. The loadings on SMB are somewhat more stable: 9% change by at least 0.1, and the standard deviation of the difference represents about 14% of the cross-sectional standard deviation.

Next, we evaluate the sensitivity of the effect of vintages on alpha estimates to the estimation horizon and the sample period. In Table VI, we estimate alphas using one, three, and five years of data (Panels A, B, C) and partition the full sample into three subperiods: the 1980s, 1990s, and 2000s.[11] Consistent with the dramatic cross-vintage differences in factor returns in the latter part of the sample (Figure 1), the variation

[11] The 2000s subsample includes the period from 2000 through 2004.

in mutual fund alphas is concentrated towards the latter part of the sample period. This is significant because the later period is also more frequently used in mutual fund studies. While the alpha estimates in the 1980s display some sensitivity to the choice of factor vintages, the results from the 1990s and 2000s are much more dramatic. Fifty-four percent of one-year alphas in the 1990s—and 48% of those in the 2000s—change by more than 100 bps. Approximately a third of alphas in these two later periods lose significance.

This pattern is present at all estimation horizons. While alphas are predictably less sensitive to factor vintages when estimated using longer windows, changes in factor vintages continue to have substantial effects, especially in the latter part of the sample. For example, focusing on the 5-year estimation horizons in the 2000s, the volatility of the difference across vintages is equivalent to approximately a fifth of the cross-sectional standard deviation of alphas. Twenty-eight percent of these five-year alphas change by more than 100 bps, and 28% of them lose statistical significance simply due to switching factor vintages.

In Table VII, we compare the differences in annual mutual fund alphas for each pair of factor vintages. As with the analogous comparison for stocks (presented in Table III), we find that the differences in alphas generally increase with the time between vintages. However, as with stocks, even adjacent vintages can produce substantially different alpha estimates. For example, 33% of alphas change by over 100 bps when we switch between the 2016 and 2017 vintages, and 15% of significant alphas lose their significance when we switch between 2020 and 2021, the two most recent vintages.

B. Fund Characteristics

Next, we ask whether funds with different characteristics are affected differently by factor noise. We focus on three characteristics that are commonly considered in the literature: fund size (i.e., assets under management), and exposure to two dimensions of style: size and value. We measure the style tilt of each fund using its lagged loadings on the size and value factors, estimated from three-factor regressions on five years of monthly data. We estimate these loadings using every factor vintage and average SMB (HML) loading across vintages to approximate the fund's exposure to size (value) styles.[12]

We sort funds cross-sectionally into quintiles on the basis of assets under management and exposures to size and value factors. For each quintile, Table VIII summarizes the differences in alphas calculated using the 2005 and 2021 factor vintages. Since the average alphas discussed above weigh all funds equally, small funds may disproportionately drive our results despite accounting for only a small proportion of total mutual fund assets under management. We find that this concern is unwarranted in our context. Panel A shows that all funds, irrespective of the value of the assets under their management, are substantially affected by factor noise. The proportion of alphas that change by more than 100 bps is surprisingly stable across quintiles, falling in a tight range between 0.46 and 0.47. If anything, the largest funds might be slightly more sensitive to factor noise: the volatility of the difference in alphas between the two vintages as a share of total cross-sectional volatility increases monotonically with fund size: from 14% for the smallest quintile to 17% for the largest group. The share of fund alphas that lose significance also increases with fund size, from 0.26 to 0.38. This is noteworthy for two reasons. First, by definition, the largest funds represent a disproportionate amount of the assets in the mutual fund industry. Second, estimates of the alphas of large funds are commonly viewed as being less susceptible to noise compared to those of smaller funds. This is not the case when it comes to the effect of factor noise.

Panel B of Table VIII shows that the effect of factor vintage on alphas is not limited to funds with a particular size style tilt. Across the size exposure quintiles, the standard deviation of the difference in alphas represents between 14% and 18% of the cross-sectional standard deviation in alphas. Funds that are more tilted towards small stocks (high size factor exposures) are more likely to have alphas that change by more than

[12]Alternatively, style can be inferred from objective codes on CRSP. In untabulated results, we confirm that our results are not sensitive to defining size and value styles of the fund using Lipper classifications available from CRSP.

100 bps (56%). However, the percentage of such observations remains large for funds that are tilted towards large stocks (36%) and those in the middle (46%). Consistent with the evidence in Panel A, it is the large-cap funds that are more likely to see their statistically significant alphas become insignificant (37% compared to 32% and 31% for the middle and top quintiles, respectively).

Finally, Panel C shows that funds with deeper growth or value tilts (those in the Low and High quintiles, respectively) are more likely to experience large changes in alphas when the factor vintage changes (55% and 51%, respectively). This result is consistent with the value factor experiencing the most dramatic changes across vintages, in turn causing the alphas of funds with large positive or negative exposure to the factor to change by a large amount. Average alphas of growth funds experience the greatest proportional change when switching factor vintages from 2005 to 2021: the average alpha across all such funds drops from -1.28% per year to -1.49%, or about one-sixth. This drop is driven by a combination of these funds' negative exposure to the value factor and a higher average return of HML in the 2021 vintage. In terms of changes in inferences about statistical significance, all value factor exposure quintiles are affected roughly equally, with between 30% and 36% of significant alpha observations turning insignificant from one vintage to the next.

Taken together, the results in this section indicate that the choice of factor vintage has a substantial impact on performance evaluation across the mutual fund market. Alphas and betas of individual funds can change dramatically with factor vintages, and large funds are not immune to the effects of factor noise. The effects are present across style tilts, with particularly pronounced changes in alphas occurring in funds with large tilts to value or growth.

IV. Are The Factors Improving?

The factors can change across vintages for two reasons: either the underlying data from CRSP and / or Compustat changes (perhaps to correct errors in the data), or the method used to construct them (including the definitions of key variables such as book equity) changes. To the extent that these changes in the data or construction reduce noise or otherwise lead to better proxies for the true unobservable factors, more recent factor vintages should represent an "improvement" relative to older vintages. Usually, the literature discusses improvement in the context of comparing two competing models, such as the CAPM and the three-factor model, each with its own set of factors. Here, we keep the model fixed and instead compare the performance of the same model with different factor vintages.

We employ three model comparison techniques. First, we analyze how changes to the factors affect the Sharpe ratios of those factors. Second, we use the classic "GRS" test of Gibbons et al. (1989) to study whether more recent factors better explain the returns of standard test assets. Finally, we employ the "BKRS" test of Barillas et al. (2020) to evaluate improvements in the squared Sharpe ratios across factor vintages.

A. Changes in Sharpe Ratios of Factors

We begin by computing the change in Sharpe ratios between adjacent vintages. For each factor, and each pair of adjacent vintages, we compute the difference in Sharpe ratios of the successor and predecessor vintages using the sample period common to both vintages. For example, the 2005 factor vintage covers the sample from July 1926 to August 2005. This predecessor vintage is succeeded by the 2006 vintage, which spans July 1926 to June 2006. To evaluate how the Sharpe ratio of the HML factor changes due to this vintage update, we compute two values for its Sharpe ratio, one from each vintage, using the same July 1926 through August 2005 sample period. The resulting Sharpe ratios are 0.1121 and 0.1137, suggesting that as the factor changed from 2005 to 2006 vintages, the Sharpe ratio of the HML factor increased by 0.0016. This change is represented by the solid black line in Figure 8 increasing to 0.0016 in year 2006.

Following a similar process, we find that the Sharpe ratio of the HML factor increased by a further 0.0006, from 0.1137 in the 2006 vintage to 0.1143 in the 2007 vintage, where both Sharpe ratios are computed using the data through June 2006. This additional increase results in the cumulative change in the Sharpe ratio of HML between 2005 and 2007 of $0.0016 + 0.0006 = 0.0024$, indicated by a dot on the black line in Figure 8 in year 2007.

The Figure shows that the vast majority of changes between vintages increased the Sharpe ratio of the HML factor. The cumulative increase is substantial: given that the average Sharpe ratio across vintages is 0.108, it represents an increase of over 10%. By this simple metric, the HML factor has, on average, improved over time.

The dashed red line in Figure 8 shows that while the SMB factor experienced some modest increases in its Sharpe ratio due to changes in vintages in the first half of the sample, subsequent changes resulted in a lower ratio. The decrease cumulated over all vintage changes (-0.0044) accounts for a reduction of approximately 6% in the Sharpe ratio of SMB. At the same time, as the dash-dotted blue line shows, the Sharpe ratio of the market factor remained essentially unchanged.

B. GRS Tests

To more formally test whether more recent factor vintages perform better, we next run a series of GRS tests. These tests are widely used to rank asset pricing models by relative performance (e.g., Fama and French, 2015). Here, we treat each factor vintage as a "model," and compare the performance of each such model against the others. We restrict the sample to July 1926 through August 2005, the period common to all factor vintages.

GRS tests require a set of test assets. The first set of test assets that we employ are the widely used 25 value-weighted portfolios sorted on size and book-to-market. As we noted in Section II.C, portfolio vintages also undergo changes. We consider all vintages of factors and portfolios available, resulting in 132 GRS tests (12 factor vintages $\times$ 11 portfolio vintages).

Panel A of Table IX summarizes the GRS F-test statistics from the 132 tests. Lower values indicate superior "models." Rows and columns of the table correspond to different vintages of portfolios and factors, respectively, and the highlighted cells show statistics from tests that use contemporaneous portfolio and factor vintages. There are several striking features of the results. First, there is substantial variation in the test statistics across vintages, with the largest value (3.47) exceeding the smallest (2.46) by over 40%. In other words, "model" performance is highly unstable, varying substantially due solely to changes in vintages.

Second, the model does not perform better when the factors and the portfolios are from the same vintage. That is, there is no evidence that the F-statistics are systematically lower along the highlighted diagonal. This is particularly surprising given that portfolios are presumably formed using same versions of CRSP and Compustat data used to compute the factors.

Third, the F-statistics tend to decline as we move down across the diagonal of highlighted cells, with the lowest value appearing in the vintages from 2021. In other words, changes in vintages generally result in better model performance. Curiously, these improvements seem to be due to changes in *portfolio*—rather than factor—vintages: test statistics are fairly stable as we move across rows but tend to decline as we move down the columns.

We repeat the analysis using the 17 value-weighted industry portfolios. The resulting F-statistics are summarized in Panel B of Table IX. Again, we see wide variation in F-statistics across vintages: from a low of 3.41 to a high of 4.41. One important difference, however, is in the analysis of contemporaneous vintages of factors and portfolios. While vintage updates appear to have resulted in better model performance when the 25 size and book-to-market portfolios are used as test assets in Panel A, the updates led to a deterioration in performance when using industry portfolio as test assets in Panel B, where F-statistics increase from 3.54 (in 2005) to 4.20 (in 2021).

Overall, the results in Table IX are inconclusive, and suggest that there is no consistent pattern to the differences across factor (and portfolio) vintages. Certainly, there is no systematic evidence that the updates represent an improvement. The fluctuations, however, can significantly affect the interpretation of standard asset pricing tests. To give just one example, suppose a researcher ran the tests using the 2006 factor and portfolio vintages. She would find that the 3-factor model performs about the "same"—in terms of the F-statistics—in pricing the 25 size and book-to-market portfolios as it does in pricing the 17 industry portfolios (3.34 versus 3.49). Were she to try to replicate this result in 2021, however, she would find a very different picture: using this vintage, the model performs markedly better for the size and book-to-market portfolios than the industry portfolios (F-statistics of 2.50 versus 4.20).

C. BKRS Tests of Model Comparison with Sharpe Ratios

Building on Barillas and Shanken (2017), Barillas et al. (2020) develop tests that permit model comparison on the basis of the squared Sharpe ratio. When comparing two models, the one whose factors produce a higher squared Sharpe ratio is viewed as dominating the other. A particularly attractive feature of these tests is that they require only the factors themselves as inputs, and do not rely on test assets.

Table X presents the results of BKRS tests under the null hypothesis that the squared Share ratios for each pair of factor vintages are equal. Panel A shows differences in squared Sharpe ratios, and Panel B provides the corresponding p-values. A few observations are noteworthy. First, all differences of the squared Sharpe ratios in Panel A are non-negative, suggesting that vintage updates do not result in the deterioration of model performance. Second, the difference in squared Sharpe ratios generally increase with the time between vintages. In particular, Panel A shows that the differences are smallest close to the diagonal, corresponding to comparisons of immediately adjacent or otherwise close vintages, and largest in the top right corner, which compares some of the most recent and oldest vintages. Fourth, none of p-values in Panel B are less than 5%, although some are smaller than 10%. However, even when the differences are statistically significant at 10%, the economic difference between square Sharpe ratios is small.[13]

[13]In addition to pairwise tests, we also consider multiple-model comparison tests of Barillas et al. (2020). The null hypothesis of these tests is that none of the other models is superior to the benchmark model. Given that no vintage dominates another economically or statistically in pairwise tests, it is not surprising that no

The results of all three of these approaches to model comparison tell a similar story: there is no robust evidence that one factor vintage is preferable to any other. While there some evidence that factors improve across vintages, those improvements tend to be economically small. We do note, however, that we find no evidence that updates to the factors lead to *worse* model performance.

V. Implications and Recommendations

We have shown that the changes to the Fama-French factors have large effects on the economic magnitudes and statistical significance of empirical research in asset pricing, corporate finance, and mutual funds. Our results have obvious implications for discussions about the state of replicability in financial economics. However, the impact of these retroactive changes is not confined to academic research. The Fama-French model is frequently taught as a "gold standard" to undergraduate and MBA students, and has been widely accepted by industry where its use may receive less critical evaluation than in academic contexts. The model is used to evaluate the performance of mutual funds and therefore affects allocation of investment capital and career outcomes of managers. Firms use the model to calculate their cost of capital in capital budgeting decisions. Single-firm event studies are commonly used by expert witnesses to determine liability and damages in securities litigation.

One obvious implication of our findings is that empiricists should indicate which vintage of the factors they are using in their analysis and should be aware that if they update their data, the factors will likely change at least a little. Moreover, authors may find it helpful to keep a log of their download dates for posterity. Authors who attempt to replicate findings should be aware that attempts to replicate papers that use the Fama-French factors may yield different results purely because of revisions to the factors. Going forward, empiricists may wish to verify the robustness of their results to the use of different factor vintages prior to disseminating their findings. Similarly, users

single model emerges as dominant in multiple-model comparison tests. We omit these tests for brevity. We thank Raymond Kan for offering the code for the pairwise and multiple-model tests.

of the factors in industry should consider evaluating the robustness of their estimates to using different factor vintages.

For alphas and betas of individual stocks and portfolios, our results suggest that regressions estimated using short-horizons—such as one-, or even five-years—may be unreliable. Longer-horizon estimates are modestly less susceptible to factor noise and, if time variation in betas does not introduce additional issues in analysis, may be preferable to short-horizon estimates.

Our results also provide support for the approach to mutual fund performance evaluation taken by Berk and Van Binsbergen (2015). Unlike factors, the returns on financial assets like index funds and ETFs are much less likely to be subject to retroactive changes.

Many models build on the three-factor model by incorporating other factors. For example, Carhart (1997) adds the momentum factor, Fama and French (2015) instead includes the profitability and investment factors, and many authors combine all of these in a six-factor model. All these settings still use market, HML, and SMB factors, and so the issues we identify for these factors extend to these contexts, where the overall noise from updates to multiple factors can be substantial. For example, we showed in Figure 1 that revisions in momentum and profitability factors cause nontrivial retroactive changes. The commonly used international factors also undergo revisions. In all these settings, empiricists may wish to communicate the vintages they use in their analysis and verify sensitivity of the results to different vintages.

In future revisions to the manuscript, we will provide more specific guidance on how empiricists may wish to adjust standard errors when conducting inference to account for the noisy factors.

VI. Conclusion

The returns on the Fama-French factors—which are among the most ubiquitous inputs in empirical finance—differ substantially depending on when the data were downloaded.

These differences stem from large retroactive changes to the data, which in turn have large effects on standard applications of the factors. We show this in several contexts, in all cases holding the sample period fixed to restrict attention to the effects of the changes in the factors. In the cross-section of equities, changes in factor vintages have a substantial effect on estimated alphas and factor loadings of both individual stocks and portfolios. Affected portfolios include well-known "anomaly" portfolios: changing factor vintage is enough to cause unconditional alphas of a third of long-short anomaly portfolios to lose statistical significance. Mutual funds are also affected: annual alphas of almost half of individual funds and even portfolios of funds change by more than 1%.

We next turn to tests of model fit. F-statistics from GRS tests of the three-factor model on standard test portfolios vary by up to 40% due only to changes in factor vintages. Our results do not suggest that the factors are improving over time, or that any particular factor vintage is dominant. Rather, they point to a source of latent noise that is not being accounted for in conventional tests. Our findings have significant implications for the replicability and robustness of finance research and have a direct bearing on a variety of legal contexts, including securities fraud, corporate valuation, and estimation of damages. They also suggest that analysis of other factors, including international ones, could be a fruitful endeavor.

References

Aiken, Adam L, Christopher P Clifford, and Jesse Ellis, 2013, Out of the dark: Hedge fund reporting biases and commercial databases, *The Review of Financial Studies* 26, 208–243.

Atanasov, Vladimir, and Bernard Black, 2021, The trouble with instruments: The need for pretreatment balance in shock-based instrumental variable designs, *Management Science* 67, 1270–1302.

Atanasov, Vladimir A, and Bernard S Black, 2016, Shock-based causal inference in corporate finance and accounting research, *Critical Finance Review* 5, 207–304.

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.

Berg, Florian, Kornelia Fabisik, and Zacharias Sautner, 2020, Rewriting history ii: The (un) predictable past of esg ratings, *European Corporate Governance Institute Working Paper* .

Berk, Jonathan B, and Jules H Van Binsbergen, 2015, Measuring skill in the mutual fund industry, *Journal of Financial Economics* 118, 1–20.

Boguth, Oliver, Murray Carlson, Adlai Fisher, and Mikhail Simutin, 2011, Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas, *Journal of Financial Economics* 102, 363–389.

Bowen, Donald E, Laurent Frésard, and Jérôme P Taillard, 2017, What's your identification strategy? innovation in corporate finance research, *Management Science* 63, 2529–2548.

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

Chen, Andrew Y, 2020, The limits of p-hacking: Some thought experiments, *The Journal of Finance* .

Chen, Andrew Y, and Tom Zimmermann, 2021, Open source cross-sectional asset pricing, *Critical Finance Review* .

Chordia, Tarun, Amit Goyal, and Alessio Saretto, 2020, Anomalies and false rejections, *The Review of Financial Studies* 33, 2134–2179.

Erickson, Timothy, and Toni M Whited, 2000, Measurement error and the relationship between investment and q, *Journal of political economy* 108, 1027–1057.

Evans, Richard B, 2010, Mutual fund incubation, *The Journal of Finance* 65, 1581–1611.

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, 2015, A five-factor asset pricing model, *Journal of financial economics* 116, 1–22.

Fisch, Jill E, and Jonah B Gelbach, 2021, Power and statistical significance in securities fraud litigation, *Harvard Business Law Review* 11, 55.

Fisch, Jill E, Jonah B Gelbach, and Jonathan Klick, 2017, The logic and limits of event studies in securities fraud litigation, *Tex. L. Rev.* 96, 553.

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

Gormley, Todd A, and David A Matsa, 2014, Common errors: How to (and not to) control for unobserved heterogeneity, *The Review of Financial Studies* 27, 617–661.

Graham, John R, and Campbell R Harvey, 2001, The theory and practice of corporate finance: Evidence from the field, *Journal of financial economics* 60, 187–243.

Harvey, Campbell R, 2017, Presidential address: The scientific outlook in financial economics, *The Journal of Finance* 72, 1399–1440.

Harvey, Campbell R, and Yan Liu, 2021, Lucky factors, *Journal of Financial Economics* .

Harvey, Campbell R, Yan Liu, and Alessio Saretto, 2020, An evaluation of alternative multiple testing methods for finance applications, *The Review of Asset Pricing Studies* 10, 199–248.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *The Review of Financial Studies* 29, 5–68.

Heath, Davidson, Matthew Ringgenberg, Mehrdad Samadi, and Ingrid M Werner, 2020, Reusing natural experiments, *Fisher College of Business Working Paper* 021.

Heider, Florian, and Alexander Ljungqvist, 2015, As certain as debt and taxes: Estimating the tax sensitivity of leverage from state tax changes, *Journal of financial economics* 118, 684–712.

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, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28, 650–705.

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

Jennings, Jared N, Jung Min Kim, Joshua A Lee, and Daniel J Taylor, 2020, Measurement error and bias in causal models in accounting research, *Available at SSRN 3731197* .

Jensen, Theis Ingerslev, Bryan T Kelly, and Lasse Heje Pedersen, 2021, Is there a replication crisis in finance?, Technical report, National Bureau of Economic Research.

Karpoff, Jonathan M, Allison Koester, D Scott Lee, and Gerald S Martin, 2014, Database challenges in financial misconduct research, *University of Washington working paper* .

Linnainmaa, Juhani T, and Michael R Roberts, 2018, The history of the cross-section of stock returns, *The Review of Financial Studies* 31, 2606–2649.

Ljungqvist, Alexander, Christopher Malloy, and Felicia Marston, 2009, Rewriting history, *The Journal of Finance* 64, 1935–1960.

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

Mitton, Todd, 2020a, Economic significance in corporate finance, *Available at SSRN 3667830* .

Mitton, Todd, 2020b, Methodological variation in empirical corporate finance, *Available at SSRN 3304875* .

Patton, Andrew J, Tarun Ramadorai, and Michael Streatfield, 2015, Change you can believe in? hedge fund data revisions, *The Journal of Finance* 70, 963–999.

Pénasse, Julien, 2020, Understanding alpha decay, *Available at SSRN 2953614* .

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

Schwarz, Christopher G, and Mark E Potter, 2016, Revisiting mutual fund portfolio disclosure, *The Review of Financial Studies* 29, 3519–3544.

Sensoy, Berk A, 2009, Performance evaluation and self-designated benchmark indexes in the mutual fund industry, *Journal of Financial Economics* 92, 25–39.

Spamann, Holger, 2019, On inference when using state corporate laws for identification, *Harvard Law School John M. Olin Center Discussion Paper* .

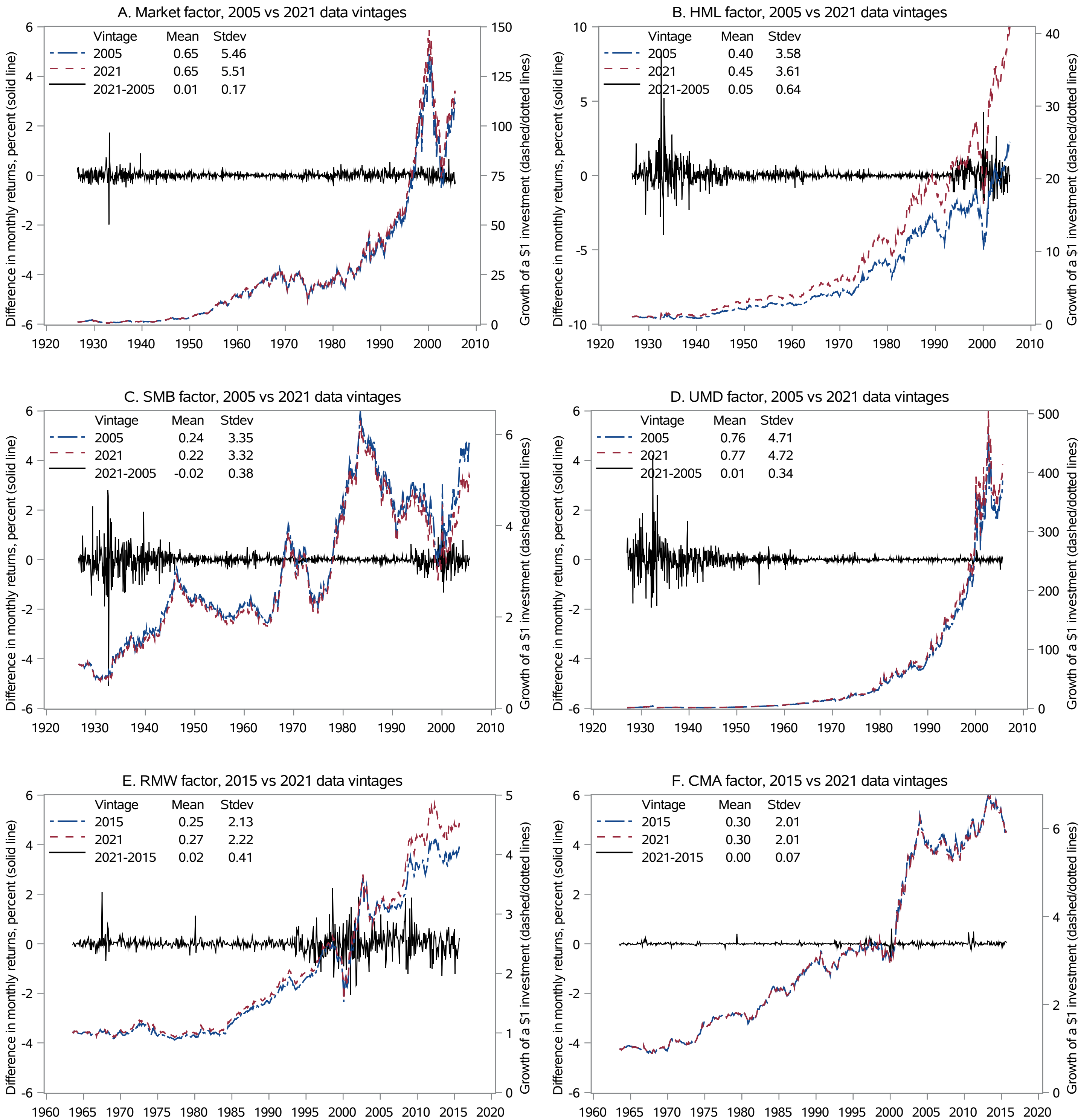

Figure 1. Factor returns from different vintages

This figure plots differences between monthly returns of factors from two vintages (solid black line). It also shows cumulative returns of factors from the two vintages (dashed and dash-dotted lines). Top left of each panel reports means and standard deviations of factor returns, in percent per month.

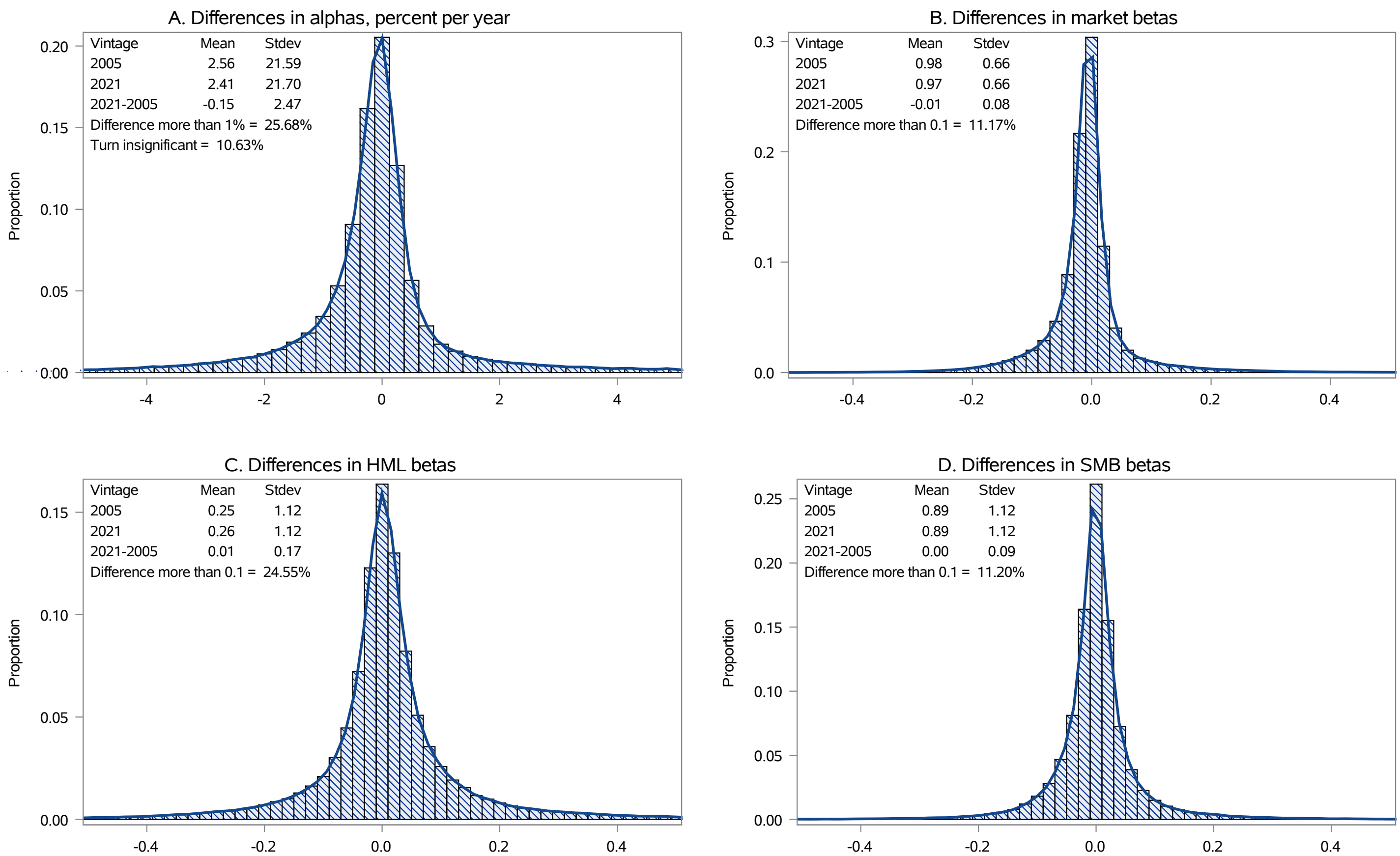

Figure 2. Differences in stock-level alphas and betas: 2005 vs 2021 factor vintages

This figure plots histograms and kernel densities of differences in alphas (Panel A) and betas (B, C, D) of individual stocks estimated using 2005 and 2021 factor vintages. Alphas, in percent per year, and betas are estimated at the end of every calendar year using three-factor regressions on five years of monthly data. Top left of each panel reports means and standard deviations of estimates from the two vintages and of their differences. It also shows the fraction of observations with absolute differences above a certain threshold and in Panel A the proportion of alphas that are significant in one vintage but not in the other.

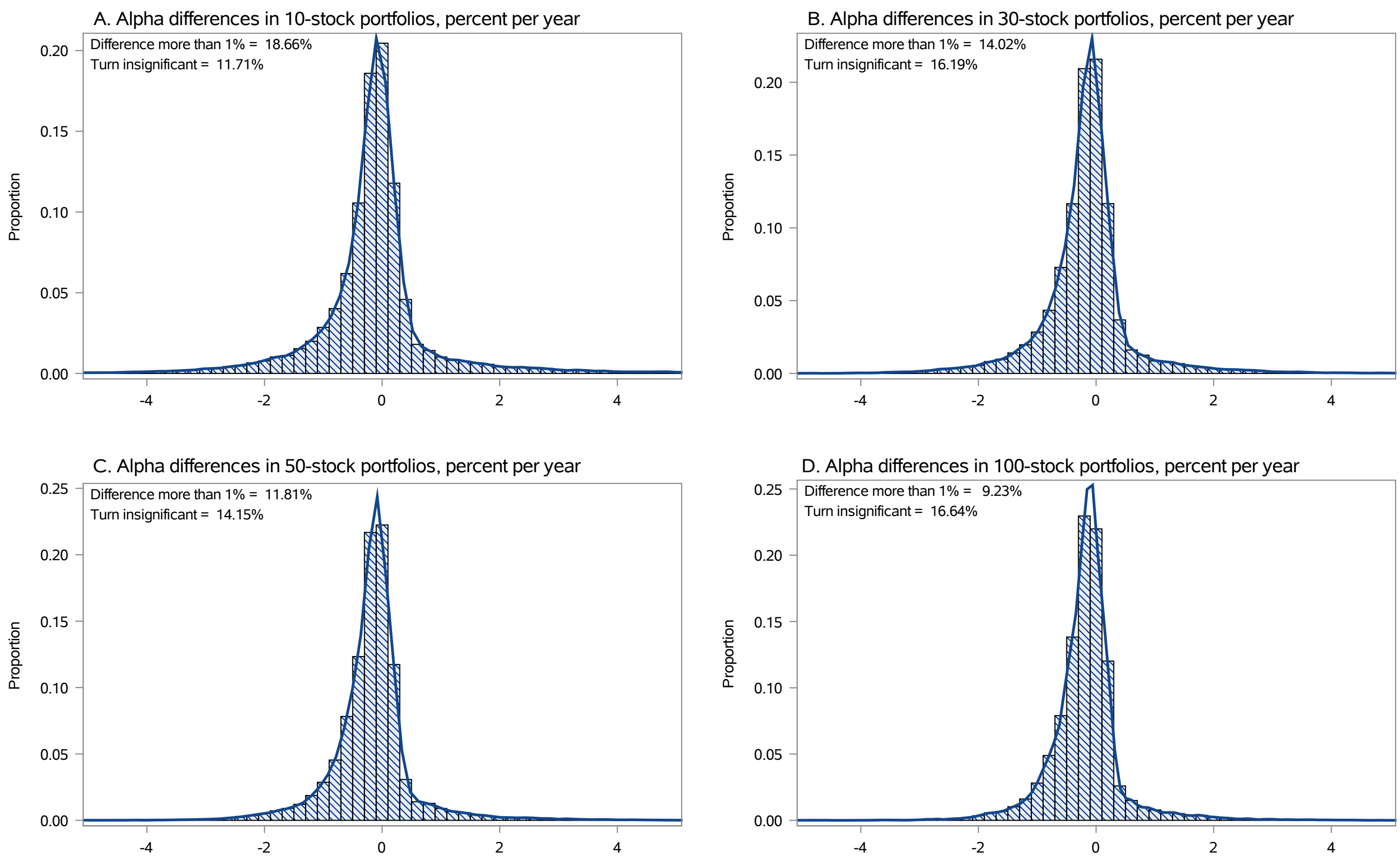

Figure 3. Differences in alphas of random stock portfolios: 2005 vs 2021 factor vintages
This figure plots histograms and kernel densities of differences in alphas estimated using 2005 and 2021 factor vintages for portfolios of randomly chosen stocks. At the beginning of every five-year period, random portfolios are created to contain between 10 (Panel A) and 100 (D) stocks. Alphas, in percent per year, are estimated for each portfolio and factor vintage using regressions of five years of monthly value-weighted portfolio returns in excess of the risk-free rate on the three factors. Top left of each panel shows the fraction of observations with absolute differences above a certain threshold and the proportion of alphas that are significant in one vintage but not in the other.

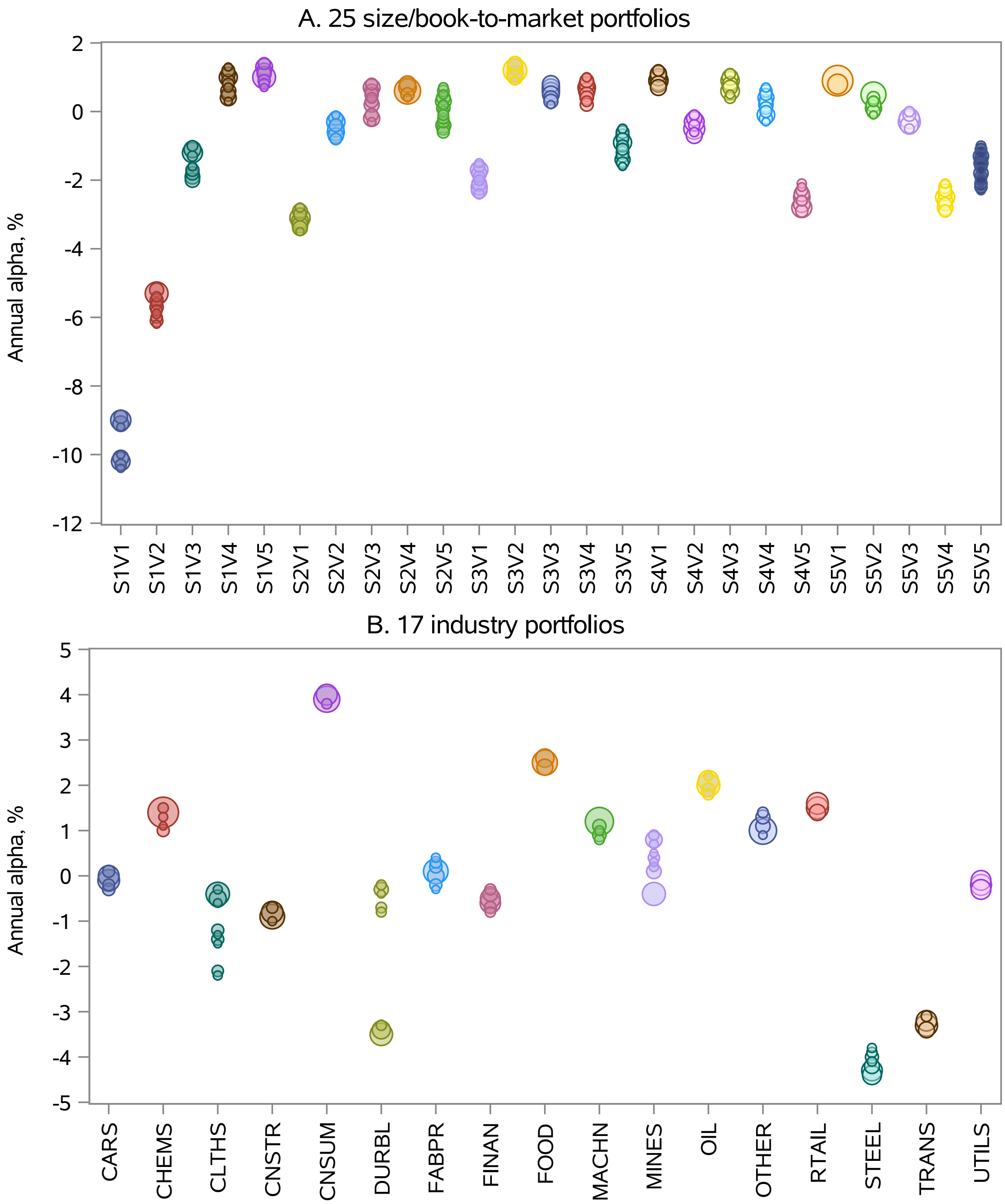

Figure 4. Alphas of 25 size/book-to-market and 17 industry portfolios estimated using different factor and portfolio vintages

This figure plots unconditional three-factor alphas, in percent per year, of 25 size and book-to-market sorted portfolios (Panel A) and industry portfolios (B). Alphas are estimated for each of 12 factor vintages and 11 portfolio vintages using the sample common to all vintages: 07/1926-08/2005 in Panel A and 07/1926-05/2005 in Panel B. Alphas are rounded to the nearest 0.1%, and the size of the bubbles represents the relative frequency of estimates. All data are from Kenneth French's website.

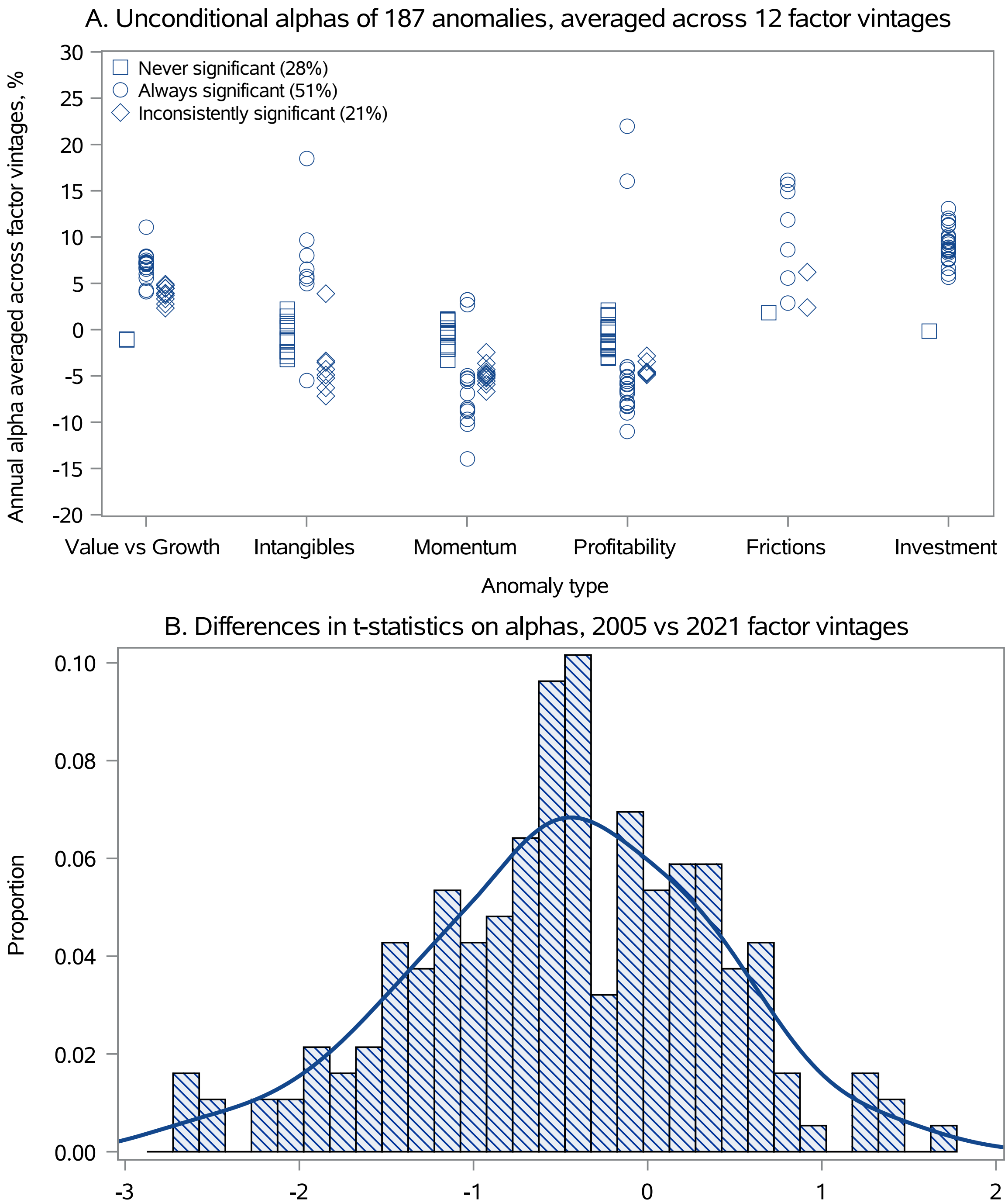

Figure 5. Alphas and t-statistics of high-low portfolios of 187 anomalies

This figure plots in Panel A unconditional alphas, in percent per year, of 187 anomalies, averaged across 12 factor vintages. Anomalies are the high-low portfolios from Lu Zhang's website. For each anomaly, alphas are estimated using unconditional regressions using the sample period common to all factor vintages. Different markers indicate anomalies whose alphas are not significant in any vintage (squares), significant in all vintages (circles), or significant in some but not all vintages (diamonds). Panel B shows the differences in t-statistics of alphas estimated for anomalies using 2005 and 2021 factor vintages.

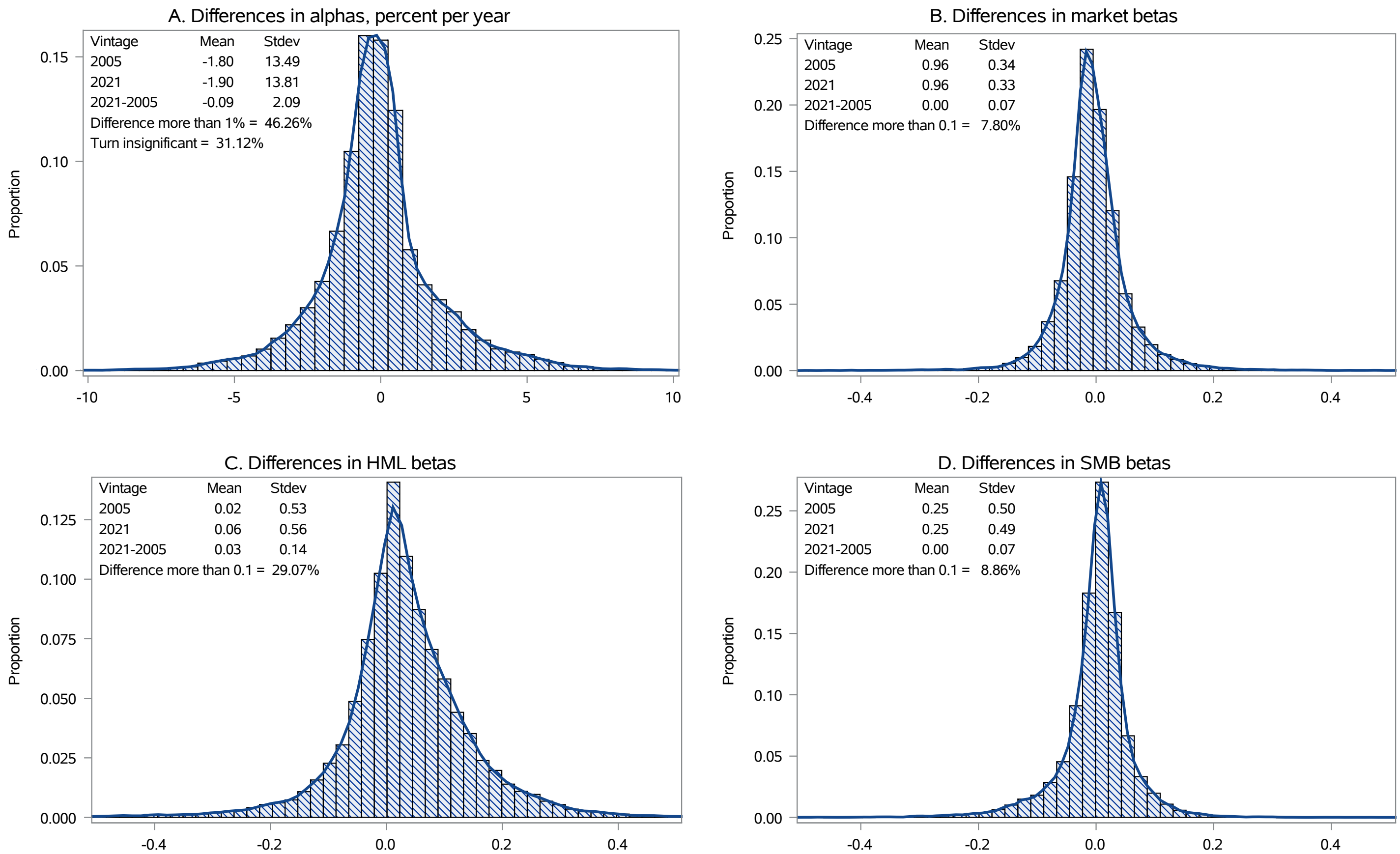

Figure 6. Differences in mutual fund alphas and betas: 2005 vs 2021 factor vintages

This figure plots histograms and kernel densities of differences in alphas (Panel A) and betas (B, C, D) of individual mutual funds estimated using 2005 and 2021 factor vintages. Alphas, in percent per year, and betas are estimated in every calendar year using three-factor regressions on monthly data. Top left of each panel reports means and standard deviations of estimates from the two vintages and of their differences. It also shows the fraction of observations with absolute differences above a certain threshold and in Panel A the proportion of alphas that are significant in one vintage but not in the other.

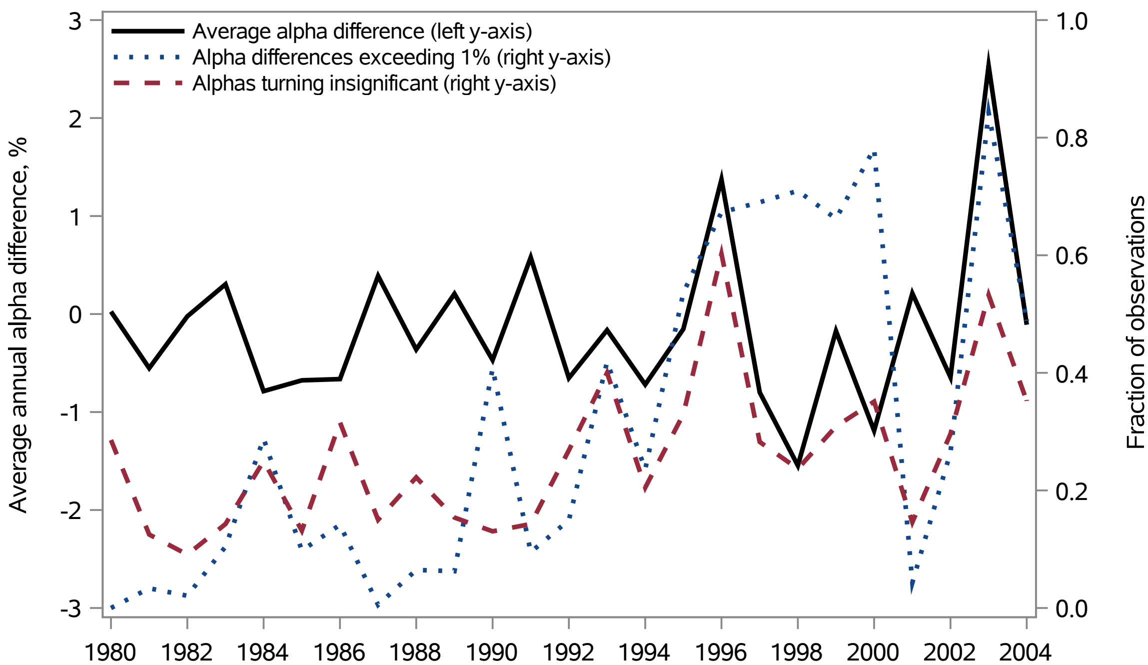

Figure 7. Evolution of differences in mutual fund alphas: 2005 vs 2021 factor vintages
This figure plots the time series of statistics of estimates of mutual fund alphas obtained using 2005 and 2021 factor vintages. Alphas are estimated in every calendar year using three-factor regressions on monthly data. The solid line shows average alpha, in percent per year. The dotted line plots the fraction of funds whose alphas from the two vintages differ by more than 1% annualized. The dashed line indicates the proportion of alphas that are significant in one vintage but not in the other.

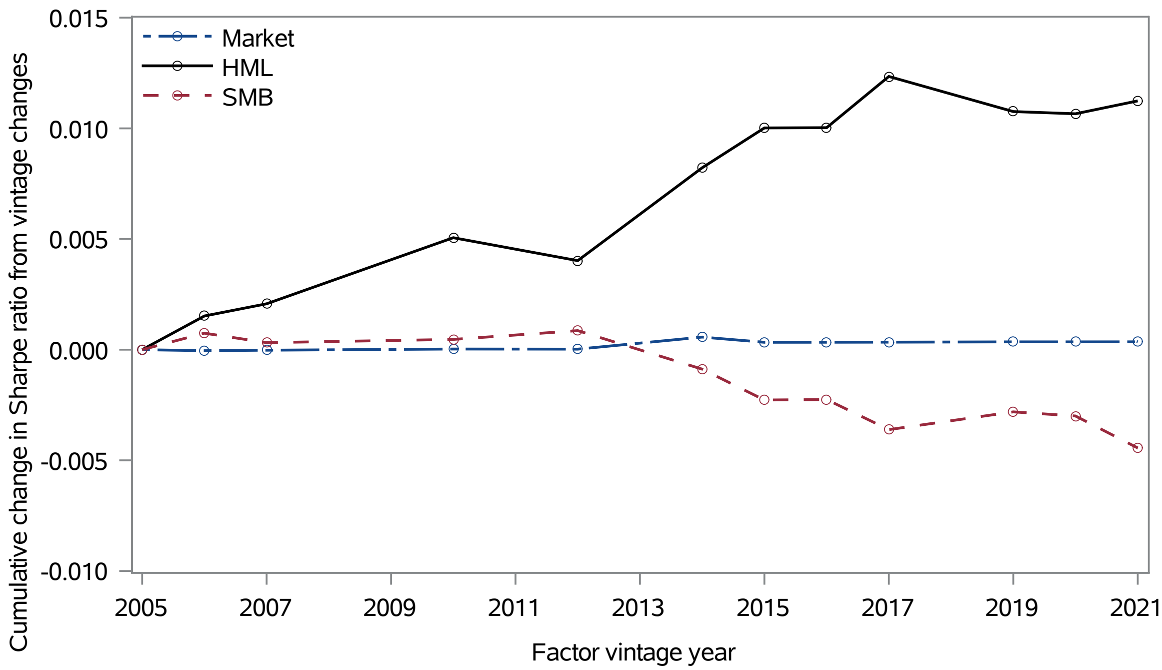

Figure 8. Cumulative changes in Sharpe ratios of factors across vintages
This figure plots cumulative changes in Sharpe ratios of the three factors that arise due to updating factor vintages. For each factor and each adjacent pair of vintages, Sharpe ratios are calculated using data common to both vintages. The differences in the two Sharpe ratio estimates are then cumulated over time.

Table I
Differences in returns of factors across vintages

This table reports statistics for differences in returns of market (Panel A), HML (B), and SMB (C) factors from different factor vintages. Upper and lower triangular entries reflect the results using monthly and daily data, respectively. Mean $\vert$Diff$\vert$ is the average absolute difference in factor returns, in percent monthly. SD Diff is the standard deviation of the difference, in percent monthly. $\vert\text{Diff}\vert > 1\%/\text{yr}$ is the proportion of observations where the absolute difference exceeds 1% per year, which translates into 1%/12 in monthly data and $1\%/(12 \times 21)$ in daily data. The row labeled Not same shows the proportion of factor return observations that is different in the two compared vintages. When comparing vintages, all data common to both vintages is used.

		Vintage 1											
Vintage 2	Variable	2005	2006	2007	2010	2012	2014	2015	2016	2017	2019	2020	2021
A. Market factor													
2005	Mean $\vert$Diff$\vert$		0.02	0.02	0.03	0.03	0.10	0.10	0.10	0.10	0.10	0.10	0.10
	SD Diff		0.10	0.10	0.10	0.10	0.16	0.17	0.17	0.17	0.17	0.17	0.17
	$\vert$Diff$\vert$ > 1%/yr		0.07	0.07	0.08	0.08	0.42	0.43	0.43	0.43	0.43	0.43	0.43
	Not same		0.39	0.39	0.42	0.44	0.96	0.95	0.95	0.95	0.95	0.95	0.95
2006	Mean $\vert$Diff$\vert$	0.00		0.00	0.00	0.00	0.09	0.09	0.09	0.09	0.09	0.09	0.09
	SD Diff	0.01		0.00	0.02	0.02	0.13	0.13	0.13	0.13	0.13	0.13	0.13
	$\vert$Diff$\vert$ > 1%/yr	0.00		0.00	0.01	0.01	0.38	0.39	0.39	0.39	0.39	0.39	0.39
	Not same	0.00		0.01	0.14	0.16	0.95	0.95	0.95	0.95	0.95	0.95	0.95
2007	Mean $\vert$Diff$\vert$	0.00	0.00		0.00	0.00	0.09	0.09	0.09	0.09	0.09	0.09	0.09
	SD Diff	0.01	0.02		0.02	0.02	0.13	0.13	0.13	0.13	0.13	0.13	0.13
	$\vert$Diff$\vert$ > 1%/yr	0.00	0.00		0.01	0.01	0.38	0.39	0.39	0.39	0.39	0.39	0.39
	Not same	0.00	0.00		0.13	0.16	0.95	0.95	0.95	0.95	0.95	0.95	0.95
2010	Mean $\vert$Diff$\vert$	0.01	0.01	0.01		0.00	0.10	0.10	0.10	0.10	0.10	0.10	0.10
	SD Diff	0.05	0.05	0.08		0.00	0.16	0.16	0.16	0.16	0.16	0.16	0.16
	$\vert$Diff$\vert$ > 1%/yr	0.02	0.02	0.02		0.00	0.40	0.41	0.41	0.41	0.41	0.41	0.41
	Not same	0.02	0.02	0.02		0.04	0.96	0.95	0.95	0.95	0.95	0.95	0.95
2012	Mean $\vert$Diff$\vert$	0.01	0.01	0.01	0.00		0.10	0.11	0.11	0.11	0.11	0.11	0.11
	SD Diff	0.06	0.06	0.08	0.03		0.17	0.17	0.17	0.17	0.17	0.17	0.17
	$\vert$Diff$\vert$ > 1%/yr	0.04	0.04	0.04	0.02		0.41	0.42	0.42	0.42	0.42	0.42	0.42
	Not same	0.04	0.04	0.04	0.02		0.95	0.95	0.95	0.95	0.95	0.95	0.95
2014	Mean $\vert$Diff$\vert$	0.45	0.45	0.46	0.55	0.57		0.01	0.01	0.01	0.01	0.01	0.01
	SD Diff	0.66	0.67	0.69	0.90	0.92		0.04	0.04	0.04	0.04	0.04	0.04
	$\vert$Diff$\vert$ > 1%/yr	0.80	0.80	0.80	0.81	0.82		0.02	0.02	0.02	0.02	0.02	0.02
	Not same	0.80	0.80	0.80	0.81	0.82		0.25	0.25	0.27	0.29	0.29	0.29
2015	Mean $\vert$Diff$\vert$	0.45	0.45	0.46	0.55	0.57	0.06		0.00	0.00	0.00	0.00	0.00
	SD Diff	0.66	0.67	0.69	0.90	0.92	0.28		0.00	0.00	0.00	0.00	0.00
	$\vert$Diff$\vert$ > 1%/yr	0.80	0.80	0.80	0.81	0.82	0.14		0.00	0.00	0.00	0.00	0.00
	Not same	0.80	0.80	0.80	0.81	0.82	0.14		0.00	0.02	0.04	0.04	0.05
2016	Mean $\vert$Diff$\vert$	0.45	0.45	0.46	0.55	0.57	0.06	0.00		0.00	0.00	0.00	0.00
	SD Diff	0.66	0.67	0.69	0.90	0.92	0.28	0.00		0.00	0.00	0.00	0.00
	$\vert$Diff$\vert$ > 1%/yr	0.80	0.80	0.80	0.81	0.82	0.14	0.00		0.00	0.00	0.00	0.00
	Not same	0.80	0.80	0.80	0.81	0.82	0.14	0.00		0.02	0.04	0.04	0.05
2017	Mean $\vert$Diff$\vert$	0.45	0.45	0.46	0.55	0.57	0.06	0.00	0.00		0.00	0.00	0.00
	SD Diff	0.66	0.67	0.69	0.90	0.92	0.28	0.01	0.01		0.00	0.00	0.00
	$\vert$Diff$\vert$ > 1%/yr	0.80	0.80	0.80	0.81	0.82	0.14	0.00	0.00		0.00	0.00	0.00
	Not same	0.80	0.80	0.80	0.81	0.82	0.14	0.00	0.00		0.02	0.02	0.03
2019	Mean $\vert$Diff$\vert$											0.00	0.00
	SD Diff											0.00	0.00
	$\vert$Diff$\vert$ > 1%/yr											0.00	0.00
	Not same											0.00	0.01
2020	Mean $\vert$Diff$\vert$	0.45	0.45	0.46	0.55	0.57	0.06	0.00	0.00	0.00			0.00
	SD Diff	0.66	0.67	0.69	0.90	0.92	0.28	0.02	0.02	0.02			0.00
	$\vert$Diff$\vert$ > 1%/yr	0.80	0.80	0.80	0.81	0.82	0.15	0.01	0.01	0.01			0.00
	Not same	0.80	0.80	0.80	0.81	0.82	0.15	0.01	0.01	0.01			0.01
2021	Mean $\vert$Diff$\vert$	0.45	0.45	0.46	0.55	0.57	0.06	0.00	0.00	0.00		0.00	
	SD Diff	0.66	0.67	0.69	0.90	0.92	0.28	0.02	0.02	0.02		0.01	
	$\vert$Diff$\vert$ > 1%/yr	0.80	0.80	0.80	0.81	0.82	0.15	0.01	0.01	0.01		0.00	
	Not same	0.80	0.80	0.80	0.81	0.82	0.15	0.01	0.01	0.01		0.00	

		Vintage 1											
Vintage 2	Variable	2005	2006	2007	2010	2012	2014	2015	2016	2017	2019	2020	2021
B. HML factor													
2005	Mean $\lvert\text{Diff}\rvert$		0.11	0.12	0.15	0.14	0.21	0.23	0.23	0.29	0.30	0.30	0.32
	SD Diff		0.30	0.31	0.35	0.35	0.46	0.55	0.55	0.61	0.61	0.61	0.64
	$\lvert\text{Diff}\rvert > 1\%$/yr		0.32	0.33	0.42	0.43	0.49	0.50	0.50	0.59	0.62	0.61	0.62
	Not same		0.90	0.92	0.96	0.96	0.97	0.97	0.97	0.98	0.97	0.97	0.97
2006	Mean $\lvert\text{Diff}\rvert$	0.13		0.03	0.07	0.07	0.17	0.20	0.20	0.27	0.27	0.27	0.30
	SD Diff	0.23		0.08	0.18	0.18	0.36	0.44	0.44	0.51	0.51	0.51	0.54
	$\lvert\text{Diff}\rvert > 1\%$/yr	0.47		0.07	0.23	0.23	0.42	0.47	0.47	0.56	0.58	0.58	0.62
	Not same	0.47		0.43	0.83	0.81	0.95	0.96	0.96	0.97	0.97	0.97	0.97
2007	Mean $\lvert\text{Diff}\rvert$	0.17	0.10		0.06	0.06	0.17	0.20	0.20	0.27	0.28	0.28	0.30
	SD Diff	0.31	0.25		0.17	0.17	0.37	0.45	0.45	0.51	0.52	0.52	0.55
	$\lvert\text{Diff}\rvert > 1\%$/yr	0.54	0.34		0.20	0.20	0.41	0.47	0.47	0.55	0.58	0.58	0.62
	Not same	0.54	0.34		0.81	0.79	0.95	0.95	0.95	0.97	0.97	0.97	0.98
2010	Mean $\lvert\text{Diff}\rvert$	0.29	0.26	0.26		0.03	0.16	0.19	0.19	0.28	0.29	0.29	0.31
	SD Diff	0.44	0.39	0.41		0.05	0.35	0.43	0.43	0.52	0.55	0.55	0.58
	$\lvert\text{Diff}\rvert > 1\%$/yr	0.73	0.70	0.69		0.07	0.36	0.43	0.43	0.54	0.58	0.58	0.61
	Not same	0.73	0.70	0.69		0.59	0.94	0.94	0.94	0.96	0.96	0.97	0.97
2012	Mean $\lvert\text{Diff}\rvert$	0.30	0.27	0.27	0.21		0.15	0.18	0.18	0.27	0.29	0.29	0.31
	SD Diff	0.42	0.39	0.41	0.32		0.35	0.43	0.43	0.52	0.55	0.55	0.59
	$\lvert\text{Diff}\rvert > 1\%$/yr	0.75	0.73	0.71	0.63		0.31	0.38	0.38	0.51	0.54	0.54	0.60
	Not same	0.75	0.73	0.71	0.63		0.94	0.94	0.94	0.96	0.96	0.96	0.97
2014	Mean $\lvert\text{Diff}\rvert$	0.31	0.28	0.29	0.29	0.21		0.11	0.11	0.21	0.23	0.23	0.26
	SD Diff	0.45	0.42	0.45	0.44	0.32		0.28	0.28	0.41	0.43	0.43	0.48
	$\lvert\text{Diff}\rvert > 1\%$/yr	0.76	0.73	0.72	0.73	0.63		0.24	0.24	0.42	0.47	0.47	0.54
	Not same	0.76	0.73	0.72	0.73	0.63		0.76	0.76	0.85	0.93	0.94	0.96
2015	Mean $\lvert\text{Diff}\rvert$	0.39	0.38	0.39	0.40	0.35	0.69		0.00	0.12	0.14	0.14	0.18
	SD Diff	0.58	0.59	0.60	0.67	0.60	1.68		0.00	0.31	0.34	0.34	0.40
	$\lvert\text{Diff}\rvert > 1\%$/yr	0.79	0.79	0.78	0.79	0.73	0.70		0.00	0.22	0.28	0.29	0.36
	Not same	0.79	0.79	0.78	0.79	0.73	0.70		0.00	0.42	0.66	0.68	0.72
2016	Mean $\lvert\text{Diff}\rvert$	0.39	0.38	0.39	0.40	0.35	0.69	0.00		0.12	0.14	0.15	0.18
	SD Diff	0.58	0.59	0.60	0.67	0.60	1.68	0.01		0.31	0.34	0.34	0.40
	$\lvert\text{Diff}\rvert > 1\%$/yr	0.79	0.79	0.78	0.79	0.73	0.70	0.00		0.22	0.28	0.29	0.36
	Not same	0.79	0.79	0.78	0.79	0.73	0.70	0.00		0.42	0.66	0.68	0.72
2017	Mean $\lvert\text{Diff}\rvert$	0.87	0.86	0.88	1.05	1.04	1.10	0.53	0.53		0.04	0.04	0.11
	SD Diff	1.66	1.65	1.66	2.17	2.18	2.28	1.57	1.57		0.13	0.14	0.25
	$\lvert\text{Diff}\rvert > 1\%$/yr	0.84	0.83	0.83	0.84	0.80	0.76	0.29	0.30		0.10	0.11	0.32
	Not same	0.84	0.83	0.83	0.84	0.80	0.76	0.29	0.30		0.60	0.63	0.73
2019	Mean $\lvert\text{Diff}\rvert$											0.01	0.09
	SD Diff											0.03	0.21
	$\lvert\text{Diff}\rvert > 1\%$/yr											0.01	0.27
	Not same											0.23	0.69
2020	Mean $\lvert\text{Diff}\rvert$	0.88	0.87	0.89	1.07	1.06	1.15	0.61	0.62	0.14			0.09
	SD Diff	1.67	1.66	1.67	2.18	2.19	2.30	1.60	1.61	0.35			0.21
	$\lvert\text{Diff}\rvert > 1\%$/yr	0.85	0.84	0.84	0.84	0.82	0.82	0.48	0.48	0.37			0.27
	Not same	0.85	0.84	0.84	0.84	0.82	0.82	0.48	0.48	0.37			0.67
2021	Mean $\lvert\text{Diff}\rvert$	1.11	1.11	1.13	1.31	1.32	1.31	0.79	0.79	0.48		0.43	
	SD Diff	2.07	2.05	2.06	2.56	2.60	2.52	1.90	1.90	1.07		1.02	
	$\lvert\text{Diff}\rvert > 1\%$/yr	0.87	0.87	0.87	0.87	0.85	0.86	0.57	0.57	0.56		0.51	
	Not same	0.87	0.87	0.87	0.87	0.85	0.86	0.57	0.57	0.56		0.51	

		Vintage 1											
Vintage 2	Variable	2005	2006	2007	2010	2012	2014	2015	2016	2017	2019	2020	2021
C. SMB factor													
2005	Mean |Diff|		0.08	0.08	0.10	0.10	0.14	0.16	0.16	0.18	0.18	0.18	0.19
	SD Diff		0.19	0.19	0.22	0.22	0.33	0.36	0.36	0.37	0.37	0.37	0.38
	|Diff| > 1%/yr		0.23	0.25	0.30	0.31	0.37	0.38	0.39	0.45	0.49	0.49	0.50
	Not same		0.88	0.88	0.94	0.93	0.94	0.94	0.94	0.96	0.96	0.96	0.97
2006	Mean |Diff|	0.09		0.02	0.05	0.05	0.13	0.14	0.14	0.16	0.16	0.16	0.17
	SD Diff	0.18		0.04	0.12	0.12	0.30	0.32	0.32	0.33	0.33	0.33	0.34
	|Diff| > 1%/yr	0.36		0.05	0.16	0.17	0.32	0.32	0.32	0.40	0.43	0.43	0.44
	Not same	0.36		0.42	0.84	0.80	0.94	0.95	0.95	0.95	0.95	0.95	0.96
2007	Mean |Diff|	0.13	0.08		0.05	0.04	0.12	0.14	0.14	0.16	0.16	0.16	0.17
	SD Diff	0.23	0.18		0.12	0.12	0.30	0.32	0.32	0.33	0.33	0.33	0.34
	|Diff| > 1%/yr	0.45	0.31		0.13	0.14	0.32	0.33	0.33	0.40	0.42	0.42	0.45
	Not same	0.45	0.31		0.81	0.80	0.93	0.94	0.94	0.94	0.95	0.95	0.96
2010	Mean |Diff|	0.21	0.19	0.18		0.02	0.11	0.13	0.13	0.16	0.17	0.17	0.18
	SD Diff	0.32	0.28	0.28		0.03	0.28	0.30	0.30	0.33	0.35	0.35	0.36
	|Diff| > 1%/yr	0.65	0.62	0.60		0.03	0.28	0.30	0.30	0.40	0.42	0.42	0.44
	Not same	0.65	0.62	0.60		0.59	0.93	0.94	0.94	0.94	0.94	0.94	0.95
2012	Mean |Diff|	0.22	0.20	0.19	0.16		0.11	0.12	0.12	0.15	0.16	0.16	0.17
	SD Diff	0.33	0.29	0.29	0.24		0.28	0.30	0.30	0.32	0.34	0.34	0.36
	|Diff| > 1%/yr	0.67	0.64	0.63	0.57		0.25	0.27	0.27	0.38	0.41	0.41	0.42
	Not same	0.67	0.64	0.63	0.57		0.90	0.89	0.89	0.92	0.92	0.92	0.93
2014	Mean |Diff|	0.22	0.20	0.19	0.21	0.15		0.07	0.07	0.11	0.12	0.12	0.13
	SD Diff	0.32	0.29	0.30	0.31	0.24		0.18	0.18	0.22	0.24	0.24	0.26
	|Diff| > 1%/yr	0.67	0.65	0.62	0.65	0.53		0.18	0.18	0.31	0.34	0.34	0.36
	Not same	0.67	0.65	0.62	0.65	0.53		0.69	0.69	0.79	0.90	0.89	0.92
2015	Mean |Diff|	0.38	0.37	0.37	0.43	0.40	0.59		0.00	0.05	0.06	0.06	0.08
	SD Diff	0.68	0.70	0.69	0.79	0.77	1.47		0.00	0.13	0.15	0.16	0.19
	|Diff| > 1%/yr	0.76	0.75	0.75	0.78	0.73	0.71		0.00	0.14	0.17	0.18	0.21
	Not same	0.76	0.75	0.75	0.78	0.73	0.71		0.01	0.39	0.60	0.63	0.68
2016	Mean |Diff|	0.38	0.37	0.37	0.43	0.40	0.59	0.00		0.05	0.06	0.06	0.08
	SD Diff	0.68	0.70	0.69	0.79	0.77	1.47	0.01		0.13	0.15	0.16	0.19
	|Diff| > 1%/yr	0.76	0.75	0.75	0.78	0.73	0.71	0.00		0.15	0.18	0.18	0.21
	Not same	0.76	0.75	0.75	0.78	0.73	0.71	0.00		0.39	0.60	0.63	0.68
2017	Mean |Diff|	0.53	0.52	0.52	0.62	0.59	0.70	0.22	0.22		0.02	0.02	0.06
	SD Diff	0.95	0.94	0.93	1.16	1.12	1.59	0.72	0.72		0.09	0.09	0.14
	|Diff| > 1%/yr	0.80	0.79	0.78	0.81	0.77	0.74	0.25	0.25		0.05	0.05	0.17
	Not same	0.80	0.79	0.78	0.81	0.77	0.74	0.25	0.25		0.54	0.58	0.66
2019	Mean |Diff|											0.00	0.04
	SD Diff											0.02	0.11
	|Diff| > 1%/yr											0.01	0.13
	Not same											0.22	0.61
2020	Mean |Diff|	0.54	0.53	0.53	0.64	0.60	0.73	0.27	0.27	0.09			0.04
	SD Diff	0.95	0.95	0.94	1.17	1.13	1.61	0.76	0.76	0.23			0.11
	|Diff| > 1%/yr	0.80	0.80	0.79	0.81	0.78	0.78	0.39	0.39	0.28			0.12
	Not same	0.80	0.80	0.79	0.81	0.78	0.78	0.39	0.39	0.28			0.57
2021	Mean |Diff|	0.63	0.62	0.61	0.73	0.70	0.79	0.35	0.35	0.23		0.19	
	SD Diff	1.09	1.07	1.07	1.33	1.30	1.68	0.94	0.94	0.57		0.52	
	|Diff| > 1%/yr	0.82	0.81	0.81	0.83	0.80	0.80	0.45	0.45	0.43		0.35	
	Not same	0.82	0.81	0.81	0.83	0.80	0.80	0.45	0.45	0.43		0.35	

Table II
Stock alphas estimated using 2005 and 2021 factor vintages: Varying sample periods and estimation horizons

This table reports statistics for alphas of individual stocks estimated using 2005 and 2021 factor vintages. Three-factor alphas are estimated at the end of every calendar year using one, three, or five years of monthly data (Panels A, B, and C, respectively). The columns show results in subperiods and for the full sample. Reported are annualized means and standard deviations of alphas from the two vintages, as well as of the difference in alphas. |Difference| > 1% indicates the proportion of estimated alpha differences that exceed 1% in magnitude. The row labeled Lose significance shows the proportion of alphas that are significant in one vintage but not in the other.

	1926-1940	1941-1955	1956-1970	1971-1985	1986-2004	1926-2004
A. 1-year estimation horizon						
Mean 2005	3.42	0.18	-0.88	-1.14	2.05	0.74
Mean 2021	3.49	0.13	-0.82	-1.11	2.36	0.91
Mean difference	0.07	-0.05	0.07	0.03	0.31	0.17
SD 2005	50.5	28.6	39.6	54.5	74.5	62.8
SD 2021	50.7	28.9	39.2	54.7	76.4	64.0
SD difference	11.24	4.20	2.39	2.33	10.31	7.83
|Difference| > 1%	0.81	0.63	0.50	0.49	0.76	0.65
Lose significance	0.27	0.19	0.11	0.08	0.19	0.15
B. 3-year estimation horizon						
Mean 2005	4.00	0.27	-0.42	-0.03	3.76	2.02
Mean 2021	3.61	0.22	-0.52	-0.09	3.75	1.97
Mean difference	-0.39	-0.05	-0.10	-0.06	-0.01	-0.05
SD 2005	26.5	14.9	19.1	27.5	33.7	29.5
SD 2021	26.5	14.9	19.1	27.5	34.1	29.8
SD difference	3.45	1.08	0.77	0.81	3.90	2.91
|Difference| > 1%	0.65	0.20	0.13	0.15	0.48	0.34
Lose significance	0.18	0.11	0.07	0.05	0.14	0.10
C. 5-year estimation horizon						
Mean 2005	4.32	0.51	-0.31	0.43	4.42	2.56
Mean 2021	3.85	0.49	-0.44	0.37	4.24	2.41
Mean difference	-0.47	-0.01	-0.13	-0.06	-0.19	-0.15
SD 2005	19.7	11.5	14.3	19.7	24.6	21.6
SD 2021	19.8	11.5	14.3	19.7	24.8	21.7
SD difference	2.45	0.53	0.52	0.49	3.35	2.47
|Difference| > 1%	0.53	0.05	0.06	0.05	0.41	0.26
Lose significance	0.14	0.08	0.06	0.04	0.16	0.11

Table III
Stock alphas estimated using different factor vintages

This table reports statistics for alphas of individual stocks estimated using different factor vintages. Three-factor alphas are estimated at the end of every calendar year using five years of monthly data. Reported are annualized means and standard deviations of alphas from the two vintages (Panel A), as well as of the difference in alphas (B). |Diff| > 1% indicates the proportion of estimated alpha differences that exceed 1% in magnitude. The row labeled Lose signif shows the proportion of alphas that are significant in one vintage but not in the other.

		Vintage 1											
Vintage 2	Variable	2005	2006	2007	2010	2012	2014	2015	2016	2017	2019	2020	2021
A. Moments of alphas in different factor vintages													
	Mean	2.56	2.49	2.49	2.45	2.45	2.35	2.34	2.34	2.35	2.36	2.38	2.41
	SD	21.6	21.5	21.5	21.5	21.5	21.5	21.5	21.5	21.6	21.6	21.6	21.7
B. Statistics for differences in alphas between factor vintages													
2006	Mean	-0.07											
	SD	1.56											
	Lose signif	0.04											
	$\|Diff\| > 1\%$	0.05											
2007	Mean	-0.07	-0.02										
	SD	1.86	1.49										
	Lose signif	0.05	0.03										
	$\|Diff\| > 1\%$	0.05	0.03										
2010	Mean	-0.11	-0.10	-0.17									
	SD	1.89	2.54	3.75									
	Lose signif	0.06	0.06	0.09									
	$\|Diff\| > 1\%$	0.07	0.05	0.05									
2012	Mean	-0.11	-0.10	-0.17	-0.05								
	SD	1.89	2.53	3.75	1.87								
	Lose signif	0.06	0.06	0.09	0.03								
	$\|Diff\| > 1\%$	0.07	0.05	0.05	0.03								
2014	Mean	-0.21	-0.19	-0.24	-0.09	-0.09							
	SD	1.94	2.59	3.78	1.94	2.24							
	Lose signif	0.08	0.08	0.11	0.06	0.06							
	$\|Diff\| > 1\%$	0.11	0.10	0.10	0.11	0.10							
2015	Mean	-0.22	-0.20	-0.25	-0.09	-0.09	-0.02						
	SD	1.98	2.61	3.79	1.96	2.26	1.70						
	Lose signif	0.08	0.09	0.11	0.06	0.07	0.04						
	$\|Diff\| > 1\%$	0.12	0.11	0.11	0.12	0.11	0.04						
2016	Mean	-0.22	-0.20	-0.25	-0.09	-0.09	-0.02	-0.01					
	SD	1.98	2.61	3.79	1.97	2.26	1.78	0.92					
	Lose signif	0.08	0.09	0.11	0.06	0.07	0.04	0.02					
	$\|Diff\| > 1\%$	0.12	0.11	0.11	0.12	0.11	0.04	0.02					
2017	Mean	-0.21	-0.19	-0.24	-0.07	-0.07	0.00	0.01	0.01				
	SD	2.14	2.78	3.91	2.24	2.47	2.08	1.66	2.00				
	Lose signif	0.09	0.10	0.13	0.09	0.09	0.08	0.07	0.08				
	$\|Diff\| > 1\%$	0.21	0.20	0.21	0.21	0.23	0.19	0.17	0.17				
2019	Mean	-0.19	-0.17	-0.22	-0.05	-0.05	0.02	0.03	0.03	0.01			
	SD	2.15	2.79	3.92	2.26	2.50	2.10	1.70	2.14	1.82			
	Lose signif	0.09	0.11	0.13	0.09	0.09	0.09	0.08	0.09	0.04			
	$\|Diff\| > 1\%$	0.21	0.21	0.22	0.22	0.24	0.20	0.19	0.19	0.03			
2020	Mean	-0.18	-0.16	-0.21	-0.05	-0.04	0.03	0.04	0.04	0.02	0.02		
	SD	2.14	2.79	3.92	2.26	2.49	2.10	1.70	2.14	1.82	1.76		
	Lose signif	0.09	0.11	0.13	0.09	0.09	0.09	0.08	0.09	0.04	0.03		
	$\|Diff\| > 1\%$	0.21	0.20	0.21	0.22	0.23	0.20	0.19	0.19	0.03	0.02		
2021	Mean	-0.15	-0.12	-0.16	0.02	0.01	0.08	0.09	0.09	0.07	0.07	0.06	
	SD	2.47	3.05	4.09	2.57	2.80	2.43	2.09	2.46	2.11	2.28	1.55	
	Lose signif	0.11	0.12	0.14	0.11	0.11	0.11	0.10	0.11	0.08	0.07	0.05	
	$\|Diff\| > 1\%$	0.26	0.25	0.26	0.28	0.29	0.26	0.25	0.25	0.17	0.15	0.15	

Table IV
Stock alphas estimated using 2005 and 2021 factor vintages: Varying stock characteristics

This table reports statistics for alphas of individual stocks estimated using 2005 and 2021 factor vintages. Three-factor alphas are estimated at the end of every calendar year using five years of monthly data. Stocks are grouped into quintiles on the basis of characteristics shows in panel headings using most recent characteristic available prior to the alpha estimation window. Reported are annualized means and standard deviations of alphas from the two vintages, as well as of the difference in alphas. |Difference| > 1% indicates the proportion of estimated alpha differences that exceed 1% in magnitude. The row labeled Lose significance shows the proportion of alphas that are significant in one vintage but not in the other.

	Low	Q2	Med	Q4	High
A. Market equity					
Mean 2005	9.50	2.82	0.92	0.36	0.24
Mean 2021	9.30	2.70	0.83	0.17	-0.03
Mean difference	-0.20	-0.12	-0.09	-0.19	-0.27
SD 2005	24.86	21.62	19.82	17.24	13.64
SD 2021	24.94	21.73	19.90	17.25	13.63
SD difference	3.30	2.74	2.44	2.08	1.64
|Difference| > 1%	0.29	0.25	0.24	0.23	0.21
Lose significance	0.12	0.11	0.09	0.10	0.13
B. Book-to-market ratio					
Mean 2005	2.64	2.20	2.41	3.16	6.23
Mean 2021	2.64	2.09	2.22	2.90	5.97
Mean difference	0.00	-0.12	-0.19	-0.26	-0.27
SD 2005	24.56	20.86	19.06	18.63	21.56
SD 2021	24.71	20.97	19.16	18.65	21.54
SD difference	3.20	2.59	2.51	2.50	3.08
|Difference| > 1%	0.30	0.27	0.26	0.26	0.27
Lose significance	0.10	0.11	0.12	0.14	0.14
C. Asset growth					
Mean 2005	7.32	3.10	2.83	2.33	1.07
Mean 2021	7.16	2.87	2.60	2.11	1.00
Mean difference	-0.16	-0.23	-0.23	-0.23	-0.07
SD 2005	25.04	18.25	17.01	18.56	22.73
SD 2021	25.12	18.25	17.06	18.58	22.82
SD difference	3.64	2.54	2.29	2.40	2.91
|Difference| > 1%	0.30	0.25	0.25	0.26	0.29
Lose significance	0.13	0.13	0.13	0.13	0.09
D. Profitability					
Mean 2005	3.31	1.60	2.35	3.65	5.72
Mean 2021	3.00	1.43	2.28	3.50	5.59
Mean difference	-0.31	-0.17	-0.07	-0.15	-0.13
SD 2005	20.98	20.37	21.34	20.88	21.98
SD 2021	21.16	20.41	21.38	20.92	22.02
SD difference	2.92	2.48	2.70	2.91	2.92
|Difference| > 1%	0.29	0.25	0.27	0.27	0.28
Lose significance	0.14	0.13	0.11	0.11	0.10

Table V
Stock portfolio alphas estimated using different factor vintages

This table reports statistics for alphas of characteristic-sorted value-weighted decile portfolios from Kenneth French's website. The earliest vintage of both the factors and the portfolios ('vintage 1') is compared to the latest vintage of each ('vintage 2'). The earliest vintages of factors and portfolios sorted on market equity and book-to-market ratio is 2005. For runup portfolios, the earliest vintage is 2007, and for all other portfolios it is 2015. The latest vintage is 2021 for factors and all portfolios. Three-factor alphas are estimated at the end of every calendar year using five years of monthly data. Reported are annualized means and standard deviations of alphas from the two vintages, as well as of the difference in alphas. |Difference| > 1% indicates the proportion of estimated alpha differences that exceed 1% in magnitude. The row labeled Lose significance shows the proportion of alphas that are significant in one vintage but not in the other.

	Mkt equity	BM ratio	Runup	Profitability	Investment
Mean vintage 1	0.09	0.03	-0.49	-0.24	0.71
Mean vintage 2	-0.05	-0.13	-0.69	-0.30	0.55
Mean difference	-0.14	-0.16	-0.20	-0.06	-0.17
SD vintage 1	2.15	2.90	5.43	3.12	2.91
SD vintage 2	2.08	3.06	5.46	3.12	2.80
SD difference	0.90	1.64	0.93	1.75	0.45
|Difference| > 1%	0.14	0.33	0.11	0.35	0.05
Lose significance	0.51	0.54	0.11	0.56	0.24

	Accruals	Beta	Issuance	Variance	Residual var
Mean vintage 1	0.67	0.45	-0.22	0.00	-0.16
Mean vintage 2	0.54	0.27	-0.35	-0.15	-0.26
Mean difference	-0.12	-0.18	-0.13	-0.15	-0.11
SD vintage 1	3.40	2.83	4.02	4.81	4.92
SD vintage 2	3.15	2.86	3.92	4.80	4.85
SD difference	1.31	0.48	0.55	0.64	0.89
|Difference| > 1%	0.20	0.07	0.06	0.08	0.15
Lose significance	0.33	0.19	0.15	0.08	0.18

Table VI
Mutual fund alphas estimated using 2005 and 2021 factor vintages: Varying sample periods and estimation horizons

This table reports statistics for alphas of individual mutual funds estimated using 2005 and 2021 factor vintages. Three-factor alphas are estimated at the end of every calendar year using one, three, or five years of monthly data (Panels A, B, and C, respectively). The columns show results in subperiods and for the full sample. Reported are annualized means and standard deviations of alphas from the two vintages, as well as of the difference in alphas. |Difference| > 1% indicates the proportion of estimated alpha differences that exceed 1% in magnitude. The row labeled Lose significance shows the proportion of alphas that are significant in one vintage but not in the other.

	1980s	1990s	2000s	1980-2004
A. 1-year estimation horizon				
Mean 2005	1.53	-2.10	-2.33	-1.80
Mean 2021	1.35	-2.44	-2.16	-1.90
Mean difference	-0.19	-0.34	0.18	-0.09
SD 2005	9.73	15.78	11.57	13.49
SD 2021	9.80	16.27	11.75	13.81
SD difference	0.57	2.12	2.26	2.09
|Difference| > 1%	0.08	0.54	0.48	0.46
Lose significance	0.18	0.31	0.34	0.31
B. 3-year estimation horizon				
Mean 2005	1.56	-1.96	-0.97	-1.10
Mean 2021	1.31	-2.07	-0.58	-0.96
Mean difference	-0.25	-0.11	0.39	0.14
SD 2005	6.64	7.75	7.61	7.64
SD 2021	6.67	7.72	7.78	7.72
SD difference	0.31	0.88	1.20	1.06
|Difference| > 1%	0.01	0.16	0.28	0.21
Lose significance	0.12	0.22	0.21	0.21
C. 5-year estimation horizon				
Mean 2005	1.41	-1.37	-1.18	-1.01
Mean 2021	1.14	-1.48	-0.88	-0.90
Mean difference	-0.27	-0.11	0.30	0.12
SD 2005	5.58	6.08	6.10	6.09
SD 2021	5.58	6.09	6.27	6.19
SD difference	0.27	0.58	1.18	0.99
|Difference| > 1%	0.00	0.08	0.28	0.19
Lose significance	0.12	0.19	0.28	0.24

Table VII
Mutual fund alphas estimated using different factor vintages

This table reports statistics for alphas of individual mutual funds estimated using different factor vintages. Three-factor alphas are estimated in every calendar year using monthly data. Reported are annualized means and standard deviations of alphas from the two vintages (Panel A), as well as of the difference in alphas (B). |Diff| > 1% indicates the proportion of estimated alpha differences that exceed 1% in magnitude. The row labeled Lose signif shows the proportion of alphas that are significant in one vintage but not in the other.

		Vintage 1											
Vintage 2	Variable	2005	2006	2007	2010	2012	2014	2015	2016	2017	2019	2020	2021
A. Moments of alphas in different factor vintages													
	Mean	-1.80	-1.84	-1.83	-1.86	-1.85	-1.90	-1.88	-1.88	-2.06	-2.05	-2.06	
	SD	13.5	13.5	13.5	13.5	13.5	13.6	13.5	13.5	13.8	13.8	13.8	
B. Statistics for differences in alphas between factor vintages													
2006	Mean	-0.03											
	SD	0.25											
	Lose signif	0.03											
	|Diff| > 1%	0.01											
2007	Mean	-0.02	0.01										
	SD	0.28	0.16										
	Lose signif	0.04	0.02										
	|Diff| > 1%	0.01	0.00										
2010	Mean	-0.06	-0.02	-0.02									
	SD	0.49	0.34	0.35									
	Lose signif	0.05	0.05	0.06									
	|Diff| > 1%	0.05	0.03	0.02									
2012	Mean	-0.05	-0.02	-0.02	0.00								
	SD	0.46	0.30	0.32	0.26								
	Lose signif	0.06	0.05	0.05	0.05								
	|Diff| > 1%	0.04	0.02	0.01	0.01								
2014	Mean	-0.10	-0.03	-0.08	0.18	0.18							
	SD	0.84	0.78	0.77	1.03	1.02							
	Lose signif	0.19	0.19	0.21	0.23	0.25							
	|Diff| > 1%	0.21	0.18	0.18	0.28	0.31							
2015	Mean	-0.08	-0.02	-0.06	0.17	0.18	0.01						
	SD	0.81	0.79	0.77	1.03	1.02	0.21						
	Lose signif	0.19	0.19	0.21	0.22	0.25	0.03						
	|Diff| > 1%	0.19	0.17	0.17	0.27	0.30	0.00						
2016	Mean	-0.08	-0.02	-0.06	0.17	0.18	0.01	0.00					
	SD	0.81	0.79	0.77	1.03	1.02	0.21	0.00					
	Lose signif	0.19	0.19	0.21	0.22	0.25	0.03	0.00					
	|Diff| > 1%	0.19	0.17	0.17	0.27	0.30	0.00	0.00					
2017	Mean	-0.26	-0.26	-0.32	-0.23	-0.22	-0.24	-0.25	-0.25				
	SD	1.68	1.76	1.75	1.99	1.94	1.79	1.74	1.71				
	Lose signif	0.24	0.26	0.29	0.31	0.32	0.24	0.22	0.23				
	|Diff| > 1%	0.39	0.40	0.42	0.46	0.47	0.35	0.34	0.33				
2019	Mean	-0.25	-0.26	-0.32	-0.23	-0.22	-0.24	-0.24	-0.24	0.01			
	SD	1.69	1.77	1.75	1.99	1.95	1.83	1.78	1.74	0.15			
	Lose signif	0.25	0.26	0.29	0.31	0.32	0.25	0.23	0.23	0.03			
	|Diff| > 1%	0.39	0.40	0.42	0.46	0.47	0.36	0.34	0.33	0.00			
2020	Mean	-0.25	-0.26	-0.32	-0.23	-0.22	-0.23	-0.24	-0.24	0.01	0.00		
	SD	1.69	1.76	1.74	1.99	1.94	1.82	1.77	1.73	0.16	0.07		
	Lose signif	0.25	0.26	0.29	0.31	0.32	0.25	0.23	0.23	0.03	0.01		
	|Diff| > 1%	0.38	0.40	0.42	0.46	0.47	0.36	0.35	0.34	0.00	0.00		
2021	Mean	-0.09	-0.07	-0.15	-0.22	-0.21	-0.21	-0.22	-0.22	0.03	0.01	0.01	
	SD	2.09	2.16	2.11	2.98	2.91	2.78	2.69	2.63	1.62	1.55	1.52	
	Lose signif	0.31	0.33	0.35	0.37	0.38	0.31	0.28	0.28	0.16	0.16	0.15	
	|Diff| > 1%	0.46	0.47	0.48	0.52	0.53	0.42	0.41	0.39	0.20	0.18	0.18	

Table VIII
Mutual fund alphas estimated using 2005 and 2021 factor vintages: Varying fund characteristics

This table reports statistics for alphas of individual mutual funds estimated using 2005 and 2021 factor vintages. Three-factor alphas are estimated in every calendar year using monthly data. Funds are grouped into quintiles on the basis of characteristics shows in panel headings using most recent characteristic available prior to the alpha estimation window. Reported are annualized means and standard deviations of alphas from the two vintages, as well as of the difference in alphas. |Difference| > 1% indicates the proportion of estimated alpha differences that exceed 1% in magnitude. The row labeled Lose significance shows the proportion of alphas that are significant in one vintage but not in the other.

	Low	Q2	Med	Q4	High
A. Fund size					
Mean 2005	-0.48	-1.76	-2.17	-2.24	-2.40
Mean 2021	-0.54	-1.82	-2.28	-2.37	-2.52
Mean difference	-0.06	-0.05	-0.10	-0.12	-0.12
SD 2005	15.03	14.22	13.49	12.86	11.53
SD 2021	15.35	14.61	13.82	13.15	11.81
SD difference	2.18	2.15	2.09	2.04	2.01
|Difference| > 1%	0.47	0.47	0.46	0.47	0.46
Lose significance	0.26	0.30	0.30	0.32	0.38
B. Size factor exposure					
Mean 2005	-1.00	-2.10	-3.60	-2.22	-1.50
Mean 2021	-1.02	-2.22	-3.75	-2.24	-1.59
Mean difference	-0.01	-0.12	-0.14	-0.02	-0.09
SD 2005	9.73	11.12	14.21	15.18	13.49
SD 2021	9.87	11.43	14.79	15.69	13.49
SD difference	1.41	1.79	2.17	2.36	2.37
|Difference| > 1%	0.36	0.40	0.46	0.51	0.56
Lose significance	0.37	0.34	0.32	0.29	0.31
C. Value factor exposure					
Mean 2005	-1.28	-2.60	-2.33	-2.06	-2.16
Mean 2021	-1.49	-2.78	-2.39	-2.04	-2.12
Mean difference	-0.21	-0.18	-0.06	0.02	0.04
SD 2005	14.29	12.06	11.70	12.76	13.66
SD 2021	14.31	12.41	12.13	13.24	14.05
SD difference	2.28	1.87	1.87	1.98	2.22
|Difference| > 1%	0.55	0.42	0.38	0.43	0.51
Lose significance	0.31	0.32	0.30	0.36	0.34

Table IX
GRS F-test statistics from different vintages of factors and portfolios

This table reports F-statistics from the GRS tests using different vintages of the three factors and portfolios. All tests use data common to all vintages: 07/1926-08/2005 for portfolios sorted on size and book-to-market, and 07/1926-05/2005 for industry portfolios. Highlighted cells indicate contemporaneous vintages of factors and portfolios.

Portfolio	Factor vintage											
vintage	2005	2006	2007	2010	2012	2014	2015	2016	2017	2019	2020	2021
A. 25 portfolios sorted on size and book-to-market ratio												
2005	2.91	2.93	2.92	2.94	2.93	2.94	2.96	2.97	3.00	2.98	2.97	2.98
2006	3.30	3.34	3.33	3.37	3.35	3.37	3.42	3.42	3.47	3.44	3.43	3.45
2007	3.19	3.22	3.21	3.27	3.24	3.27	3.32	3.32	3.38	3.34	3.33	3.35
2010	3.08	3.10	3.09	3.13	3.11	3.12	3.15	3.15	3.20	3.18	3.17	3.19
2012	3.09	3.11	3.10	3.14	3.12	3.14	3.18	3.18	3.23	3.20	3.20	3.22
2014	2.83	2.82	2.82	2.81	2.81	2.81	2.81	2.81	2.83	2.82	2.82	2.83
2015	2.69	2.67	2.68	2.66	2.66	2.66	2.64	2.64	2.66	2.65	2.64	2.67
2016	2.69	2.67	2.68	2.66	2.66	2.67	2.64	2.64	2.66	2.65	2.64	2.67
2017	2.77	2.78	2.79	2.75	2.76	2.72	2.69	2.69	2.71	2.70	2.70	2.74
2020	2.75	2.74	2.76	2.73	2.73	2.70	2.67	2.67	2.70	2.69	2.69	2.72
2021	2.55	2.55	2.57	2.53	2.53	2.49	2.46	2.46	2.48	2.47	2.47	2.50
B. 17 Industry portfolios												
2005	3.54	3.58	3.58	3.64	3.63	3.48	3.53	3.53	3.59	3.56	3.55	3.56
2006	3.45	3.49	3.49	3.55	3.54	3.41	3.45	3.46	3.51	3.48	3.47	3.48
2007	3.64	3.69	3.69	3.75	3.73	3.59	3.64	3.64	3.70	3.67	3.67	3.67
2010	4.30	4.35	4.35	4.41	4.40	4.28	4.35	4.35	4.40	4.38	4.37	4.37
2012	4.28	4.33	4.33	4.40	4.38	4.26	4.34	4.34	4.39	4.37	4.36	4.36
2014	4.12	4.17	4.17	4.23	4.21	4.12	4.20	4.20	4.25	4.22	4.21	4.22
2015	4.09	4.13	4.13	4.19	4.18	4.10	4.19	4.19	4.23	4.21	4.20	4.19
2016	4.09	4.13	4.13	4.19	4.18	4.10	4.19	4.19	4.23	4.21	4.20	4.19
2017	4.09	4.13	4.13	4.19	4.18	4.10	4.19	4.19	4.23	4.21	4.20	4.19
2020	4.09	4.13	4.13	4.19	4.18	4.11	4.19	4.19	4.24	4.21	4.20	4.20
2021	4.10	4.13	4.13	4.19	4.18	4.11	4.19	4.19	4.24	4.21	4.20	4.20

Table X
Tests of Equality of Squared Sharpe Ratios

This table reports results of pairwise tests of equality of the squared Sharpe ratios of the three-factor model with different factor vintages. Panel A reports the difference between the bias-adjusted sample squared Sharpe ratios of the models based on vintages showing in columns and rows. Panel B shows the associated p-values.

	Vintage 1										
Vintage 2	2006	2007	2010	2012	2014	2015	2016	2017	2019	2020	2021
A. Differences in squared Sharpe ratio											
2005	0.000	0.000	0.001	0.000	0.001	0.001	0.001	0.002	0.002	0.002	0.002
2006		0.000	0.001	0.000	0.001	0.001	0.001	0.002	0.002	0.002	0.002
2007			0.001	0.000	0.001	0.001	0.001	0.002	0.002	0.001	0.002
2010				0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001
2012					0.001	0.001	0.001	0.001	0.001	0.001	0.001
2014						0.000	0.000	0.001	0.000	0.000	0.000
2015							0.000	0.000	0.000	0.000	0.000
2016								0.000	0.000	0.000	0.000
2017									0.000	0.000	0.000
2019										0.000	0.000
2020											0.000
B. p-values											
2005	0.963	0.970	0.406	0.536	0.161	0.145	0.144	0.110	0.148	0.155	0.158
2006		0.673	0.060	0.118	0.064	0.064	0.063	0.059	0.080	0.086	0.099
2007			0.060	0.132	0.071	0.076	0.075	0.066	0.091	0.097	0.109
2010				0.132	0.261	0.237	0.234	0.163	0.259	0.275	0.263
2012					0.185	0.172	0.170	0.129	0.202	0.215	0.215
2014						0.740	0.731	0.384	0.658	0.699	0.594
2015							0.206	0.369	0.767	0.827	0.677
2016								0.376	0.776	0.837	0.684
2017									0.260	0.223	0.707
2019										0.585	0.767
2020											0.694