

Time Series Variation in the Factor Zoo*

Hendrik Bessembinder

W.P. Carey School of Business, Arizona State University

Aaron Burt

Michael F. Price College of Business, University of Oklahoma

Christopher Hrdlicka

Michael G. Foster School of Business, University of Washington

Initial Draft: December 2021

This Draft: December 2022

Abstract

The ability of three- to six-factor models to explain the cross-section of stock returns varies substantially over time, providing scope for time-varying numbers of additional factors. We show that additional factors are relevant and non-redundant, as out-of-sample Sharpe ratios formed from principal components of factors identified in-sample are economically substantive and continue to increase up to more than twenty factor principal components. The numbers of significant factors are strongly related to variation in economic conditions and measures of diversity in firm characteristics. These results suggest that time variation in the number of significant factors reflects time-varying economic complexity.

*We are grateful to Turan Bali, Michael Brennan, Stephen Brown, Andrew Chen, Tarun Chordia, Campbell Harvey, Theis Jensen, Bryan Kelly, Scott Murray, Stefan Nagel, Jeffrey Pontiff, Seth Pruitt, and seminar participants at the University of Washington and the University of Oklahoma. Author emails are hb@asu.edu, aaronburt@ou.edu, and hrdlicka@uw.edu, respectively.

In the beginning there was chaos. . . Then came the CAPM. . . Then anomalies erupted, and there was chaos again...Fama and French (1993, 1996) brought order once again...Alas, the world again is descending into chaos.....I did not say it will be easy! But we must address the factor zoo.

John Cochrane, AFA Presidential Address, 2011

1. Introduction.

The literature has identified hundreds of empirical variables, including firm characteristics and “factors” constructed as returns to long-short portfolios, that appear to have significant explanatory power for the cross-section of stock returns.¹ However, as the preceding quotation from the former President of the American Finance Association illustrates, there is a widespread perception that finance researchers have collectively identified too many factors. Indeed, foundational asset pricing models such as the CAPM or the consumption-based CAPM imply that a single factor should be sufficient to explain the cross-section of returns if it is measured correctly.

While we do not resolve the question of whether researchers have collectively identified too many factors, we posit and provide empirical evidence that the number of economically relevant factors varies over time as a function of firm complexity and economic conditions. We use rolling sixty-month specifications to allow for time variation, and show that the number of factors with significant explanatory power for the cross-section of returns varies substantially over time and is related to measures of firm diversity and economic complexity. We examine out-of-sample Sharpe ratios for portfolios constructed from the principal components of the factors that are significant in-sample. These Sharpe ratios are economically large, and on average continue to increase as the number of factor principal

¹ The literature has not always been consistent in usage of the terms “characteristic” and “factor.” To be precise, we use the term “characteristic” to refer to a firm-level attribute, such as firm size or profitability, and we use the word “factor” to refer to returns on a long-short portfolio. More specifically, each factor is the time series of returns on a portfolio that is long a set of stocks selected with either left or right tail outcomes on a given characteristic, e.g., firms of small size or high profitability, and short a set of stocks with outcomes in the opposite tail, e.g., stocks of large size or low profitability. We do not herein use the term factor to refer to outcomes obtained by the statistical technique of factor analysis.

components increases to twenty or more, implying that the significant factors are not redundant of each other.

Variation in the economic relevance of individual factors can arise because of changes in individual factors' return premium per unit of factor risk or in the quantity of factor risk (i.e., factor betas).² It is broadly recognized that the return premia associated with canonical factors such as firm size or value (market-to-book ratio) have varied substantively over time.³ Haddad, Kozak, and Santosh (2020) broaden this avenue of inquiry by showing that time variation in the top few principal components associated with a set of fifty "anomaly" portfolios can be identified based on the market-to-book ratios of the factors themselves. Our results are consistent with Haddad, Kozak, Santosh (2020) in that both their study and ours document that conditional return premia can be substantially higher than unconditional premia. However, we focus on time variation in the relevance of individual factors rather than a constant set of principal components formed from a fixed set of factors. We document that time variation in the number of significant factor premia is more the rule than the exception across a large sample of over two hundred factors, demonstrate the economic relevance and non-redundancy of the significant factors both in- and out-of-sample, and assess the economic determinants of variation in the number of significant factors.

Cochrane (2011) observes that essentially all variation in price-to-dividend ratios is attributable to changes in discount rates, i.e., expected returns. If factor models determine expected returns it follows that variation in discount rates is attributable to time variation in interest rates and factor return premia. There are numerous reasons that factor premia can vary through time. In contrast to the assumptions of representative agent models, investors are diverse in terms of both their sophistication and their investment objectives, which allows that the identity of the marginal investor can differ across stocks and, in a given stock, can change over time. Some individual investors may seek to form mean-variance

² We do not take a stand as to whether the return premium associated with a given factor arises because of investor aversion to undesirable factor outcomes, mispricing in the face of barriers to arbitrage, or a combination thereof.

³ See, for example, , e.g. Conrad and Kaul (1988), Ferson and Harvey (1991), Cochrane (1999), van Dijk (2011) and Ehsani and Linnainmaa (2021).

efficient portfolios, while others seek out positive skewness or “lottery” payoffs, and yet others trade in response to comments on discussion boards such as “WallStreetBets.”⁴ Some may hold positions for long periods, others periodically rebalance to target weights, and yet others trade episodically in response to wealth shocks, opportunities to provide liquidity or to correct perceived mispricings. Betermier, Calvet, Knüpfer, and Kvaerner (2021) show that the cross-section of expected stock returns depends in part on the proportion of individual investors that are younger as well as the proportion that are wealthier. Some investors trade directly, while others delegate portfolio decisions to professional managers, whose objectives can differ from those of their investors due to agency issues arising, for example, from specific compensation plans (e.g., Kashyap, Kovrijnykh, Li and Pavlova, 2021) or as a function of career horizons. Further, the trades of professional investors can depend on considerations such as the funding liquidity of their employing firms, and return premia have been shown to also depend also on the leverage of financial sector firms.⁵

Investor learning may also be relevant. Brennan (1998) for example, explores how investors’ utility-maximizing portfolio decisions depend on their stock market experience, and Pastor and Veronesi (2009) explore how estimates of firm values evolve as investors learn about firms’ growth prospects. Martin and Nagel (2022) show how investors’ incorrect priors regarding parameters of the return distribution in combination with complexity in the form of a large number of relevant firm characteristics allows for out-of-sample return predictability as investors learn about the true distribution, even if the underlying economic structure is stable. Chinco, Neuhierl, and Weber (2021), also assuming constant underlying parameters, assess how active traders’ optimally combine their prior assessments with the emergence of empirical evidence to determine which signals they will attempt to trade on. We reason that a dynamic economic environment with time-varying parameters only heightens the importance of investor

⁴ Recent studies documenting the diversity of individual trading approaches include Barber, Huang, Odean, and Schwarz (2021), Chen, Kumar, and Zhang (2021) and Bali, Brown, Murray Tang (2017).

⁵ See, for example, Kojien, and Yogo, (2019), He, Kelly and Manela (2017), Tobias, Etula, and Muir (2014) and He and Krishnamurthy (2013).

learning as a barrier to arbitrage, thereby providing scope for factor premia to persist for a time even after they become economically relevant.

In addition to the diversity of investors in terms of their experience, knowledge, strategies, and objectives, the economic characteristics of newly listed firms can differ from those of existing firms, as shown by Campbell (2001), Fama and French (2004), and Kahle and Stulz (2017). We construct a measure of cross-sectional diversity in the observable characteristics that are collectively known to be related to expected returns, and show the number of factors that are significant in explaining stock returns increases with such diversity.

At first glance, our findings may appear to contrast with conclusions reached by earlier authors. Kelly, Pruitt, and Su (2019) present evidence indicating that as few as five latent factors identified by the technique of Instrumented Principal Components Analysis can outperform existing factor models and lead to insignificant alpha estimates. Kozak, Nagel, and Santosh (2020) report that a stochastic discount factor formed from a small number of factor principal components performs well in terms of the pricing model's out-of-sample R-squared statistic, and Kozak, Nagel, and Santosh (2018) report that factor principal components beyond the first few do not contribute meaningful to out-of-sample Sharpe ratios. However, our rolling estimation approach allows for flexible time variation in factor return premia in a manner that these studies do not.⁶ We apply our methods to the same data studied by Kozak, Nagel and Santosh (2018) and show that the divergence of our outcomes from theirs is attributable to our allowance for time variation in factor return premia. Further, we use the data and programs posted by Kozak, Nagel and Santosh (2020) to show that, even though they do not allow for time variation in factor premia, out-of-sample Sharpe ratios implied by their analysis continue to increase as more factor principal components are used to form portfolios, even while the R-squared statistic on which they focus rises only modestly.

Our analysis is also related to that of Chinco, Neuhierl, and Weber (2021), who focus on an economic environment characterized by a constant set of predictive variables and associated parameters,

⁶ Kelly, Pruitt, and Su (2019) do allow for factor return premia to vary over time, but in a less flexible manner, as factor alphas and betas are constrained to be time-invariant linear functions of observable characteristics.

e.g., factor premia, where the number of variables optimally included in a trading strategy varies over time as a function of agents' priors from evolving sample evidence. They further show that such time variation in priors is largely unrelated to changes in the macroeconomic environment. In contrast, we focus on time variation in the premia associated with specific factors, and directly link such time variation to the dynamic complexity of the economic environment. Our analysis of time varying relations between factor outcomes and the cross-section of stock returns also complements that of Farmer, Lawrence Schmidt, and Timmermann (2022), who show that the ability of macroeconomic variables such as the interest rate term structure to forecast aggregate market returns is not time-invariant, but rather is concentrated in adjacent "pockets" of time.

Our contributions relative to the related literature include the following. First, we focus attention on the substantial time variation in the number of significant factors, and document that this variation has significant out-of-sample predictive power for portfolio Sharpe ratios. Second, we show that variation in the number of economically relevant factors is significantly related to a set of variables measuring the complexity of firms and economic conditions. Third, we focus on the cross-sectional variation in mean returns to individual stocks rather than focusing only on characteristic-sorted portfolios. We confirm that, when studying mean returns to size and book-to-market portfolios, the widely used three- to six-factor models outperform the CAPM, a finding which may contribute to a perception that a few factors are sufficient. However, when considering individual stocks, the CAPM outperforms these models. These results imply scope for additional relevant factors beyond those included in the widely used three- to six-factor models. We further show that out-of-sample Sharpe ratios for portfolios constructed from the principal components of factors that are significant in sample are economically large, and the number of principal components that yields the largest out-of-sample Sharpe ratios varies over time and often includes more than twenty principal components. These results imply that factors beyond the first few are not simply redundant.

To assess the scope for multiple factors to explain the cross-section of individual stock returns, we also estimate *firm-specific* alphas from rolling sixty-month market-model regressions. We then study

the cross-sectional standard deviation of the resulting firm-level alpha estimates for each month. Time periods with more dispersion in firm-level CAPM alphas indicate greater scope for factors beyond the market to have explanatory power for expected stock returns. We document that the number of factors with significant alphas is indeed positively related to the standard deviation of firm-level CAPM alpha estimates, even while the number of significant factors is not related to the average standard error of alpha estimates or average idiosyncratic volatility. These results support the reasoning that the number of significant factors is related to the extent to which expected stock returns are left unexplained by the CAPM.

We also contribute to the literature that studies whether factors have significant explanatory power during time periods outside those studied by original authors. This literature typically focuses on outcomes during authors' original sample periods versus results obtained during subsequent periods, and as such do not allow for time variation within these periods. In contrast, using rolling 60-month windows, we show that many factors are statistically significant in periods both before and after the range of data studied by the original authors. We recognize that time variation in the estimated premia associated with specific factors can arise due to three alternative hypotheses. First, this outcome could simply reflect random noise in a stable economic environment. That is, a factor with a constant, but economically modest, true premium could be associated with significant estimates during some intervals and insignificant estimates during other intervals due to random variation (Jensen, Kelly and Pedersen, 2021). Second, it could arise due to collective data mining as others have argued. Third, it could reflect time variation in the magnitude of actual factor premia.

We differentiate between these possibilities with two novel empirical approaches. First, we simulate the distribution of the number and average length of "spells" of factor significance under the null hypothesis that true factor premia are time-invariant, and show that the observed sample outcomes are exceptionally unlikely under the null hypothesis. Second, we assess the extent to which variation in the number of significant factors is related to measures of changes in the economic environment. We document that the number of significant factors is related to a recession indicator variable, interest rates,

the percentage of firms that pay dividends, mean institutional ownership rates, and an economic complexity index, and is particularly strongly related to the number of firms that are publicly listed. The link to the number of publicly-listed firms is subsumed, in turn, by a measure of cross-sectional diversity in firm characteristics. Periods with increased dispersion in characteristics across firms are those where the underlying firms themselves are more differentiated. The positive and significant relation between the number of significant factors and such measures of economic complexity support the conclusion that factor premia themselves vary over time.

On balance, our findings suggest that a time-varying number of non-redundant factors are required to price the cross-section of returns as the economy evolves dynamically and diverse firms are listed and delisted. Further, in a dynamic economy a factor can be significant in explaining returns during some periods but not others. This suggests a degree of caution in interpreting the results of existing out-of-sample tests, as insignificant out-of-sample outcomes need not imply that the factor was unpriced in the original sample period, will remain unpriced, or that arbitrageurs permanently eliminated a mispricing after becoming aware of it.

However, accommodating such time variation may also provide additional scope for specification searches or other sources of bias. We have attempted avoid amplifying any such bias in this study by focusing only on previously-identified factors and the sixty-month estimation window that is used in many prior studies and that corresponds roughly to the horizon over which Keloharju, Linnainmaa, and Nyberg (2021) document persistence in cross-sectional variation in expected stock returns. Further, a requirement to link variation in estimated factor premia to variation in observable economic variables (such as diversity in firm characteristics) that were not directly employed to identify the factors imposes a degree of discipline. Finally, the results here call for the developing of econometric methods specifically tailored to accommodate such a dynamic environment.

2. Data and Key Variables

a. Data Sources

We rely on two main data sources: monthly returns to individual common stocks and monthly returns to 205 factors derived from cross-sectional characteristics previously documented in the literature. The individual stock returns are obtained from CRSP, and include all stocks listed on the NYSE, AMEX and NASDAQ markets with a share code of 10 or 11 during the period July 1926 to December 2020. The factors are the 161 “clear predictors” and 44 “likely predictors” identified by Chen and Zimmerman (2021).⁷ We estimate factor exposures and alphas based on rolling sixty-month regressions. As a consequence, to enter our sample a stock or factor must have 60 prior months of non-missing returns, and our assessment of alpha estimates for stocks and factors begins with estimates obtained for June 1931. We obtain industry, size and book-to-market sorted portfolios along with the market excess returns from Kenneth French’s website.

b. Assessing the number of significant factors

We first seek to assess the number of economically relevant factors at each point in time. For each month t and for each factor f , we estimate 60-month rolling CAPM alphas for the factors themselves by means of the following regression over the period $t-59$ to t :

$$R_{ft} = \alpha_{ft} + \beta_{ft}R_{MKTRF,t} + \epsilon_{ft}$$

where R_{ft} is the month t return on factor f and $R_{MKTRF,t}$ is the month t value-weighted market excess return, obtained from Kenneth French’s website. A positive and significant alpha estimate indicates that the factor has explanatory power for the cross-section of stock returns beyond that which is explained by returns to the overall market. We identify a factor as significant for a given time period if the t -statistic

⁷ The authors graciously posted their data to <https://www.openassetpricing.com/>. We find qualitatively and quantitatively similar results using the factor data of Jensen, Kelly, and Pedersen (2021). We mainly focus on Chen and Zimmerman factors for two reasons. First, they more closely follow the factor construction methods employed by authors of the original papers, and second, they provide a larger set of factors to evaluate. Bessembinder, Burt, and Hrdlicka (2022) provide additional analyses of the factors constructed by these authors.

for the alpha estimate exceeds positive 3.00, the level recommended by Harvey, Liu and Zhu (2016) to allow for potential effects of multiple hypotheses testing and specification searches in the prior literature.⁸ Having done so, we display the number of factors with significant prior-sixty-month CAPM alphas as of each month. The orange solid line in Figure 1 panel A displays time series variation in the number of significant factors. For comparison, we also display with the dotted blue and dashed grey lines the numbers of factors that are significant based on alternative t-statistics hurdles of 1.96 and 4.00, respectively.⁹ While we focus on outcomes based on a t-statistic of 3.00, all three measures are highly correlated and support similar conclusions regarding the importance of allowing for time variation in the number of factors.

The literature has noted that anomalous returns to long-short portfolios are often attributable to the short leg, presumably due to higher costs of shorting shares. We display on Figure A1 in the Appendix the number of factors for which the long and short legs are individually significant. The grey dashed line indicates the number of factors that are significant based on the alpha of the factor's long-only portfolio, and the blue dotted line the number of significant factors based on the short-only portfolio. The figure shows that long-only factor portfolios have significant alphas more often than short-only portfolios. That is, the relevance of a substantial number of factors is not simply attributable to high costs of obtaining short positions.

c. Measuring stocks' unexplained mean return variation using CAPM alphas

We assess the extent to which cross-sectional variation in mean stock returns allows scope for multiple factors to be relevant. To do so, we estimate alphas from simple market-model regressions of excess firm returns on excess market returns, in each month using data drawn from the prior sixty

⁸ The factors are typically constructed by the original authors to have a positive mean return (e.g., the size factor is defined as return the return on a small firm portfolio less the return on a large firm portfolio, not vice versa). As a consequence, fewer than 4% of the alphas we estimate are negative.

⁹ In their replication studies, Hou, Xue and Zhang (2020) and Chen and Zimmerman (2021) rely on a t-statistic of 1.96. As noted, Harvey, Liu and Zhu (2016) recommend reliance of t-statistics of 3.0 or greater, while Chordia, Goyal and Saretto (2020) argue for a threshold of 3.78.

calendar months. In particular, letting R_{it} denote the month t excess return for stock i , we estimate for rolling sixty-month intervals

$$R_{it} = \alpha_{it} + \beta_{it}R_{MKTRF,t} + \epsilon_{it}.$$

Having done so, we compute the cross-sectional standard deviation of the α_i estimates for each month t .¹⁰ The standard deviation or variance of alpha estimates can be thought of as a simpler, unweighted, analog to the Gibbons, Ross, and Shanken (1989) test statistic, except for the focus on deviations from the sample mean alpha estimate rather than deviations from zero.¹¹ The key advantage to using the standard deviation measure is that the variance-covariance matrix need not be estimated and inverted, which would be impractical in light of the number of individual stocks. The degree of variation across stocks in these alpha estimates provides a measure of the extent to which mean stock returns over the sixty months diverge across stocks in a manner not explained by stocks' betas with respect to the overall market. Time series variation in the degree of cross-sectional variation in alpha estimates, in turn, give indication of the scope for the number of significant factors to vary over time.

While our main focus is on simple market-model alphas, we also consider cross-sectional and time series variation in firm-specific alphas that are estimated with respect to the well-known three- to six-factor models presented by Fama and French (1993), Fama and French (2015), Fama and French (2018), Carhart (1997), Pastor and Stambaugh (2003), Stambaugh and Yuan (2017), Barillas and Shanken (2018), and Hou, Xue, and Zhang (2015).

¹⁰ For ease of interpretability, we standardize the cross-sectional standard deviation such that it has mean zero and standard deviation equal to one across months.

¹¹ In practice, the information contained in deviations of alpha from zero as opposed to the sample mean is essentially identical. Figure IA-1 in the Internet Appendix displays uncentered and centered measures over time. The two measures are highly correlated; Pearson correlation coefficients are 0.93 and 0.95 for the equal- and value-weighted measures. We rely on the centered measure throughout the paper as it is less noisy than the uncentered measure.

3. The Evolution in the Number of Factors Over Time

a. The number of Identified Factors and Unexplained Variation in Mean Stock Returns

Figure 2 displays information regarding the scope for multiple factors to explain returns. Panel A of Figure 2 displays the number of factors amongst the 205 studied by Chen and Zimmermann (2021) that were identified in the CRSP data for each of the indicated dates. The dotted blue line displays the factor count starting from the earliest data used in the original studies, while the solid black line includes factors as of the (often earlier) date for which all data necessary to construct the factors is now available.¹² All 205 factors draw on data from 1995 or earlier, and approximately 200 of these factors draw on data from 1991 or earlier. In contrast, only about 10 factors employed data from years prior to 1961 in the studies that originally identified the factors. At present, however, sufficient data is available to implement over fifty factors in data drawn from June 1931 or later, and to implement nearly 120 factors in data drawn from 1961 or later. The key point conveyed by Panel A of Figure 2 is that the literature has identified a substantial number of factors that can be studied even in data from the earlier decades covered by the CRSP dataset.

Panel B of Figure 2 displays information regarding CAPM alpha estimates for factors as well as individual stocks. The dotted orange line displays the number of factors with statistically significant (t -statistic greater than 3.00) alpha estimates based on return data for the prior sixty months. The solid blue line displays the cross-sectional standard deviation of estimated individual stock CAPM alphas over the same periods. As noted, we view the cross-sectional variation in CAPM alphas to comprise a useful measure of the amount of variation in mean stock returns that can potentially be explained by pricing factors other than the overall market, i.e., as a measure of the scope for additional factors to be relevant.

¹² The single most common reason that we can now construct factors for time periods that were not included in the original studies is that additional accounting data has become available in the intervening years. The second most common reason is that daily data necessary to construct some measures of trading and liquidity for periods prior to 1962 were added to the CRSP data in 2006. Bryzgalova, Lerner, Lettau, and Pelger (2022) provide a comprehensive analysis of the nature and consequences of the fact that characteristic data is not available for all stock/months.

The two curves displayed on Panel B of Figure 2 diverge during the 1930s and 1940s, decades when returns were unusually volatile and for which there is even now insufficient data to implement many factor models. However, since approximately 1950 the number of statistically significant factors and the cross-sectional variability of individual firm market-model alphas appear to move reasonably closely together. We study this relation more rigorously in Section 4 below, focusing in particular on whether the comovement in the number of significant factors and cross-sectional variation in firm-level alphas represents variation in expected returns that require more factors versus the alternative that time variation in aggregate idiosyncratic volatility could explain the relation.

Figure 3 displays data informative as to the extent to which the use of prominent multi-factor models improves on the CAPM in terms of reducing the cross-sectional standard deviation of individual firm alpha estimates. Large reductions in the variability of firm alpha estimates, if observed, would be indicative that the factors employed in these workhorse models have substantive explanatory power for mean firm returns, implying limited scope for additional factors. The alternative factor models we assess include the Fama and French (1993) 3-factor model (FF3F), the Fama and French (2015) 5-factor model (FF5F), the Fama and French (2018) 6-factor model (FF6F), the FF3F model augmented with the Carhart (1997) momentum factor (FF3F+UMD), this model augmented with the Pastor and Stambaugh (2003) liquidity factor (FF3F+UMD+PSLIQ), the Stambaugh and Yuan (2017) four-factor model (M4), the Barillas and Shanken (2018) 6-factor model (BS6F), and the Hou et al. (2015) q-factor model (Q4).

The most noteworthy result that can be observed on Figure 3 is that the multifactor models do not outperform the CAPM in terms of reducing the cross-sectional variability of individual stock alpha estimates. The variability of individual stock CAPM alphas, displayed as the solid black line, has been the lowest or among the lowest as compared to the multi-factor models, particularly since about 1961. More specifically, the variability of CAPM alphas is smaller than the variability of alphas from *any* of the six other factor models considered in 65% of the individual months from 1963 to 2020.

That is, the prominent three- to six-factor models typically leave more, not less, unexplained variation in mean individual stock returns, as compared to the CAPM. This reduction in explanatory

power is all the more notable because the market return is included as a factor in the three-to-six factor models. The inclusion of factors in addition to the market must necessarily improve fit (as measured by R-squared) in the time series factor regression for each individual stock. However, the inclusion of these additional factors results in stock-specific intercept estimates that are on average further from rather than nearer to the benchmark of zero that is implied by the factor models. The implication is that the non-market factors included in the three-to-six factor models degrade the ability to explain the cross-section of average individual stock returns.¹³

The preceding result stands in contrast to numerous studies, such as Hodrick and Zhang (2001), Stambaugh and Yuan (2017), and Hou, Karolyi, and Kho (2011) that document the relative success of the three-to-six factor models. However, the literature has mainly sought to explain returns to selected portfolios rather than individual stocks. In Appendix Figure 4A we report results corresponding to those in Figure 3, except that the focus is on returns to the twenty-five size and book-to-market portfolios identified by Fama and French (1993). Consistent with the prior literature, the multi-factor models virtually always outperform the CAPM when explaining returns to these portfolios, particularly in recent decades. Specifically, the CAPM is outperformed in terms of cross-sectional standard deviation of alpha estimates by at least one of the factor models in 96.8% of all sample months and 99.6% of months since 1963.¹⁴ The fact that these models outperform the CAPM when explaining mean returns to size and book-to-market portfolios may contribute to a perception that a few factors are sufficient to explain cross-sectional variation in mean returns. In contrast, the fact that these models perform more poorly than the CAPM when explaining individual stock returns implies scope for additional factors beyond those in the widely-used three- to six-factor models.

¹³ The outcome that the inclusion of additional factors can result in alphas that deviate further from zero has appeared in the literature, but does not seem to have been emphasized. For example, results in Table I of Linnainmaa (2013) show that adding Fama and French (1993) factors increases average mutual fund alpha estimates as compared to those obtained by use of the CAPM.

¹⁴ We also assess outcomes for thirty industry portfolios identified on Kenneth French's website. The results, displayed in Appendix Figure 4A, are consistent with the results reported by Ahmed, Bu, and Tsvetanov (2019) in that the CAPM often performs better than the multi-factor models when explaining cross-sectional variation in mean industry portfolio returns as well.

b. Are some factors redundant?

The data displayed on Panel B of Figure 2 shows that as many as 95 factors have statistically significant alpha estimates at certain times during the sample period. Of course, some factors are similar to each other in their construction, and the economic information contained in outcomes on similarly constructed factors could overlap substantially. To assess the extent to which the factors studied here contain overlapping information, we rely on principal component analysis, and assess Sharpe ratios for portfolios constructed from the principal components.¹⁵ This approach is similar to that employed by Gu, Kelly, and Xiu (2020) and Kozak, Nagel, and Santosh (2018), who also rely on principal component analysis; the former focuses on principal components of individual stock returns, the latter, like us, considers principal components of factor returns. However, we emphasize time variation in factor premia, while these authors do not.

We first consider in-sample results. Figure 4 Panel A displays the number of principal components required to explain 95% of the variation across all 205 factors, as well as the number of principal components required to explain 95% of the variation in the statistically significant factors, when each is assessed on a rolling sixty-month basis. Figure 4 Panel B displays more granular information, including the number of principal components required to explain 50%, 60%, 75%, 90%, and 95% of the variation in the set of all factors.

This data reveals that, consistent with the results reported by Hou, Xue, and Zhang (2020), approximately fifty to sixty percent of the variation in the factors can be explained by a small number of principal components, ranging at various times from three to eight. However, explaining a larger portion of the variation in the factors requires many more principal components. To explain 95% of the variation

¹⁵ We impose in the estimation that the sum of the absolute values of the weights in each portfolio equals one. This constraint precludes large loadings on any individual principal component. A simple alternative approach to the Sharpe ratio approach would involve cross-sectional Fama-MacBeth regressions of firm returns on returns to all factors. However, since the number of factors studied here exceeds the number of observations in our rolling sixty-month regressions, this approach is infeasible. Lopez-Lira and Roussanov (2021) apply principal component analysis to individual stock returns and show that portfolios hedged against these components earn high returns relative to their risk. In contrast, we focus on principal components of factor returns because we are interested in the extent to which various factors are redundant.

requires between 29 and 40 principal components for every rolling sixty-month window from the late 1950s through the end of the sample period.¹⁶

Further, while the incremental explanatory power of additional factor principal components decreases by construction, this does not necessarily imply that the incremental factors are of minor economic importance. To directly assess the economic significance of allowing for an increasing and time-varying number of factors, we measure portfolio Sharpe ratios as the number of principal components is increased, first on an in-sample basis and subsequently on an out-of-sample basis. More specifically, for each month t , we construct optimized portfolios based on the first, first two, first three, etc., up to the first fifty-nine principal components. In each case, portfolio weights are chosen to optimize the portfolio Sharpe ratio. On Figure 5 we display Sharpe ratios for portfolios formed from increasing numbers of principal components, in sample. Marginal Sharpe ratios are reflected in the width of the bands displayed on the figure.

Figure 5 reveals that the Sharpe ratio continues to increase in an economically meaningful manner as the number of factor principal components increases. Even the higher-order principal components provide a non-negligible marginal contribution to a portfolio's Sharpe ratio. In particular, the first twenty principal components rarely contribute more than half of the Sharpe ratio of the portfolio constructed from all available principal components. In addition, the marginal Sharpe ratio contribution of additional principal components exhibit considerable time variation in their relative magnitudes. Overall, the evidence displayed on Figure 5 supports the reasoning that many factors contribute relevant economic information not captured by the other factors.

Table 1 provides additional data that is useful in assessing the extent to which the large number of factors considered in this study contain distinct information. Specifically, we report the results of regressions where the dependent variable during each month is the number of statistically significant

¹⁶ The number of principal components estimated from monthly data is inherently limited by the fact that only sixty data points are employed for each estimate. When we repeat this procedure using daily data, the total number of principal components is nearly equal to the number of statistically significant factors, suggesting that virtually all of the factors contain significant independent information for the cross-section of stock returns.

factors as measured over the prior sixty months, and the explanatory variables are the number of principal components required to explain 95% of the variation in all factors or 95% of the variation in the statistically significant factors. Columns (1) and (2) pertain to factors measured at the monthly horizon while, to assess robustness, columns (3) and (4) report results for factors measured at the daily horizon.

The central result observed in Table 1 is that there is a strong positive and statistically significant relation between the number of statistically significant factors and the number of principal components in the factors. This result implies that, in those months where more factors have significant CAPM alphas there is also more independent variation in the factors. This result would not be anticipated if researchers had systematically identified new factors that essentially duplicated the information contained in alternative factors. The R-squared statistics for these regressions are quite high, ranging from 0.65 (column 1) to 0.94 (column 2). We conclude that time variation in the number of statistically significant factors is not primarily attributable to increases or decreases in the number of factors that essentially replicate or are replicated by the economic information contained in other factors.

c. **Out-of-sample evidence**

The results reported in the prior section support the conclusions that the number of factors with significant explanatory power for returns is relatively large and varies over time, and that the factors are both economically important and to a substantial extent not redundant of each other. However, the results to this point are in-sample, as the significance of factors and the maximum implied Sharpe ratios are always assessed within the same sixty-month period. We next assess the extent to which the factors are or are not redundant and improve Sharpe ratios on an out-of-sample basis.

For each month, t , we consider all factors that have non-missing returns for the “in-sample” months $t - 59$ to t . From these, we compute the in-sample eigenvalues and eigenvectors of the standardized factor covariance matrix, sorting the in-sample eigenvectors by decreasing order of their corresponding eigenvalues. For the out-of-sample evaluation we focus on the 36 months from $t+1$ to

$t+36$, and multiply these out-of-sample factor returns by the in-sample eigenvectors to create out-of-sample principal components.¹⁷

We then construct portfolios comprised of increasing numbers of these out-of-sample principal components. More specifically, for each month t , we construct portfolios based on the first, first two, first three, etc., up to the first 59 out-of-sample principal components. Portfolio weights are chosen in each case to maximize the portfolio's *in-sample* Sharpe ratio. We then focus on the returns earned by these portfolios during the subsequent 36-month out-of-sample period.

Panel A of Figure 6 displays for each month the number of principal components (from 1 to 59) that results in the highest out-of-sample Sharpe ratio. While the optimal number of principal components used to form the out-of-sample portfolio can only be observed on an ex-post basis, it is informative as to the extent to which factors are redundant on an out-of-sample basis. Table 2 reports provides regression-based statistical evidence regarding the predictive ability of the number of factors that are statistically significant in-sample for both maximized out-of-sample Sharpe ratios and for the number of principal components included in the portfolios that maximize Sharpe ratios. Each estimated slope coefficient is positive and statistically significant at the 0.01 level. That is, a larger number of statistically significant factors during a given sixty-month interval reliably predicts both larger out-of-sample Sharpe ratios and that the portfolios delivering the higher out-of-sample Sharpe ratios will be formed from a larger number of principal components.

Panel B of Figure 6 displays additional information regarding out-of-sample Sharpe ratios. More specifically, Panel B displays the Sharpe ratios obtained if the portfolios are formed from the first five principal components (in line with the results of Hou, Xue, and Zhang, 2020 and Kozak, Nagel and Santosh, 2018), the average Sharpe ratio obtained from all possible numbers of principal components (from one to fifty-nine), the Sharpe ratio obtained from the maximum number of principal components,

¹⁷ While we report results for a 36-month out-of-sample window, outcomes are similar for both 12- and 60-month windows. We focus on 36 months as a balance between greater noise at short horizons and a potential loss of economic relevance at long horizons attributable to time variation in the economic importance of individual factors.

and the maximum Sharpe ratio obtained for portfolios based on any of one to 59 principal components. (Note that only the last of these relies on out-of-sample information.) Panels C and D of Figure 6 are analogous to Panels A and B, respectively, but display results that are obtained when principal components are formed from only those factors with significant in-sample ($t - 59$ to t) alphas.¹⁸ Panel A of Table 3 compiles some descriptive statistics, including means and standard deviations, of the Sharpe ratios displayed on Panels B and D of Figure 6.

The data displayed on Figure 6 and the summary statistics in Table 3 indicate that the factors that are significant in-sample have substantial out-of-sample explanatory power for portfolio returns. The Sharpe ratio for out-of-sample portfolios formed from the first five principal components are virtually always positive, average 0.38 for the full sample, and reached 1.0 during some portions of the 1980s and 1990s. The average (across various numbers of principal components) Sharpe ratios are, particularly since about 1950, larger than those based on five principal components, average 0.84 over time, and approached 1.5 during portions of the 1980s and 1990s. The maximum (across any of the numbers of principal components, corresponding to the number of principal components identified on Panel A) Sharpe ratios are always positive, consistently exceeded one from about 1974 to 2005, average 1.12 over time, and reached 2.5 at times. The maximum Sharpe ratio is always greater than the Sharpe ratio for the maximum number of principal components as (unlike a purely in-sample exercise) measured performance need not increase when the portfolio includes additional principal components. More specifically, the time series average of the Sharpe ratios obtained from the greatest number of principal components is 1.12, compared to the time series average of the maximum Sharpe ratios obtained across any number of principal components, which is 1.35. The average number of principal components for the portfolio that gives the maximum out-of-sample Sharpe ratio is forty-seven.

The data displayed on Panel C of Figure 6 shows that the number of principal components contained in the portfolios with the highest out-of-sample Sharpe ratios is most often as large or nearly as

¹⁸ Some sections of the lines in Panel D are missing due to an insufficient number of significant factors (i.e., only 1 or 0) during those time periods.

large as the number of factors that were statistically significant in-sample. During approximately the 1980 to 2002 period, the number of principal components (drawn from factors significant in sample) that maximized out-of-sample portfolio Sharpe ratios consistently exceeded twenty, and averaged twenty-three.

The data displayed in Panel D in combination with the Table 2 results showing a strong correlation between out-of-sample Sharpe ratios and the number of factors that are significant in-sample support that the number of economically significant and non-redundant factors varies over time. However, since factors estimated within sample can be used to improve Sharpe ratios out-of-sample, the results also suggest that time variation in the economically relevant factors is not so rapid as to render the factors useless after the estimation period where their significance is initially assessed.

The out-of-sample Sharpe ratios displayed on Figure 6, which are based on monthly data and have not been annualized, are economically large. However, such large Sharpe ratios are not unprecedented in the literature. For example, Kelly, Pruitt, and Su (2019) report an annualized out-of-sample Sharpe ratio of 4.05 in their study of latent factors. Further, it is not clear that these large Sharpe ratios necessarily imply unexploited arbitrage opportunities that are “too good to be true,” for two reasons.

First, investors would need to be aware and respond to the time variation in the underlying factor return premia in real time. Duffie (2010) observes that arbitrage capital moves slowly due to institutional impediments that include search costs and the elapsed time to raise capital after opportunities are identified. Martin and Nagel (2022) advance an investor learning argument that arises when investors’ priors do not correspond to underlying parameters, focusing specifically on the complexity associated with a large number of relevant characteristics. We highlight not only that numerous factors may be relevant, but also the enhanced complexity of an economy where the magnitude of factor premia can vary through time. That is, while competition among arbitragers should indeed reduce large Sharpe ratios, a dynamic economy requires not only continual investor learning, but in the presence of “moving targets.”

Second, any attempt to capture the large Sharpe ratios would involve implementation costs, which would be relatively high. Bessembinder, Burt, and Hrdlicka (2022) show that the Chen and Zimmerman (2021) factors are mainly based on equal-weighted portfolio returns, where microstructure frictions for the smaller and less-liquid stocks will comprise a barrier. Whether the results here imply profit opportunities to active traders is a question worth further assessment. However, the importance of understanding the nature of cross-sectional variation in expected returns, and time series variation therein, remains in any case. As one particular example, we are interested in knowing if illiquidity affects expected returns, even if that very illiquidity implies the absence of a profit opportunity to an active trader.

To obtain insights as to the data structure that could allow for such large out-of-sample Sharpe ratios, we conduct a series of simulations with parameters calibrated to the properties of the return data, as described in detail in Section II of the Appendix. The central insight obtained based on these simulations is that two key features of the actual data—including (i) maximized out-of-sample Sharpe ratios most often exceed one and (ii) the correlation between the number of factors that are priced in-sample and the number of out-of-sample principal components contained in the largest Sharpe ratio portfolio is high—are only obtained when the correlation in returns across various priced factors is low, more specifically below approximately 0.10. In contrast, high correlations in the returns to the priced factors do not allow for these out-of-sample outcomes to be observed. That is, the simulations support the conclusion that the out-of-sample outcomes could not be observed if the factors were substantially redundant of each other.

The results we report on Figure 6 can be contrasted to those reported by Kozak, Nagel, and Santosh (2018), who do not allow for time variation in factor premia. They report that while factors beyond the first few principal components contribute substantially to in-sample Sharpe ratios, principal components beyond the first few do not substantially enhance the Sharpe ratio out of sample. They argue that this result is to be expected since even a relatively small number of arbitragers should “be sufficient to ensure that near-arbitrage opportunities—that is, trading strategies that earn extremely high Sharpe

ratios do not exist.” Yet, we document high out-of-sample Sharpe ratios, and that principal components beyond the first few contribute substantively.

We assess whether time variation in factor premia can reconcile the differences in our findings as compared to those of Kozak, Nagel and Santosh (2018). First, we replicate their results. In particular, we compute in- and out-of-sample Sharpe ratios for portfolios formed from the principal components of the same thirty long and short factors that they consider, when the in- and out-of-sample periods are defined based on the first and second half (25 years each) of their sample period.¹⁹ We then assess the effect of instead focusing on shorter subsamples, ranging from 36 to 120 months, to define the in- and out-of-sample periods.

Figure 7 displays the findings. Panel A shows the in-sample Sharpe ratios as a function of number of principal components in the portfolio, while Panel B shows the corresponding out-of-sample Sharpe ratios. The bright red lines in both panels confirm the findings of Kozak, Nagel and Santosh (2018). In particular, in-sample Sharpe ratios increase as more principal components are used to form portfolios, while out-of-sample Sharpe ratios increase modestly beyond the 5th principal component, when the in- and out-of-sample periods are each based on half of the full sample.

The additional six lines on Figure 7 display average (across time) of portfolio Sharpe ratios for shorter estimation windows. These indicate that average out-of-sample Sharpe ratios increase to a maximum of nearly 1.6 as the estimation windows are decreased from half the sample, as in Kozak, Nagel and Santosh (2018), to thirty-six months. Further, out-of-sample Sharpe ratios continue to increase beyond the first five principal components until the portfolios are constructed from approximately twenty-five principal components when shorter estimation windows are employed.

The data reported in Panel B of Table 3 quantifies these differences more specifically. In particular, we report there the difference in out-of-sample Sharpe ratios for portfolios constructed from

¹⁹ We thank Kozak, Nagel and Santosh (2018) for kindly sharing their data. For robustness, we also replicated their findings using the factor data from Chen and Zimmerman (2021) for the same anomalies. Figure A2 in the Appendix shows similar results.

thirty principal components as compared to portfolios constructed from five principal components, when the underlying assets are either the long and short legs of fifteen anomaly portfolios or twenty-five book-to-market portfolios. The first row contains results that correspond to those reported by Kozak, Nagel, and Santosh (2018), who split their sample into equal halves that comprise in- and out-of-sample subsets. The other rows report corresponding results that we obtain by rolling estimation over shorter windows. Focusing on the anomaly portfolios, the use of thirty rather than five principal components with the equal split of the sample increases the Sharpe ratio by 0.06. In contrast, our shorter-window estimation increases out-of-sample Sharpe ratios for thirty principal components as compared to five by between 0.45 and 0.62, depending on the precise lengths of the estimation windows. Outcomes for the size and book-to-market portfolios are broadly similar. That is, additional principal components are relevant out-of-sample, as Sharpe ratios for portfolios constructed from them increase more, as estimation windows become shorter to accommodate time variation in factor premia.

d. The Number of Relevant Factors and the Unconditional SDF

Kozak, Nagel and Santosh (“KNS”, 2020) report that, while a large number of factors are required to explain the cross section of returns to the fifty-anomaly based factors in their sample, a relatively sparse stochastic discount factor formed from only four principal components performs quite well, as judged by their model’s out-of-sample R^2 statistic. This finding appears to contrast with our own, though the difference is likely attributable, at least in part, to the fact that we study a larger set of factors and that we allow for time variation in factor risk premia.

We investigate further. The computer code employed by KNS computes not only the out-of-sample R^2 statistic, but also out-of-sample Sharpe ratios.²⁰ Despite the facts that the estimation procedure they employ penalizes deviations of Sharpe ratio estimates from zero and that their method does not accommodate time variation in parameters, the out-of-sample Sharpe ratios that are estimated in

²⁰ More specifically, their program computes the square root of the expected squared Sharpe ratio.

their sample are 0.75 with 4 factor principal components, 0.90 with ten factor principal components, and 1.11 with 48 factor principal components.

However, these Sharpe ratio estimates may be biased due to the fact that the KNS program computes principal component eigenvectors over the full sample period.²¹ We therefore modify their program to construct factor principal components separately during the “training folds” (in-sample subperiods) and apply the resulting eigenvectors to returns in the “evaluation folds” (the out-of-sample subperiods). Figure 8 displays the outcomes. Panel A displays the out-of-sample R2 statistic, and corresponds to Figure 3A in KNS. Panel B displays the corresponding out-of-sample Sharpe ratios. The specific out-of-sample Sharpe ratios estimated in their sample are 0.80 with four factor principal components, 0.94 with ten factor principal components, and 1.09 with 48 factor principal components. We conclude that the data and programs employed by KNS also support that employing factor principal components beyond the first few leads to greater explanatory power for the cross-section of out-of-sample stock returns.

Our results also help to understand why KNS report that many factors (as opposed to factor principal components), some with small SDF weightings, are necessary for good out-of-sample performance. In particular, we posit that time variation in factor premia contribute indirectly to their findings. Even if a particular factor is not significant during the in-sample period, keeping small non-zero weights on many factors implies that those that factors that become economically important out-of-sample contribute to portfolio performance.

We also obtain from Serhiy Kozak’s website the optimal SDF coefficients for individual factors as estimated by Kozak, Nagel and Santosh (2020). Analogous to the research approach adopted here, we then assess, for each of the fifty anomaly portfolios they study, the percentage of sample months where the portfolio has a significant (t-statistic > 3.0) alpha in rolling sixty-month regressions of portfolio returns on market returns. Finally, we study relations between the absolute value of the coefficients in

²¹ We are grateful to Stefan Nagel for identifying this bias, and suggesting the solution to eliminate it.

the SDF as reported by Kozak, Nagel and Santosh (2020), and the percentage of months where the portfolio has a significant alpha.²²

The results, displayed on Figure 9, demonstrate a strong positive relation between SDF coefficients and the frequency of significance. The figure also reports the outcome of an OLS regression of absolute SDF coefficients compiled by Kozak, Nagel and Santosh (2020) on the percentage of months with significant alphas; the slope coefficient is 0.80 with a t-statistic equal to 9.85, and the regression R-squared statistic is 0.67. These results imply their shrinkage technique produces weights related to the fraction of time a factor is significant and also places small rather than zero weights on factors that have only intermittent significance.

4. Time series variation and explanatory power outside original sample periods

It has been suggested that most empirical findings related to factors are attributable to specification searches (also referred to as “data snooping” or “p-hacking”) and a failure to incorporate appropriate multiple testing procedures. However, it has also been argued that the large majority of the factor-related findings can indeed be replicated, do not arise from specification searches, and survive adjustment for multiple testing.²³

While we do not resolve this debate, we contribute to it by providing the out-of-sample evidence described in the prior section, and in three additional ways. First, we assess the extent to which factors have statistically significant explanatory power in subperiods before and after those examined in the studies that originally identified the factors. In doing so, we extend the related results reported by McLean and Pontiff (2016), Linnainmaa and Roberts (2018) and Ilmanen, Israel, Moskowitz, Thapar, and

²² We rely on absolute t-statistics and coefficients since Kozak, Nagel and Santosh (2020), unlike some authors, do not normalize factor returns such that they have positive mean returns.

²³ Studies that conclude that factor-based evidence is largely unreliable include Harvey, Liu, and Zhu (2015), Linnainmaa and Roberts (2018), Chordia, Goyal, and Saretto (2020), and Hou, Xue, and Zhang (2020), while the studies arguing that identified factors do reliably explain returns include Chen (2021), Chen and Zimmerman (2021), and Jensen, Kelly, and Pedersen (2021).

Lee (2021) in that we study a substantially larger set of factors.²⁴ Second, while the related studies have mainly assessed whether factors do or do not have explanatory power for the pre- and post-sample periods as a whole, we assess the extent to which factors' explanatory power changes over time, both within and outside of the authors' original samples. Third, we go on to identify the economic determinants of such time series variation.

Figure 10 displays information regarding the significance of each factor over time. The figure includes one row for each factor, and a column for each sample month.²⁵ A given row and column contains a dot (either blue or grey) if the t -statistic on the alpha estimated in a market-model regression of the factor return on the overall market return over the prior sixty months is greater than 3.00. In addition, each row contains a green dot that denotes the earliest data used in the original study that identified the factor, a red dot that denotes the latest data used in the original study, and a magenta dot that indicates the earliest date for which we are able to estimate the factor's alpha based on data now available. It is, of course, not possible even now to ascertain if the factor had significant explanatory power for returns for those dates that are earlier than the magenta dots.

Two points can be observed visually on Figure 10. First, factors often display statistically significant explanatory power in data drawn from months both before and after the data used in the original study that identified the factor. Panel A of Table 4 reports on the extent, indicating that over three quarters (77%) of the factors have significant explanatory power during at least one sixty-month

²⁴ However, replication rates are not directly comparable across our study and theirs, as we focus on a set of 205 factors that were previously verified by Chen and Zimmerman to have significant explanatory power within the authors' original sample periods. In contrast, only 85 of the 97 factors studied by McLean and Pontiff (2016) have an in-sample t -statistic greater than 1.50, and only 32 of the 36 factors studied by Linnainmaa and Roberts (2018) have an in-sample t -statistic greater than 1.96. Ilmanen, Israel, Moskowitz, Thapar, and Lee (2021) study just four factors, but over a 100-year sample period, and in several distinct asset classes. They report little evidence that arbitrage reduces factor premia over time.

²⁵ For purposes of Figure 9, we follow Chen and Zimmerman (2021) in assigning factors to categories, including (1) "Price", which includes factors mainly constructed from return data, (2) Accounting, which includes factors that rely on financial statement data, (3) Analyst, which rely on analyst estimates, (4) Trading, which use volume and transactional data, (5) 13F, which use institutional holdings data, (6) Options, which use options-related data, and (7) Other, which include hand-collected or other non-standard data. Due to the small numbers of factors, we combine the last two categories as "Other".

interval prior to the range of dates used in the original studies, and a remarkable 93% have significant explanatory power during at least one sixty-month interval after the range of dates used in the original studies, when significance is assessed by a t-statistic on the alpha estimate of 1.96 or greater. If statistical significance is defined based on a larger t-statistic the proportion of factors that are significant outside of the original sample period declines, but remains large. For example, applying the t-statistic of 3.00 (p-value is .003 or less) used elsewhere in this paper, 54% of factors are significant during at least one earlier sixty-month interval and 69% are significant during at least one subsequent sixty-month interval (as compared to the original study sample period). This out-of-sample evidence supports the reasoning that the factors' success in the original studies cannot be fully attributed to data mining or specification searches.

The second observation that can be gleaned from Figure 10 is that the statistical significance of individual factor alphas varies over time; in many cases a given factor is significant for periods spanning multiple years, loses significance for a time, and then regains significance. Panel B of Table 4 reports on the distribution of the number of non-overlapping periods, or “spells” of significance, for various t-statistic cutoffs. For example, relying on a t-statistic cutoff of 3.00, the cross-factor median number of significance spells is 6.0 per factor, while the cross-factor mean is 7.7 spells per factor. Panel C of Table 4 reports on the distribution of the duration of such significance spells. Once again based on a t-statistic cutoff of 3.00, the cross-factor median length of a significance spell is 13 months, while the cross-factor mean length is 22 months.

Of course, a pattern whereby statistical significance for individual factors ebbs and flows over time could simply reflect random noise in a stable economic environment. That is, a factor with a constant premium equal to zero or an economically modest level could be associated with significant estimates during some intervals and insignificant estimates during other intervals. Alternatively, the pattern could reflect that the number of factors that earn a return premium, or the magnitude of such return premia, vary over time. We distinguish between these explanations in two ways. First, we use simulation methods to assess the distribution of the statistics reported on Panels B and C of Table 4 under

the null hypothesis that factor premia are constant over time. Second, we present evidence in Section 6 that assesses the extent to which variation in the number of significant factors is or is not related to measures of changes in the economic environment.

To assess the distribution of the statistics reported on Table 4 under the null hypothesis that factor premia are constant over time, we proceed as follows. First, we estimate each factor's constant alpha, beta and residual volatility from a regression of its returns on the market excess returns. We then create a simulated time series of market returns calibrated to the sample mean and standard deviation of the market over our sample period, and generate a simulated time series of returns for each of the 205 factors using a factor model that relies on the simulated market returns in combination with the estimated alpha, beta, and residual volatility for each factor. The length of each factor's simulated time series is matched to the number of sample observations for the factor return. We then estimate rolling 60-month regressions of simulated factor returns on simulated market returns, and obtain both the count and average length of significance spells for each simulated factor, when significance is assessed based on t-statistics ranging from 1.96 to 4.00. Having done so, we compute the cross-factor average of the spell counts and spell lengths, corresponding to the sample data reported in Table 4. We repeat the simulation 2,000 times to obtain a distribution of the cross-factor average factor spell lengths and counts.

Panel A of Figure 11 displays the simulated distributions for the cross-factor average of the average spell lengths, while Panel B corresponding cross-factor average spell counts. The red dashed lines display corresponding sample outcomes. The information displayed on Panel A of Figure 11 shows that the statistics reported on Table 4 based on the actual sample are unlikely to be observed under the null hypothesis of constant factor premia. For each of the t-statistic cutoffs (used to define significance) considered, the actual average spell length lies far in the right tail of, or entirely outside, the simulated distribution of spell lengths. That is, actual spell lengths are longer than would be observed under the null hypothesis, as would be anticipated if premia were economically large during some periods.

The information displayed on Panel B of Figure 11 shows that the cross-factor average number significant spells also (with the exception of the results obtained based on a t-statistic cutoff of 2.5)

diverges from the distribution obtained under the null of constant premia. The use of a higher t-statistic cutoff naturally leads to fewer periods of significance, both in the sample data and in the simulated distribution obtained under the null. Note, though, that with high t-statistic cutoffs of 3.5 or 4.0 the actual average count of significance spells lies far in the right tail of the simulated distribution, while with low t-statistic cutoffs of 1.96 or 2.00 the actual average count of significance spells lies to the left of the simulated distribution. That is, simulated outcomes under the null hypothesis are considerably more sensitive to the t-statistic cutoff employed as compared to sample outcomes. This result reflects that time-invariant premium estimates of modest economic magnitude are more likely to be recategorized as insignificant rather than significant as the t-statistic hurdle increases, while factor premia that are economically substantive at some times and close to zero at other times are less sensitive to the t-statistic employed. On balance, the simulation outcomes displayed on Figure 11 imply that it is exceptionally unlikely that the sample data reported on Table 4 would be observed under the null hypothesis of constant factor premia.

5. The role of Idiosyncratic Volatility

The empirical results reported in the prior sections demonstrate that (i) a substantive but time-varying number of factors have explanatory power that is both statistically and economically significant for the cross-section of stock returns during certain time periods, (ii) most of the factors are significant in periods before and after the time intervals studied by the authors who originally identified them and are useful in out-of-sample portfolio selection, (iii) the factors are generally not redundant of each other, in that the number of principal components required to explain their variation is substantial, both within and out-of-sample, (iv) the extent to which the simple CAPM explains the cross-section of mean returns to individual stocks varies substantially over time, and (v) popular three- to six-factor models most often underperform the CAPM in explaining the cross-section of mean returns to individual stocks, and, therefore, do not substantively diminish the potential role of additional factors in explaining returns.

As noted, we view cross-sectional variability in CAPM alpha estimates as being informative as to the scope for multiple factors to explain average returns. However, it has been documented, e.g., by Goyal and Santa Clara (2003), that aggregate idiosyncratic volatility, i.e., the cross-sectional average standard deviation of residuals from market-model regressions, varies over time. An alternative hypothesis is that the standard deviation of *ex post* alphas varies over time because of variation in aggregate idiosyncratic risk, not because of variation across stocks in *ex ante* expected returns. To guide our analysis, we provide in section I of the Appendix an assessment of the statistical determinants of such variation. We show that idiosyncratic return volatility is relatively more important in explaining the average standard error of the alpha estimates, while return premia associated with non-market factors are relatively more important in explaining cross-sectional variation in market-model alpha estimates. Thus, when both the volatility and the average standard error of alpha estimates are included as explanatory variables in the same regression, the coefficient on the former primarily reflects the effect of non-market return factor premia, while the coefficient on the latter primarily reflects average idiosyncratic risk.

Table 5 reports the results of time series regressions where the dependent variable in each month is the number of statistically significant factors, and the explanatory variables include the cross-sectional standard deviation of firm-specific CAPM alpha estimates. If, as we hypothesize, greater cross-sectional variation in CAPM alphas indicates greater variation in *expected* returns across stocks, and that variation is attributable to return premiums associated with factors other than the market, then we should observe a positive coefficient on this variable, with or without inclusion of the residual volatility variables. We include as control variables either the cross-sectional mean standard error of the alpha estimates, or to assess robustness, the cross-sectional average market-model idiosyncratic volatility, σ_{ϵ}^2 , itself. Columns (1) to (3) are based on estimation with equal weighting of each observation, while to assess robustness we report in Columns (4) to (6) corresponding results when each observation is weighted by the firm's market capitalization at the beginning of the sixty-month estimation period.

The results reported in Table 5 indicate positive and significant coefficient estimates on the standard deviation of alpha estimates, with or without inclusion of the control variables, and by either weighting method. While the coefficient estimates on the control variables are negative, suggesting that higher idiosyncratic volatility tends to reduce the number of statistically significant factors due to reduced statistical power, these estimates are not themselves significant. The inclusion of the control variables substantially increases the magnitude of the coefficient estimates on the cross-sectional standard deviation of the alpha estimates. That is, these results show that the number of significant factors is significantly explained by the cross-sectional standard deviation of firm-specific alphas, but not by aggregate idiosyncratic volatility. These results support the reasoning that greater cross-sectional variation in firm-specific alpha estimates results from greater dispersion in expected returns attributable to return premia associated with non-market factors which in turn allows for the empirical estimation of a larger number of such factors.

6. Variation in Significant Factors and Economic Complexity

The relevance of a given factor in terms of explaining the cross-section of stock returns can depend on the volatility of the factors, the variation across stocks in sensitivities of firm returns to factor outcomes, as well as the magnitude of the return premia per unit of risk associated with the factor. We report in this section on relations between the number of factors that are statistically significant during rolling sixty-month periods, and a number of measures related to the state of the economy and the complexity of the economic environment. We focus mainly on results for the 1968 to 2020 period, during which we can construct a larger set of such measures. However, we report corresponding results for the full 1931 to 2020 sample in the Appendix.

a. The role of the number of listed firms.

We first focus on relations between the number of factors with statistically significant CAPM alphas and the number of firms traded in the U.S. markets. We reason that large increases or decreases in the number of publicly traded firms are likely to be accompanied by shifts in the types of firms available

for public investment. Indeed, Fama and French (2004) show that the characteristics of firms newly listed on major U.S. stock markets varies substantially over time. Multiple and varied risk factors may be necessary to explain patterns in the returns of varying firm types.

Column (1) of Table 6 reports outcomes obtained from a regression of the number of statistically significant factors during months $t-59$ to t on the number of firms listed in month t , and indicate a positive and statistically significant relation.²⁶ This finding supports the reasoning that a larger number of factors are required to explain cross-sectional variation in mean returns when more firms are listed. This result need not arise mechanically. As a simple example, suppose the CAPM determined expected returns for all stocks. The addition of new stocks with unique characteristics would only require estimation of their potentially distinct market betas. The empirical fact that more factors have significant explanatory power at times when more firms are listed is consistent with the reasoning that the firms that enter and depart the CRSP database differ from other firms in that they are exposed to differing sources of priced risk, rather than simply having differential exposures to a fixed set of priced systematic risks.

A simple alternative explanation for the observed positive relation between the number of statistically significant factors and the number of listed firms is that a larger cross-sectional sample size improves statistical power, such that estimated return premia of given economic magnitudes are more likely to become statistically significant. We demonstrate in section I.C of the appendix that this possibility arises, in particular, when the non-market factors are not directly observable, and consistent with actual practice, the empirical analyses are implemented based on factors created from returns to portfolios sorted based on observable firm characteristics.

To distinguish between these possibilities, we conduct cross-sectional regressions of the number of statistically significant factors on (i) the cross-factor average standard error of the alpha estimates, (ii) the cross-factor average absolute alpha estimate, and (iii) the number of firms. We report outcomes in

²⁶ In Appendix Table A2 we report results obtained when the number of firms is assessed as of month $t-60$ and as the number firms continuously listed from time $t-60$ to t (so that alpha can be estimated). Outcomes are similar for each measure.

Appendix Table A3. As would be anticipated, the result reported in columns (1) and (5) of Table A3 indicate that fewer factors are statistically significant in periods where alpha standard errors are larger. As we show in section I.C of the appendix, one determinant of these standard errors is the number of firms in the sample. In columns (2) and (6) of Table A3 we report results that are obtained when the number of firms is included in the regression along with the average standard error. We continue to estimate negative coefficients on the mean standard error, even while we estimate positive coefficients on the number of firms. These results indicate that the positive relation between the number of statistically significant factors and the number of firms is not solely attributable to the effect of the number of firms on the standard errors.

Of course, the number of significant factors depends on the magnitude of the factor alpha estimates as well. In columns (3) and (7) of Table A3 we report results obtained when the explanatory variables include the mean absolute alpha as well as the mean standard error of the alpha estimates, while in columns (4) and (8) we report results when the number of sample firms is included as the third explanatory variable. These results confirm that, while a larger mean alpha is, as expected, associated with more statistically significant factors, the number of firms continues to have a significantly positive effect as well. We conclude that the number of listed firms has explanatory power for the number of significant factors that is distinct from the improvement in statistical power associated with a larger sample size, and that the number of firms contributes explanatory beyond any direct effect on the mean alpha estimate.

b. The State of the Economy, Economic Complexity and Diversity in Firm Characteristics

We next assess the extent to which the number of significant factors is related to aspects of economic complexity, and to the diversity of observable firm characteristics. To facilitate interpretation, we standardize each of the following variables relative to its own time series. Thus, regression coefficients are interpreted as a response to a one-standard deviation change in that variable.

We conjecture that the business cycle will be relevant, both because of potential variation in the magnitude of return premia and due to changes in firm types, with economic expansions characterized by high rates of firm entry and recessions more likely to involve net exit by firms. To capture these effects, we rely on an indicator variable equal to 1 for recession months, as defined by the National Bureau of Economic Research, and the unemployment rate reported by the US Bureau of Labor Statistics. We also consider two interest rate series, the Fed Funds rate (which begins in 1954) and the 10-year treasury note yield (which begins in 1964). Interest rates potentially capture the effects of monetary policy and funding conditions. The unemployment rate, fed funds rate and treasury yields are all obtained from the Federal Reserve Economic Data (FRED) website.

Fama and French (2001) suggest that the disappearance of dividend-paying firms reflects the changing characteristics of publicly traded firms. Thus, to further measure variation in firm types, we compute the proportion of dividend-paying common stocks as the number of firms paying at least one cash dividend in the previous 12-months relative to the total number of common stocks. Variation in firm characteristics such as the propensity to pay dividends could arise as firms respond to demand from different investor types. Further, the preferences of the marginal investor who effectively sets prices for specific stocks can depend on whether the investor is an individual or an institution.²⁷ To potentially capture the impact of changes in the composition of the investor base we measure the proportion of each firm's shares outstanding held by 13-F institutions in the Thomson-Reuters database.

We also consider the possibility that the number of significant factors may be related to market liquidity and to general economic complexity. To the extent that factor premia arise because investors are unable to profitably trade to eliminate mispricing, we should observe that more factor premia are significant when markets are less liquid. To assess this possibility, we compute on a monthly basis the average across stocks of the Amihud (2002) illiquidity measure. As a proxy for general economic

²⁷ Lewis and Santosh (2021), for example, show that an asset pricing model where betas are defined relative to the portfolios held by active institutional investors performs better than the standard CAPM where betas are defined relative to aggregate market holdings.

complexity, we use the Economic Complexity Index constructed by Simoes and Hidalgo (2011), which is a measure of “the relative knowledge intensity of an economy.”

To measure diversity in firm characteristics, we first compute the cross-sectional standard deviation for each of the 205 characteristics within each month. We then rescale each of these measures such that the time series mean is zero and the time series standard deviation equals one. Finally, we compute the sum of these standardized volatility measures across characteristics within each month.²⁸ The result can be interpreted as a measure of the cross-sectional dispersion in those characteristics observable for the available sample of firms in each month. Note that no relation necessarily exists between the number of firms and cross-sectional dispersion in characteristics; if newly listed firms were predominantly similar to the typical existing firm in terms of observable characteristics the diversity measure would decline rather than increase as more firms listed. An increase in the cumulative dispersion in characteristics across firms indicates, in contrast, that the underlying firms themselves are becoming increasingly differentiated.²⁹

Figure A5 displays the average number of characteristics that can be computed, delineated by the number of months since the firm initially appears in the database. In the first few months, less than twenty characteristics can be computed, on average. Thirty-six months after listing, approximately one hundred characteristics can be computed. This rapid growth reflects that many characteristics require

²⁸ Note that since the measures are rescaled to a zero mean there is not a mechanical relation between the sum and the number of characteristics available in a month. The correlation between the sum and the mean across characteristics is 0.96, and use of the latter for the results in Table 5 leads to results that are virtually identical, but with moderately higher standard errors on the coefficient estimates.

²⁹ In Appendix Figure A3 we present evidence regarding time series variation in the number of characteristics (among the 205 considered) for which characteristic data is available as a function of calendar time. Panel A reveals that the numbers of firms for which large numbers of factors are available grew rapidly between the initial years of the sample until the late 1990s, attributable both to increases in the number of listed firms and in the number of characteristics for which requisite data is available. To distinguish the separable effects of changes in the number of firms and changes in the set of characteristics for which data is available, Panel B of Figure A3 displays for each month the percentage of firms for which the indicated numbers of characteristics are available, and demonstrate that the percentage of firms with a large number of available characteristics has steadily increased over time. It can also be observed that the percent of firms with large numbers of characteristics decreases during the final few years of the sample, which can be attributed to the recent growth in the number of IPOs in combination with the fact that prior accounting and return data is necessary to compute some characteristics.

prior accounting statement data (which is often sparsely collected at the beginning of a firm's public life), as well as prior return history. However, the fact that a given characteristic cannot yet be computed by an econometrician need not imply that market participants are unaware of the characteristic. To accommodate the "burn in" period between the addition of a firm and the time when characteristics become observable, we focus on cross-sectional variation in characteristics in month $t+36$ to measure firm characteristic diversity as of month t .

Table 6 reports the results of regressions of the number of statistically significant factors on these measures of economic complexity. Columns (1) to (10) report results of univariate regressions for each variable in turn, while columns (11) and (12) report multivariate outcomes. We omit mean institutional ownership the final multivariate specification, because data is available only from 1980 onward.

The univariate results reported in Table 6 show that the number of statistically significant factors is related to macroeconomic conditions, decreasing during recessions and increasing during periods of higher interest rates, based both on the Federal Funds rate and the Treasury-bond rate (though the former is not statistically significant during the more recent subsample). The unemployment rate, in contrast, does not have significant explanatory power. It is, however, noteworthy that the macroeconomic variables have much less explanatory power for the number of significant factors as compared to the number of listed firms. The R-squared statistics for the statistically significant macroeconomic variables vary from 0.03 for the recession indicator to 0.17 for the Treasury bond rate, as compared to 0.50 for the number of listed firms.

The coefficient estimates reported in column (6) of Table 6 indicate that the number of statistically significant factors is negatively related to the percentage of firms that pay dividends, with a full-sample r-squared of 0.11. This result is consistent with the reasoning that the listing of non-dividend paying firms, which tend also to be younger and less familiar to investors, is associated with an increase in the number of significant factors, and more broadly with the notion that more factors are required to explain returns when listed firms are more diverse. The coefficient estimates reported in column (7) of

Table 6 indicate that the number of statistically significant factors is also strongly negatively related to mean institutional ownership, with an R-squared statistic equal to 0.46. If institutions invest with a differing objective function as compared to individuals (due, for example, to agency issues or heterogeneity across individual investors) then changes in institutional ownership can effectively alter the identity and objective of the marginal stock market investor. The negative coefficient estimates reported on Table 6 imply that increased institutional ownership reduces the number of significant factors, potentially because it effectively reduces variation in the identity of the marginal investor. The coefficient estimate on the economic complexity index in column (8) is positive and is statistically significant. The coefficient estimate for the average Amihud illiquidity measure (column 9) is not statistically significant. The coefficient estimate for the diversity of firm characteristics (column 10) is positive and statistically significant, with a relatively large r-squared statistic of 0.38. This result implies that more factors are significant during those periods when there is greater cross-sectional variability in the firm characteristics that are observable to econometricians.

Columns 11 and 12 present results for multivariate specifications. The unemployment rate remains significant in all specifications, but the recession indicator and the 10-year Treasury bond yield lose significance in the shorter sample employed for Column 12. The coefficient on cross-sectional mean Amihud illiquidity measure is positive and significant in Columns 11; that is, the multivariate outcomes support that greater illiquidity is associated with more significant factors, potentially due to reduced arbitrage activity. The proportion of firms paying dividends, and the ECI become insignificant in the multivariate setting.

Notably, the diversity of firm characteristics remains significant. That is, even after allowing for the explanatory power of macroeconomic variables such as the unemployment rate and interest rates, changes in institutional ownership, cross-sectional variation in firm characteristics has explanatory power for the number of significant factors. The number of publicly listed firms is no longer significant in the multivariate setting, which is consistent with the reasoning that the univariate significance of the number of firms is linked to the greater diversity of characteristics when the number of firms is large. The R-

squared statistics for each of the multivariate regressions exceed 0.70. In combination, the results here provide strong support for the notion that time variation in the number of significant factors is not random, but rather is linked to variation in macroeconomic conditions and observable diversity in firms' characteristics.

7. Conclusions

The reasoning that only a few factors should be necessary to explain the cross section of mean returns is attractive because parsimony is desirable. So, should the fact that the literature shows that many empirically observable factors have explanatory power for the cross-section of stock returns be viewed as a collective failure? We think not, if the reason is that financial markets and the broader economy are complex and dynamic. The characteristics of the firms that are available for investment can change through time as existing firms evolve and new firms are listed or delisted. Investors are diverse in terms of their investment horizon and objectives. Some investors trade on their own account, while others rely on professional managers whose strategies can be affected by agency issues related to their compensation. The identity of the marginal investor can differ across stocks, and in any given stock can vary through time. Return premia have been shown to depend on intermediaries' funding liquidity, leverage, and balance sheets, as well as on the state of the economy. In short, it is unclear that return premiums in actual capital markets are necessarily governed by only a small and time-invariant set of factors.

More broadly, Cochrane (2011) observes that most variation in price-to-dividend ratios is attributable to changes in discount rates, i.e., expected returns. If factor models determine expected returns, it follows that variation in discount rates is attributable to time variation in interest rates and factor return premia. Prices are, in turn, determined in the course of market trading, based on the interaction between buy and sell orders. Cochrane (2022, page 31) observes that "the standard models do not produce a hundredth of the observed trading volume." It follows, in our view, that the determinants of expected returns are not necessarily confined to those predicted by these standard models, and can vary

as market conditions and the economic environment change. The need to be mindful of the possibility of collective data mining and joint hypothesis testing notwithstanding, these considerations support allowing the data to speak on the issues.

We present a number of empirical findings relevant to these issues, showing that a substantial number of factors have significant explanatory power, and that the number of significant factors varies substantially over time. Further, the number of principal components required to explain variation in the significant factors is also large and is positively correlated with the number of significant factors, both in sample and out-of-sample, implying that the results do not simply arise because various researchers identify factors that are redundant of each other. Out-of-sample Sharpe ratios for portfolios formed from the principal components of factors that are significant in-sample are economically large and comparable to those obtained from recent machine learning applications. Further, the results of existing studies that assess whether factors identified in-sample are useful for portfolio construction out-of-sample are altered when the only substantive change in research design is to allow for time variation in factor premia.

We assess the extent to which widely-used three- to six-factor models do a better job of explaining the cross-section of returns as compared to the CAPM. While these models outperform the CAPM in terms of explaining returns to characteristic-sorted (size and market-to-book) portfolios, they do not reliably outperform for industry portfolios, and most often perform worse than the CAPM for the cross-section of stock returns. The last result is noteworthy in part because the three- to six-factor models all include the market factor, implying that the non-market factors degrade the ability of the models to explain mean returns to individual stocks. To the extent that the perception that only a few factors should matter for stocks in general is based on the performance of three- to six-factor models in explaining returns to characteristic-sorted portfolios, the perception is misplaced.

We also provide evidence that the number of significant factors varies through time. We use simulation methods to show that neither the average number of periods where a factor is significant or the average period of time with significance is consistent with the null hypothesis that factor premia are constant over time. We further show that the number of significant factors varies with measures of

economic complexity and firm diversity. In particular, the number of significant factors is related to a recession indicator variable, interest rates, the percentage of firms that pay dividends, mean institutional ownership rates, and an economic complexity index, and is particularly strongly related to the number of firms that are publicly listed, cross-sectional variation in observable firm characteristics. The finding with respect to the number of firms supports the reasoning that newly listed firms systematically differ from existing firms in terms of systematic risks relevant to investors. Finally, the finding with respect to diversity of firm characteristics suggests that more factors are relevant when firms themselves are more distinct.

On balance, our findings suggest that multiple and time-varying factors may be required to price the cross-section of returns as the economy continues to evolve dynamically and new firms are listed. Further, in a dynamic economy a factor can be significant in explaining returns during some periods but not others. This suggests the desirability of a degree of caution in interpreting the results of existing out-of-sample tests, as insignificant out-of-sample outcomes need not imply that the factor was unpriced in the original sample period, and the need for the development of econometric methods for out-of-sample tests suitable to the dynamic environment.

REFERENCES

- Adrian, Tobias, Erkki Etula, and Tyler Muir, 2014, Financial Intermediaries and the Cross-Section of Asset Returns, *Journal of Finance*, 69, 2557-2596.
- Ahmed, Shamim, Ziwen Bu, and Daniel Tsvetanov, 2019, Best of the best: a comparison of factor models, *Journal of Financial and Quantitative Analysis* 54, 1713-1758.
- Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets*, 5, 31-56.
- Bali, Turan, Stephen Brown, Scott Murray, and Yi Tang, 2017, A Lottery-Demand-Based Explanation of the Beta Anomaly, *Journal of Financial and Quantitative Analysis*, 52, 2369-2397.
- Barber, Brad M. and Huang, Xing and Odean, Terrance and Schwarz, Christopher, 2021, Attention Induced Trading and Returns: Evidence from Robinhood Users, *Journal of Finance*, Forthcoming.
- Barillas, Francisco, and Jay Shanken, 2018, Comparing asset pricing models, *Journal of Finance* 73, 715-754.
- Bessembinder, Hendrik, Aaron Burt and Christopher Hrdlicka, 2022, Factor Returns and Out-of-Sample Alphas: Factor Construction Matters, working paper, downloadable at [ssrn.com/abstract=4281769](https://papers.ssrn.com/abstract=4281769).
- Betermier, Sebastien, Laurent Calvet, Samuli Knüpfer, and Jens Kvaerner, 2021, What do the portfolios of individual investors reveal about the cross-section of equity returns?, working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3795690.
- Bryzgalova, Svetlana, Sven Lerner, Martin Lettau, and Markus Pelger, 2022, Missing Financial Data (May 11, 2022). available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4106794.
- Brennan, Michael, 1998. The role of learning in dynamic portfolio decisions, *European Finance Review*, 1, 295–306.
- Bustamante, M Cecilia, and Andres Donangelo, 2017, Product market competition and industry returns, *Review of Financial Studies* 30, 4216-4266.
- Carhart, Mark M, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.
- Chen, Andrew Y, 2021, The limits of p-hacking: Some thought experiments, *Journal of Finance* 76, 2447-2480.
- Chen, Andrew Y., and Tom Zimmermann, Open source cross sectional asset pricing, *Critical Finance Review*, Forthcoming.
- Chen, Yao, Alok Kumar, and Chendi Zhang, 2021, Searching for Gambles: Investor Attention, Gambling Sentiment, and Stock Market Outcomes, *Journal of Financial and Quantitative Analysis*, 56, 2010-2038.
- Chinco, Alex, Andreas Neuhierl, and Michael Weber, 2021, Estimating the Anomaly Base Rate, *Journal of Financial Economics*, 140, 101-126.

- Chordia, Tarun, Amit Goyal, and Alessio Saretto, 2020, Anomalies and false rejections, *Review of Financial Studies* 33, 2134-2179.
- Cochrane, John H, 2011, Presidential address: Discount rates, *Journal of Finance* 66, 10471108.
- Duffie, Darrell, 2010, Presidential address: Asset price dynamics with slow-moving capital, *Journal of Finance*, 65, 1237-1267.
- Fama, Eugene F, 1998, Determining the number of priced state variables in the ICAPM, *Journal of Financial and Quantitative Analysis* 33, 217-231.
- 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, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55-84.
- Fama, Eugene F, and Kenneth R French, 2004, New lists: Fundamentals and survival rates, *Journal of Financial Economics* 73, 229-269.
- Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1-22.
- Fama, Eugene F, and Kenneth R French, 2018, Choosing factors, *Journal of Financial Economics* 128, 234-252.
- Farmer, Leland, Lawrence Schmidt, and Alan Timmermann, 2022, Pockets of Predictability, *Journal of Finance*, Forthcoming.
- Gibbons, Michael, Stephen Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica*, 57, 1121-52.
- Goodwin, Thomas H, 1998, The information ratio, *Financial Analysts Journal* 54, 34-43.
- Goyal, Amit, and Pedro Santa-Clara, 2003, Idiosyncratic risk matters!, *Journal of Finance* 58, 975-1007.
- Grullon, Gustavo, Yelena Larkin, and Roni Michaely, 2019, Are US industries becoming more concentrated?, *Review of Finance* 23, 697-743.
- Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *Review of Financial Studies* 26, 5-68.
- He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics* 126, 1-35.
- He, Zhiguo and Arvind Krishnamurthy, 2013, Intermediary Asset Pricing, *American Economic Review*, 103, 732-770.
- Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical Asset Pricing via Machine Learning, *Review of Financial Studies*, 33, 2223-2273.

- Haddad, Valentin, Serhiy Kozak, and Shrihari Santosh, 2020, Factor Timing, *Review of Financial Studies*, 33, 1980-2018.
- Hodrick, Robert J., and Xiaoyan Zhang. "Evaluating the specification errors of asset pricing models." *Journal of Financial Economics* 62, no. 2 (2001): 327-376.
- Hou, Kewei, G. Andrew Karolyi, and Bong-Chan Kho. "What factors drive global stock returns?." *The Review of Financial Studies* 24, no. 8 (2011): 2527-2574
- Hou, Kewei, and David T Robinson, 2006, Industry concentration and average stock returns, *Journal of Finance* 61, 1927-1956.
- Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650-705.
- Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019-2133.
- Iilmanen, Antti, Ronen Israel, Tobias J Moskowitz, Ashwin K Thapar, and Rachel Lee, 2021, How do factor premia vary over time? A century of evidence, *Journal of Investment Management* 19, 15–57.
- Jensen, Theis, Bryan T Kelly, and Lasse Heje Pedersen, 2021, Is there a replication crisis in finance?, NBER Working Paper .
- Kahle, Kathleen M, and Rene M Stulz, 2017, Is the US public corporation in trouble?, *Journal of Economic Perspectives* 31, 67-88.
- Kashyap, Anil, Natalia Kovrijnykh, Jian Li, and Anna Pavlova, 2021, The benchmark inclusion subsidy, *Journal of Financial Economics*, 142, 756-774.
- Kelly, Bryan, Seth. Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics*, 134, 501-524.
- Keloharju, Matti, Juhani T. Linnainmaa, and Peter Nyberg. "Long-term discount rates do not vary across firms." *Journal of Financial Economics* 141, no. 3 (2021): 946-967.
- Koijen, Ralph SJ, and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475-1515.
- Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, Interpreting factor models, *Journal of Finance*, 133, 1183-1223.
- Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross-section, *Journal of Financial Economics*, 135, 271-292.
- Lewis, Ryan, and Shrihari Santosh, 2012, Investor betas, working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3739424.
- Linnainmaa, Juhani T, and Michael R Roberts, 2018, The history of the cross-section of stock returns, *Review of Financial Studies* 31, 2606-2649.

- Lopez-Lira, Alejandro, and Nikolai Roussanov, 2021, Do common factors really explain the cross-section of returns? Working paper, https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3628120.
- McLean, R David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *Journal of Finance* 71, 5-32.
- Pastor, Lubos, and Robert F Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642-685.
- Pastor, Lubos, and Pietro Veronesi, 2009, Learning in financial markets, *Annual Review of Financial Economics*, 1, 361-381.
- Pontiff, Jeffrey, 2006, “Costly Arbitrage and the Myth of Idiosyncratic Risk,” *Journal and Accounting and Economics*, Vol. 42, 35-52.
- Sharpe, William F, 1994, The Sharpe Ratio, *Journal of Portfolio Management* 21, 49-58
- Simoes, Alexander James Gaspar, and Cesar A Hidalgo, 2011, The economic complexity observatory: An analytical tool for understanding the dynamics of economic development, in *Workshops at the twenty-fifth AAAI conference on artificial intelligence*.
- Simpson, E., 1949, Measurement of Diversity, *Nature*, 163, 688-688.
- Stambaugh, Robert F, and Yu Yuan, 2017, Mispricing factors, *Review of Financial Studies* 30, 1270-1315.
- Stulz, Rene M, 2018, The shrinking universe of public firms: Facts, causes, and consequences, *NBER Reporter* 12-15.
- Van Reenen, John, 2018, Increasing Difference Between Firms: Market Power and the Macro Economy, Changing Market Structures and Implications for Monetary Policy, *Kansas City Federal Reserve: Jackson Hole Symposium*, 19-65.

Figure 1. Time series variation in the number of significant factors. This figure shows the time series variation in the number of significant factors based on the alphas obtained from the sample of factor returns. For each factor at each month t , we regress each of the three portfolios' monthly returns from $t - 59$ to t on the market's monthly excess returns to obtain each portfolio's CAPM alpha. To be included, portfolios must have 60 non-missing returns over the alpha estimation period. For each month t , we count the number of significant factors based on each of the three portfolios' alphas. A factor is significant at month t if the t -statistic of its CAPM alpha on a given portfolio exceeds defined thresholds. The dotted blue line shows the number of significant factors for a t -statistic cutoff of 1.96, the solid orange line uses a t -statistic cutoff of 3.00 and the dashed grey line uses a t -statistic cutoff of 4.00.

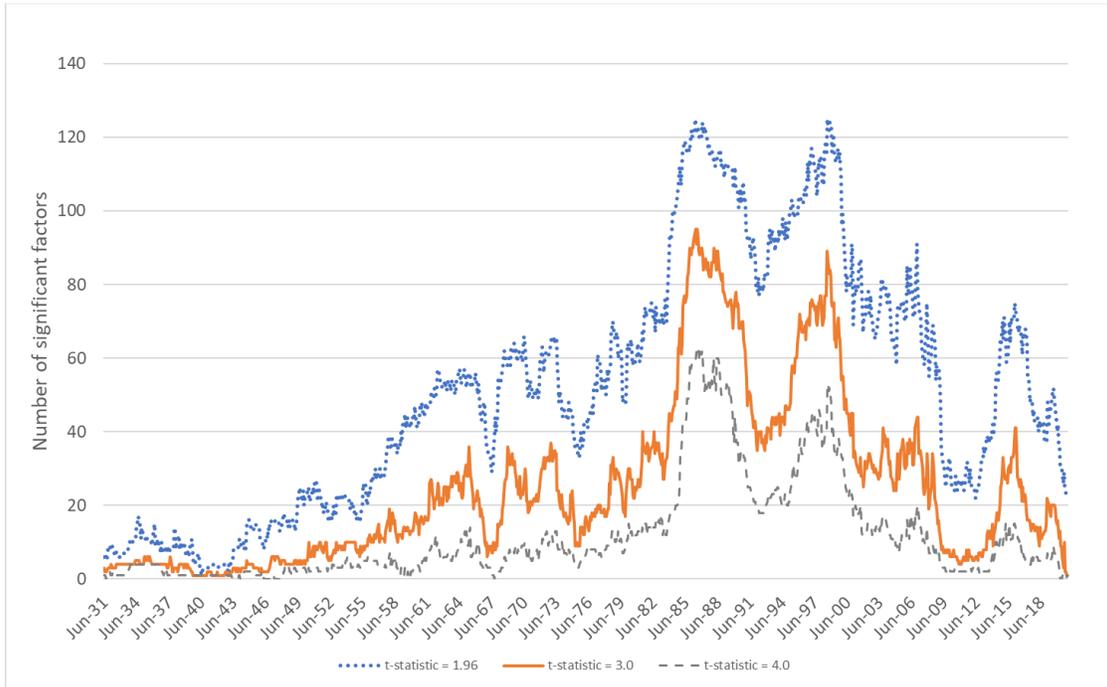
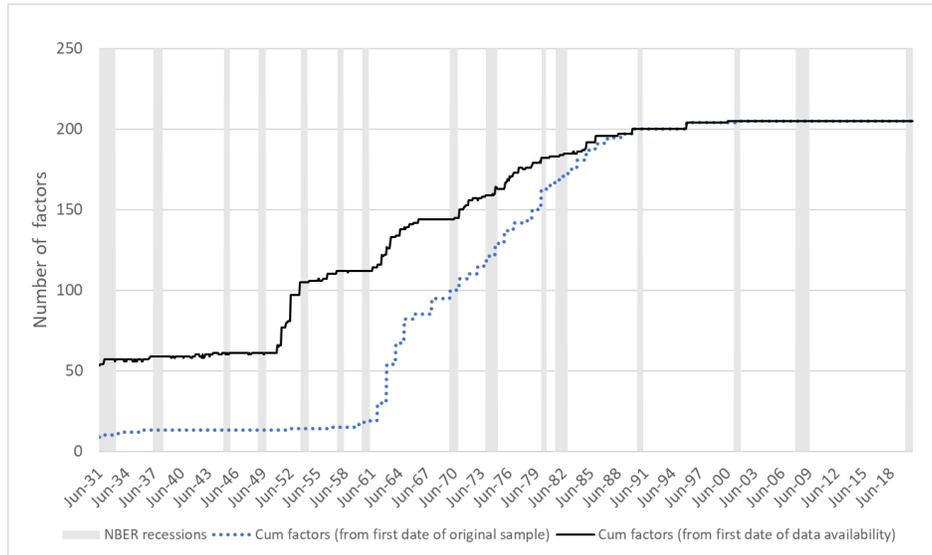


Figure 2. Time series variation in number of and significance of factors. This figure shows the cumulative factors over time as documented in the finance literature. The sample of factors comes from the set of “clear” and “likely” predictors provided by Chen and Zimmermann (Forthcoming) from 1931 to 2020. Panel A shows the cumulative number of factors over time computed in two ways. The solid black line is incremented at the date of each factor’s first available return given the data available today. The dashed blue line is incremented at the date of each factor’s first available return based on the time period of the data used in the original paper’s sample. Panel B relates the number of significant factors to the variation in the standard deviation of stock-level CAPM alphas. For each month t , we regress each factor’s (stock’s) monthly returns from $t - 59$ to t on the market’s monthly excess returns to obtain each factor’s (stock’s) CAPM alpha. To be included, factors (stocks) must have 60 non-missing returns over the alpha estimation period. A factor is counted significant at month t if the t -statistic of its CAPM alpha exceeds 3.00. The solid blue line shows the standard deviation of all stock-level CAPM alphas computed at month t . The dashed orange line shows the number of significant factors at month t . The grey vertical bars represent periods of NBER-defined recessions. A regression of the number of significant factors on the standard deviation of alphas has a beta of 9.80 and a t -statistic of 4.69.

Panel A: Cumulative factors over time



Panel B: Number of significant factors relative to stock-level alpha dispersion

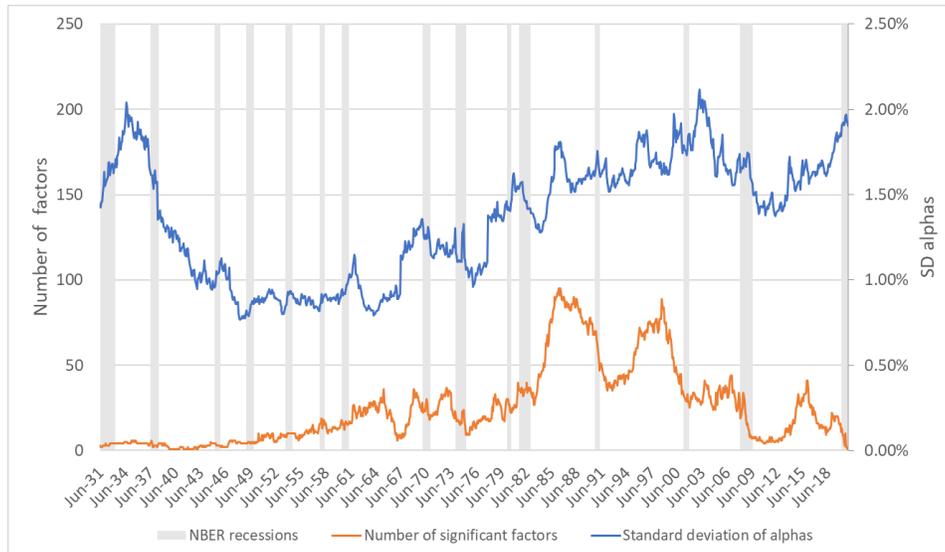


Figure 3. Time series variation in alphas of various asset pricing models. This figure shows the time series variation in the standard deviation of stock-level alphas obtained from various asset pricing models for all common stocks in the CRSP universe. For each month t , a stock's excess monthly returns from $t - 59$ to t are regressed on factors of various asset pricing models to obtain an alpha relative to that asset pricing model. To be included, a test asset is required to have 60 non-missing returns over the estimation period. At each month t , we plot the standard deviation of all alphas obtained from a specific asset pricing model. We use four samples of test assets when computing the alphas. Figure A4 shows the analog for 25 size/BM portfolios and FF30 industry portfolios. The asset pricing models include the Capital Asset Pricing Model (CAPM), the Fama and French (1993) 3-factor model (FF3F), the Fama and French (2015) 5-factor model (FF5F), the Fama and French (2018) 6-factor model (FF6F), the FF3F model augmented with Carhart (1997) momentum factor (FF3F+UMD), the FF3F+UMD model augmented with momentum and Pastor and Stambaugh (2003) liquidity factor (FF3F+UMD+PSLIQ), the Stambaugh and Yuan (2017) factor model (M4), the Barillas and Shanken (2018) 6-factor model (BS6F) and the Hou et al. (2015) q-factor model (Q4). The standard deviation of alphas is in percent per month. The grey vertical bars represent periods of NBER-defined recessions.

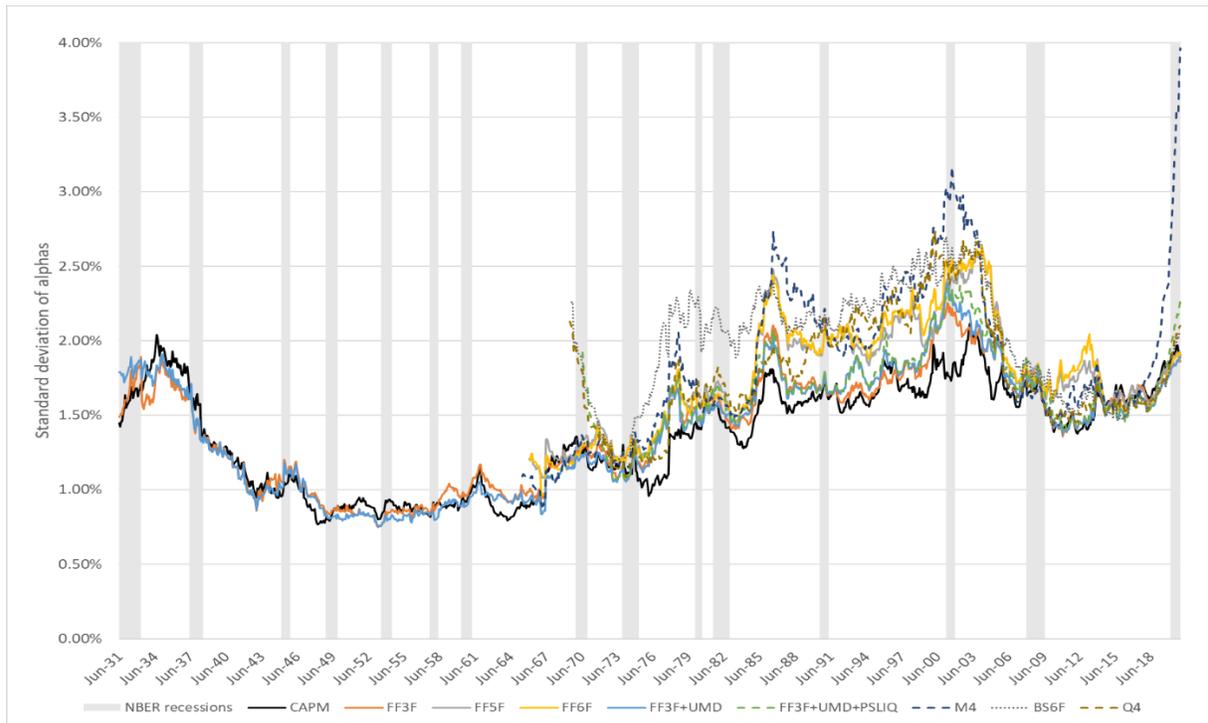
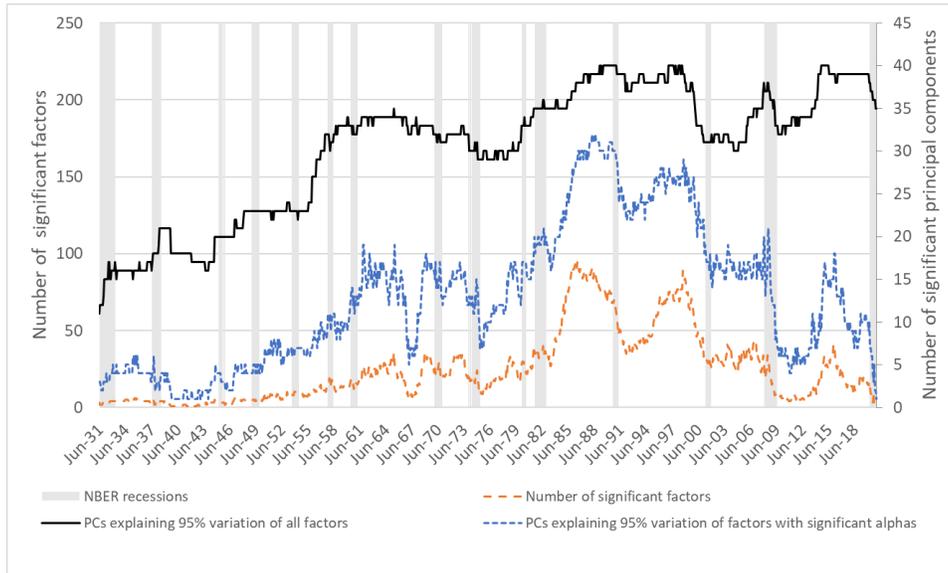


Figure 4. Time series variation in principal components of factors. Panel A shows the variation in the number of significant principal components in the sample of 205 factors across time. For each month t , we regress each factor's monthly returns from $t - 59$ to t on the market's monthly excess returns to obtain each factor's CAPM alpha. A factor is significant at month t if the t -statistic of its CAPM alpha exceeds 3.00. To be included, factors must have 60 non-missing returns over the alpha estimation period. The orange dashed line shows the number of significant factors at each date. We also compute the number of significant principal components at each month t by counting the number of principal components required to explain 95% of the cumulative variation of a set of factor returns from $t - 59$ to t . We compute the number of principal components based on two rolling samples of factors. The black solid line shows the number of significant principal components for the sample of all factors. The blue dotted line shows the number of significant principal components for the sample of factors which have a significant CAPM alpha over the previous 60 months. The grey vertical bars represent periods of NBER-defined recessions. Panel B shows the cumulative number of principal components at each date that make up different percentages of total variation.

Panel A. Time series variation in principal components of factors



Panel B. Cumulative variation explained by principal components of all factor returns

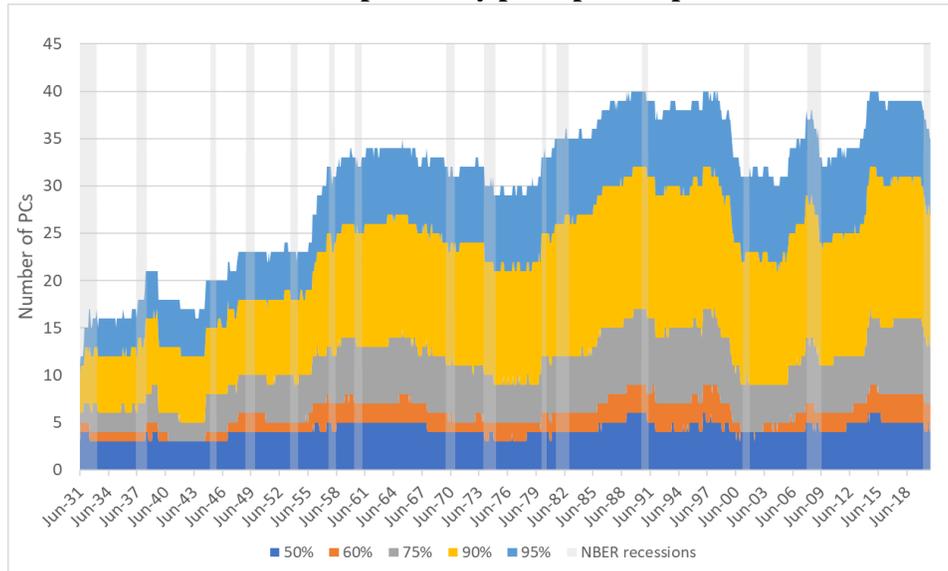


Figure 5. Cumulative in-sample Sharpe ratios of portfolios obtained from principal components. This figure shows the marginal in-sample Sharpe ratio contributions that additional principal components add to an optimal tangency portfolio. From 1968-2020, at each month t , we compute the principal components of the factors' monthly returns from $t - 59$ to t . Factors must have 60 non-missing returns over the analysis window to be included in the sample at a given month t . We construct up to 59 portfolios by incrementally adding a principal component to the portfolio. The weights in each portfolio are chosen to optimize the portfolio's Sharpe ratio.

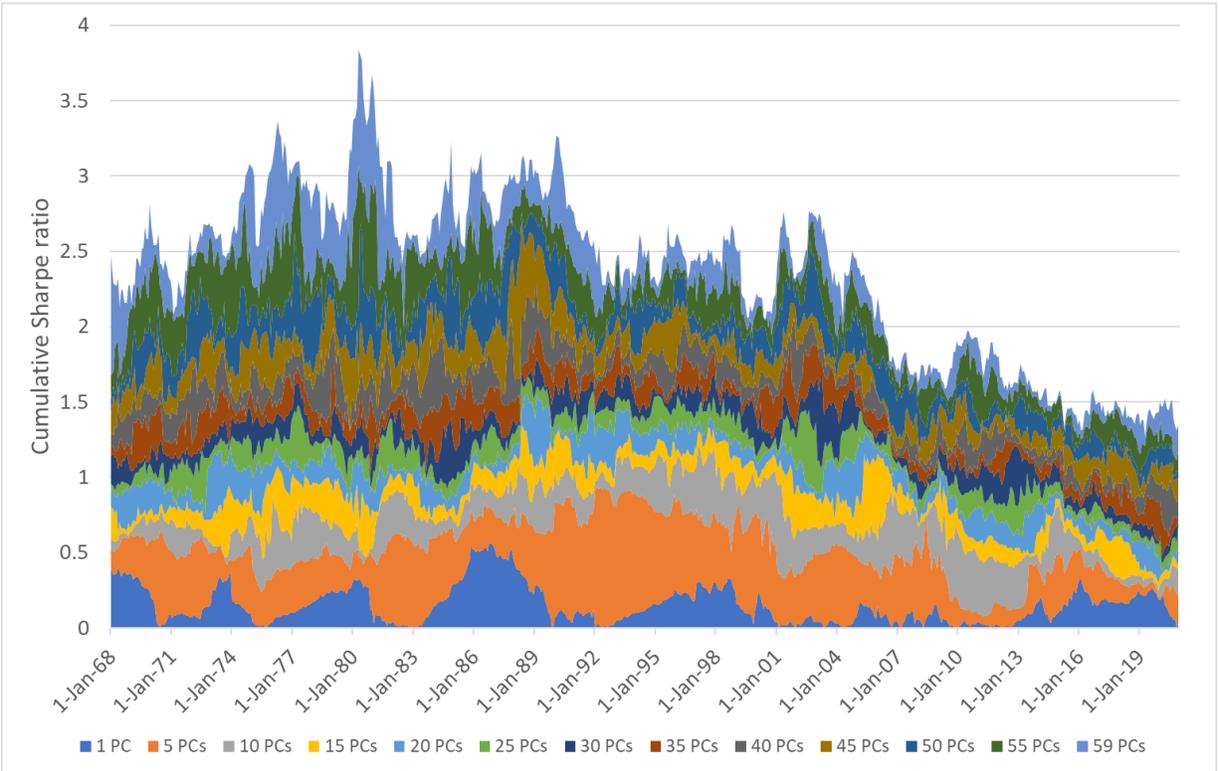


Figure 6 – Number of PCs and Sharpe ratio of portfolios constructed from out-of-sample principal components. These figures show the number of PCs and the maximum Sharpe ratio for portfolios formed from varying number of principal components. The PCs are calculated from the factor returns from $t-59$ to t . The portfolios are formed using the out-of-sample data from $t+1$ to $t+36$. See section 3.C for the full methodology. Panels A and C compare the number of principal components that form a portfolio with the highest Sharpe ratio. The orange dots in panels A and C are the number of principal components that form a portfolio with the maximum Sharpe ratio and the blue lines are the number of significant factors at month t . Panels B and D show the Sharpe ratios of portfolios formed from 4 different sets of PCs: 1) the first 5 PCs, 2) the average Sharpe ratio across all portfolios formed by increasing numbers of PCs, 3) the maximum number of principal components, and 4) the number of PCs that form the portfolio with the maximum Sharpe ratio. Panels A and B show results for PCs formed from all factors. Panels C and D shows results based on the subset of significant factors.

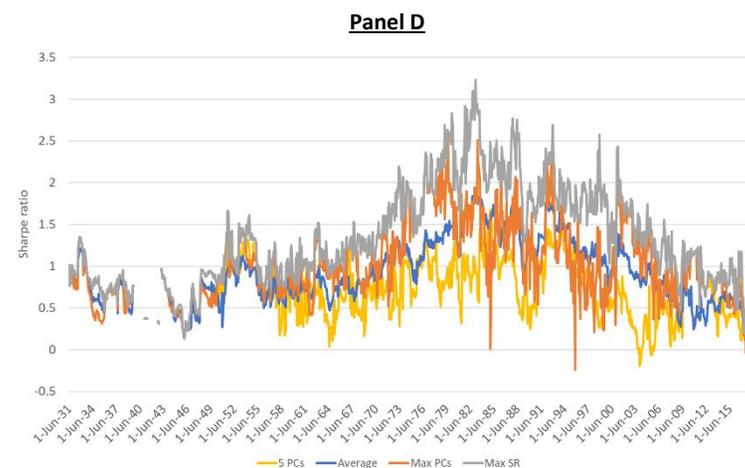
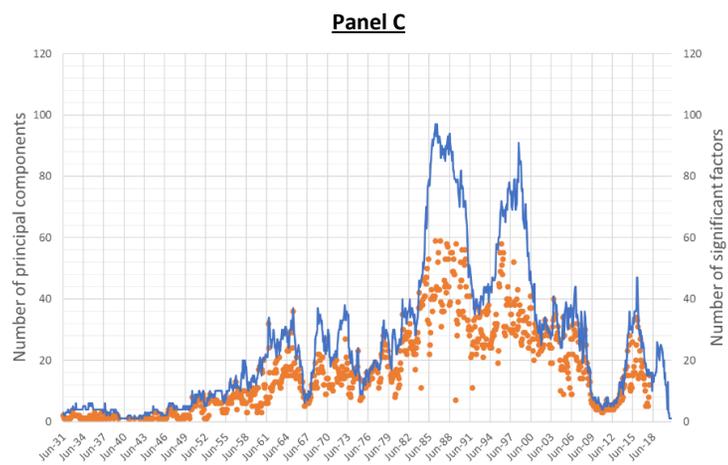
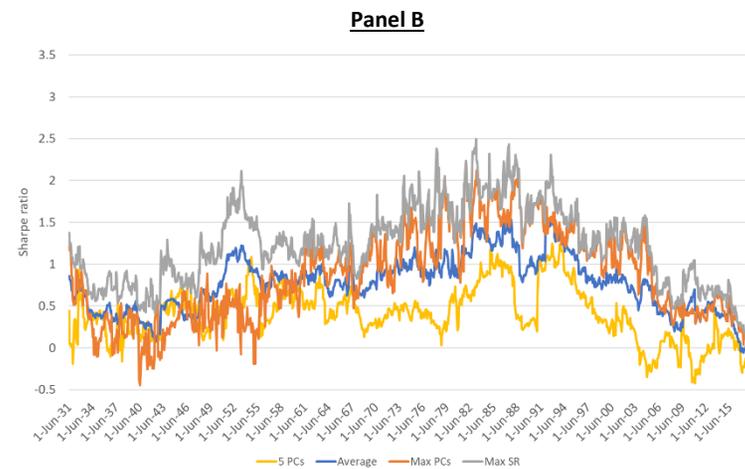
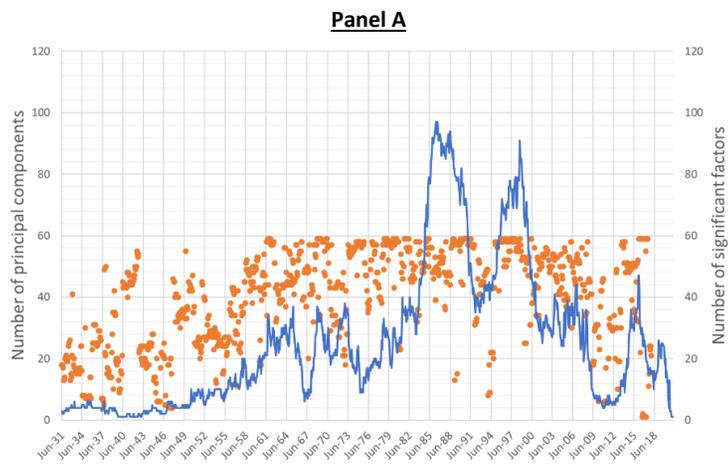
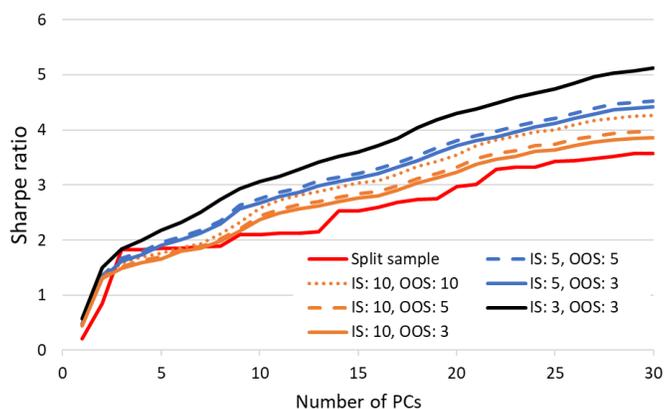
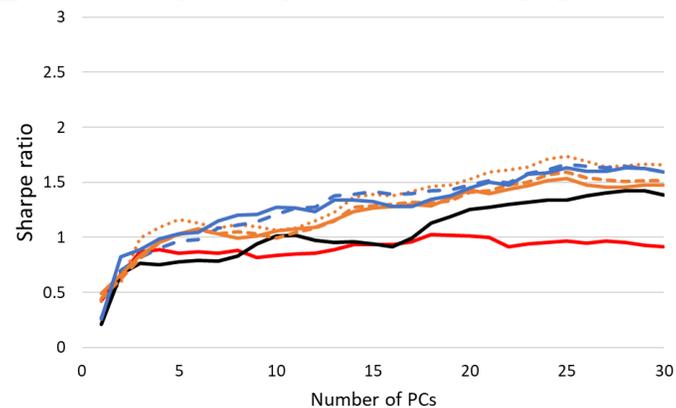


Figure 7. Effect of principal components' estimation and forecasting horizons on out-of-sample Sharpe ratios. This figure shows the average Sharpe ratios of portfolios formed from increasing numbers of principal components for different estimation windows. The data sample consists of the short and long legs of 15 anomalies from 1965 to 2015 as provided by Kozak, Nagel and Santosh (2018). Panel A shows the averages of rolling in-sample Sharpe ratios obtained from forming optimal portfolios consisting of different numbers of principal components. Panel B shows the averages of rolling out-of-sample Sharpe ratios. Panel C shows the averages of rolling out-of-sample Sharpe ratios for the 25 size and book-to-market portfolios. Panel D shows the averages of the corresponding out-of-sample Sharpe ratios for the same assets. The principal components are computed on a rolling monthly basis for different estimation windows. The out-of-sample portfolios are constructed using the in-sample optimal weights. The solid red line shows the Sharpe ratios where the sample is split in half and replicates Kozak, Nagel and Santosh (2018). The other lines present in-sample and out-of-sample windows from $t - k$ to t , where k can be 10, 5 or 3 years of rolling months of daily returns.

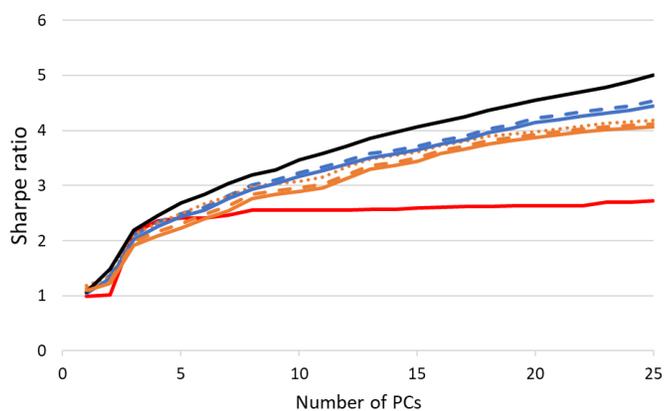
Panel A: In-sample Sharpe ratios – short/long legs of 15 anomalies



Panel B: Out-of-sample Sharpe ratios – short/long legs of 15 anomalies



Panel C: In-sample Sharpe ratios – 25 Size/BM



Panel D: Out-of-sample Sharpe ratios – 25 Size/BM

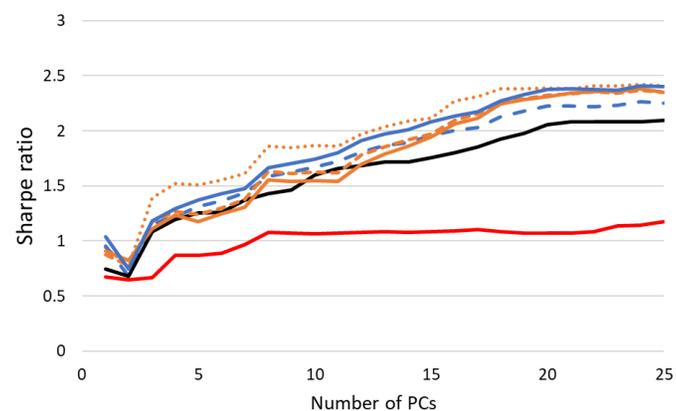


Figure 8. R-squared and Sharpe Ratios based on Kozak, Nagel, and Santosh (2020) (KNS). Panel A replicates Figure 3b in KNS, and displays the out-of-sample r-squared implied by a range of possible priors regarding the Sharpe ratio (kappa) and for a range of non-zero coefficients in an SDF formed based on the principal components of the 50 anomaly return series they study. Panel B shows the corresponding out-of-sample Sharpe ratios actually attained, based on a modified version of their computer code, as described in the text. Warmer colors indicate higher outcomes on both Panels. The red line denotes outcomes for the kappa that generates the highest out-of-sample r-squared. The red ‘+’, ‘x’ and ‘•’ denote outcomes when the SDF has non-zero coefficients on 4, 10, and the maximum number of principal components. The ‘•’ also reflects the maximum achievable out-of-sample r-squared in Panel A.

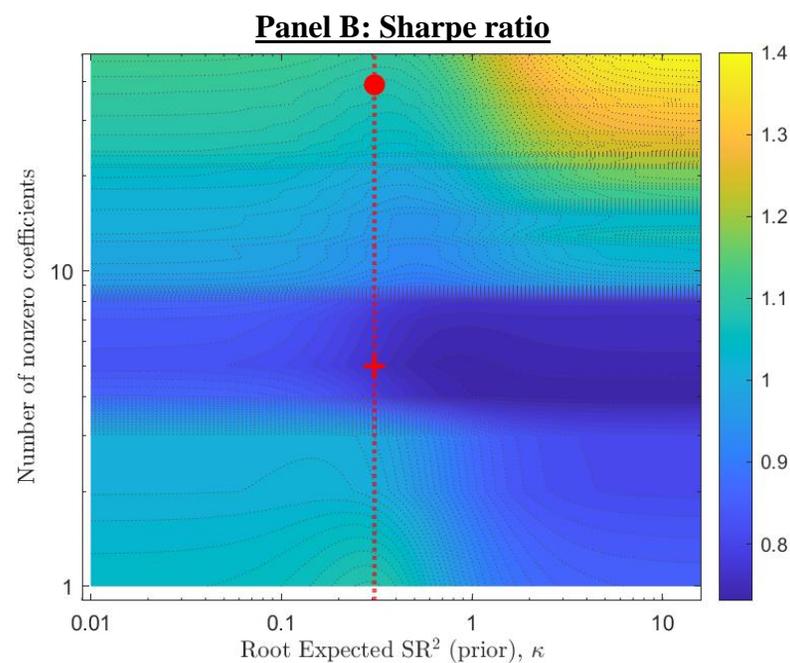
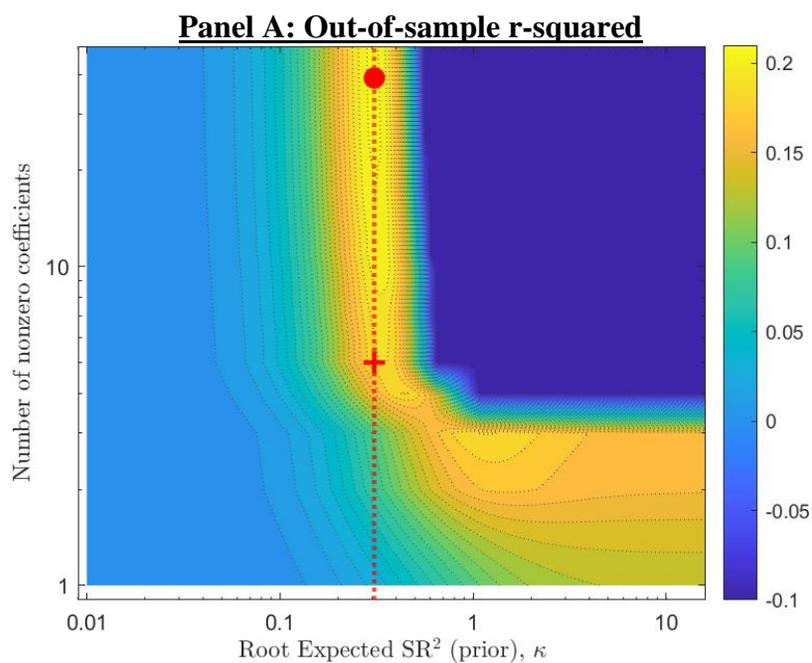


Figure 9. Factors fraction of time significant related to the size of its coefficient in optimal unconditional SDF. This figure shows that the coefficients in the optimal unconditional SDF of Kozak, Nagel and Santosh (2020) are positively related to the fraction of months that a given factor is significant in the sample. The data consists of the 50 anomaly portfolios from Kozak, Nagel and Santosh (2020) for their same sample period. Both axes are based on absolute values, as in that paper, these factor portfolios have not been normalized to have positive premia, but instead are based on long-short portfolios of high and low characteristic firms. The x-axis shows the fraction of time a given factor is significant relative to the CAPM (absolute t-stat greater than 3.0). The y-axis shows the absolute value of the coefficient in the optimal unconditional SDF as reported in Kozak, Nagel and Santosh (2020). The full set of coefficients are obtained from the code posted on Kozak’s website. The red dashed line shows the best fit line with the parameters and statistics reported in the text in the figure.

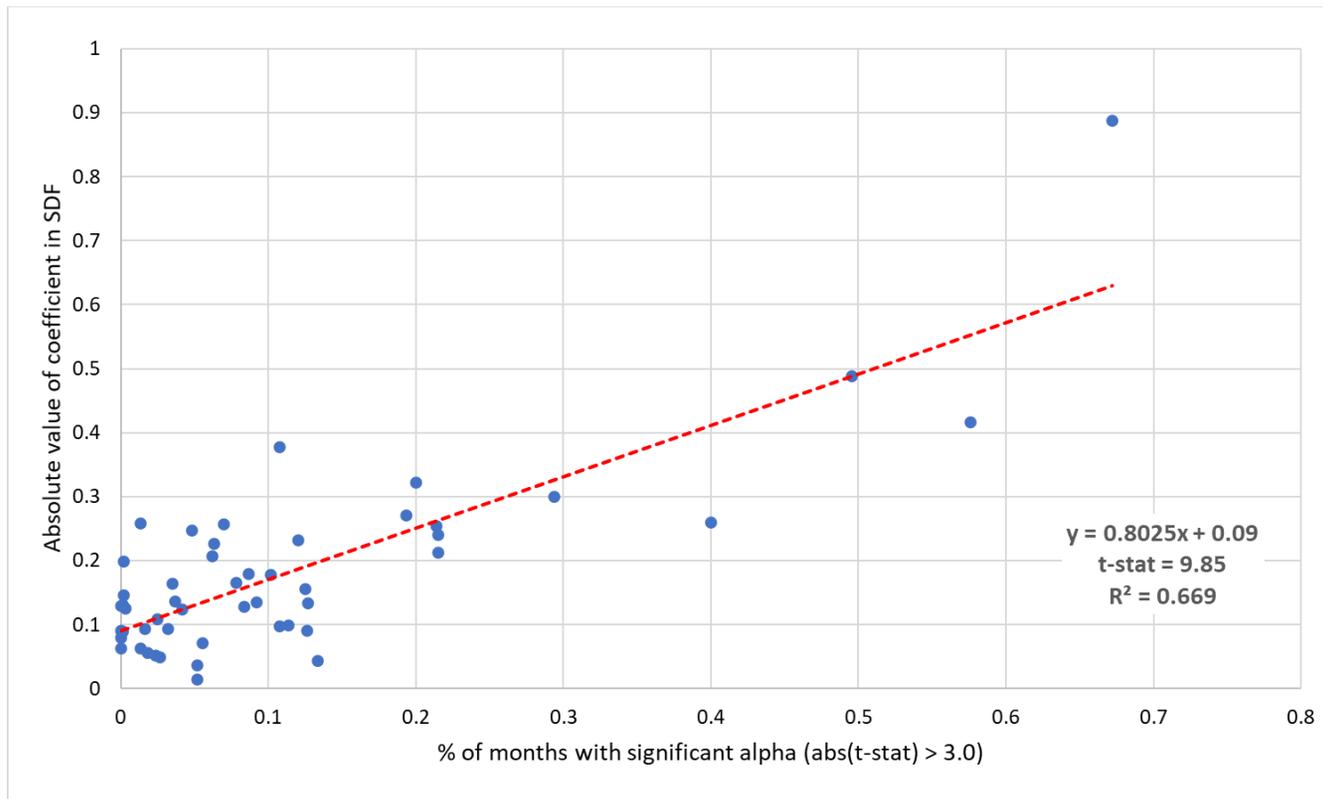


Figure 10. Time series of factor significance This figure shows whether a factor at a specific date has a statistically significant CAPM alpha over the preceding 60 months. For each month t , we regress each factor's monthly returns from $t-59$ to t on the market's monthly excess returns to obtain each factor's CAPM alpha. A factor is significant at month t if the t -statistic of its CAPM alpha exceeds 3.0. Factors must have 60 non-missing returns over the alpha estimation period. Each horizontal series represents a different factor with the blue or grey dots signifying a month in which the factor is significant. A green dot denotes the earliest data used in the original study that identified the factor. A red dot denotes the latest data used in the original study. A magenta dot indicates the earliest date for which we are able to estimate the factor's alpha based on data now available. The left vertical axis lists the category of factors according to Chen and Zimmermann (2021). Categories are assigned primarily based on the data source underlying the characteristic used to form the factor. The grey vertical bars represent periods of NBER-defined recessions.

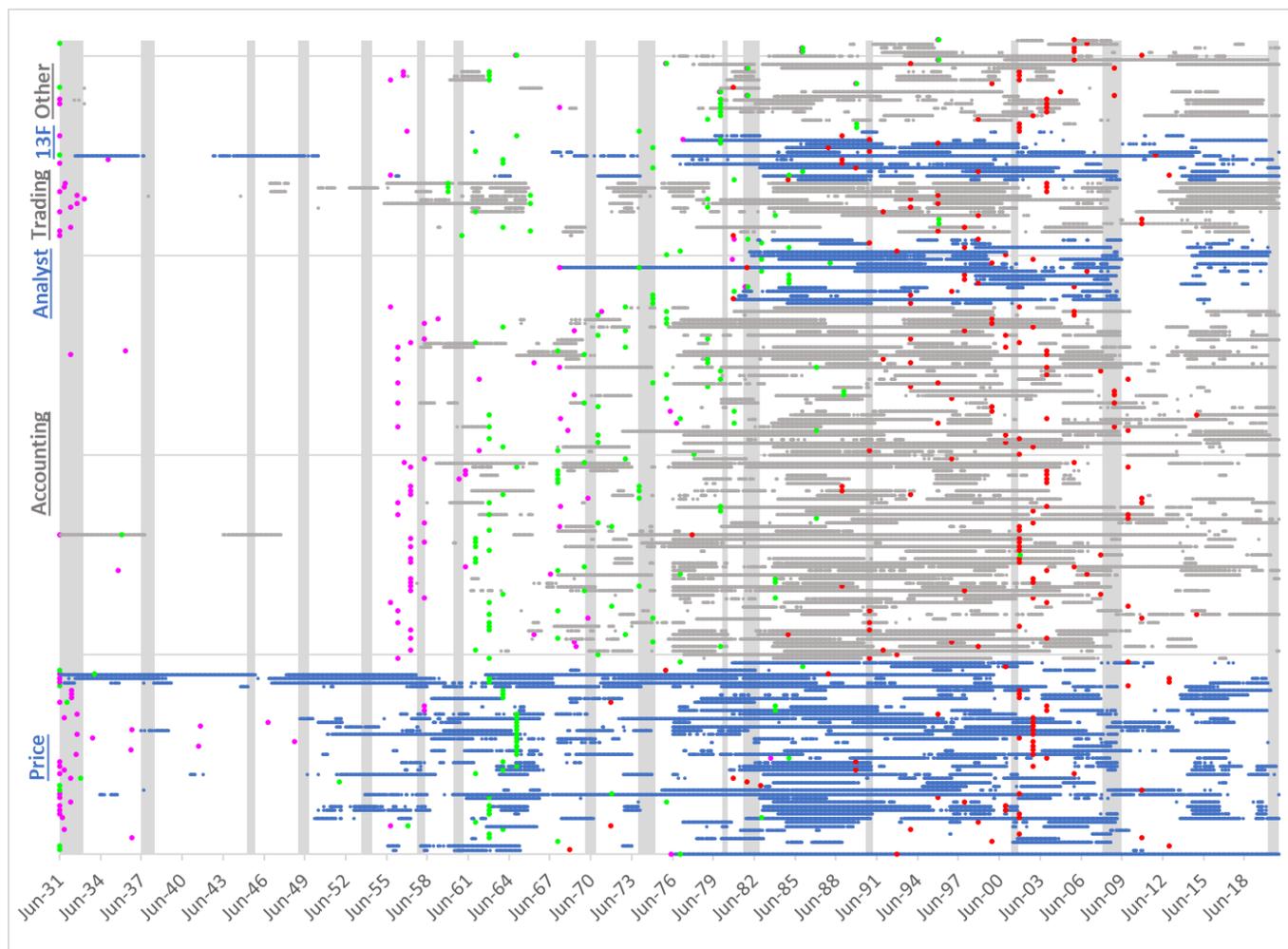


Figure 11. Simulated distributions for average significance spell lengths and counts. This figure shows simulated distributions of the cross-factor average of individual factors' average spell lengths and spell counts, when each factor has a constant return premium. We estimate for each factor a constant alpha, beta and residual volatility by means of a regression of factor returns on excess market returns. We create a simulated time series of market returns calibrated to the sample mean and standard deviation of the market over the sample period, and generate a simulated time series of returns for each of the 205 factors based on the simulated market returns, estimated factor alpha and beta, and estimated factor residual volatility, with the length of each factor's simulated time series matched to the number of sample observations for the factor return. We then estimate rolling 60-month regressions of simulated factor returns on simulated market returns, and obtain both the count and average length of significance spells for each simulated factor, when significance is assessed based on t-statistics ranging from 1.96 to 4.00. Having done so, we compute the cross-factor average of the spell counts and spell lengths (corresponding to the sample data reported in Table 4). We repeat the simulation 2,000 times to obtain a distribution of the average cross-sectional factor spell lengths and counts. Panel A displays the simulated distributions for the cross-factor average of the average spell lengths, while Panel B corresponding cross-factor average spell counts. The red dashed lines display the corresponding sample outcomes.

Panel A: Simulated distribution of average spell length

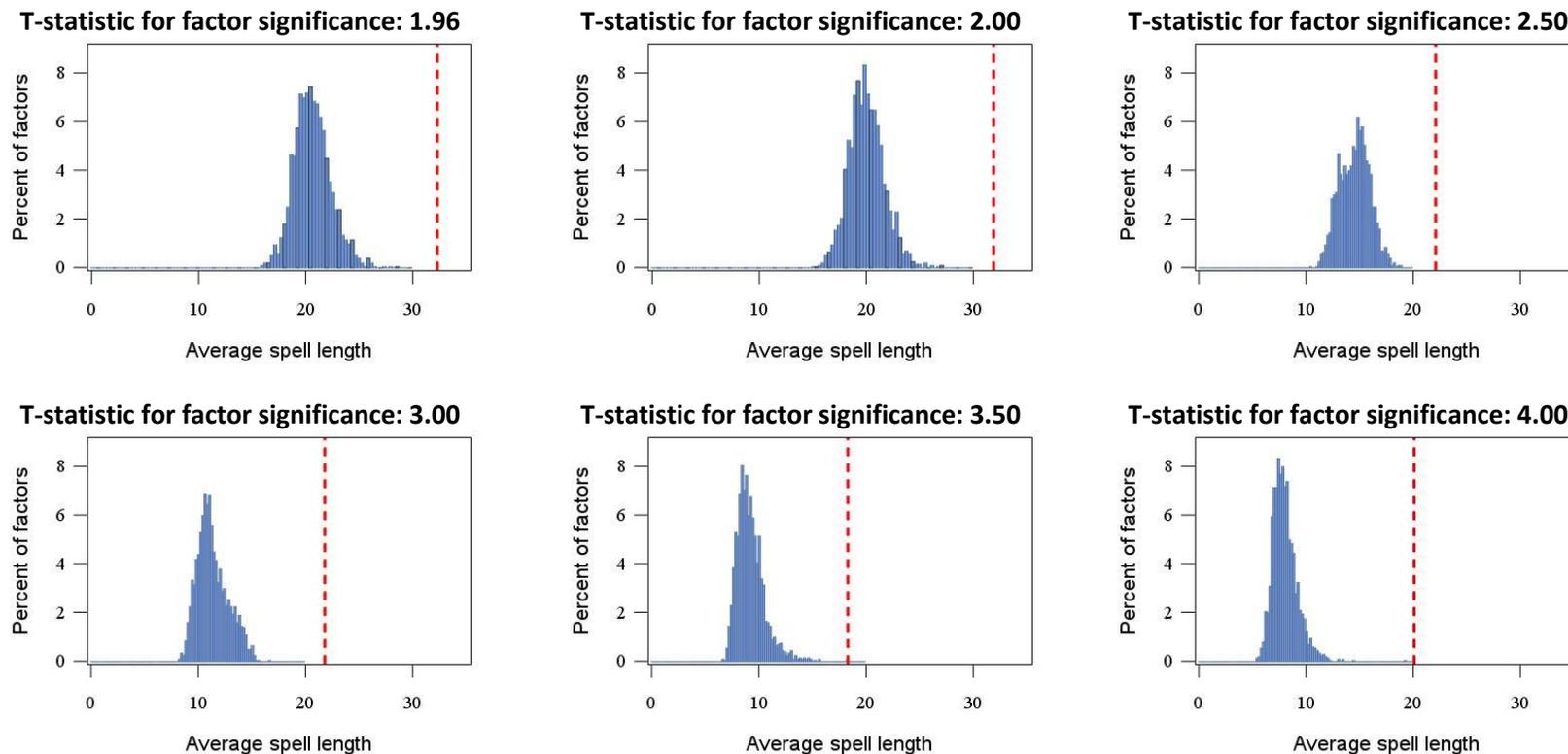


Figure 11 continued.

Panel B: Simulated distribution of average spell count

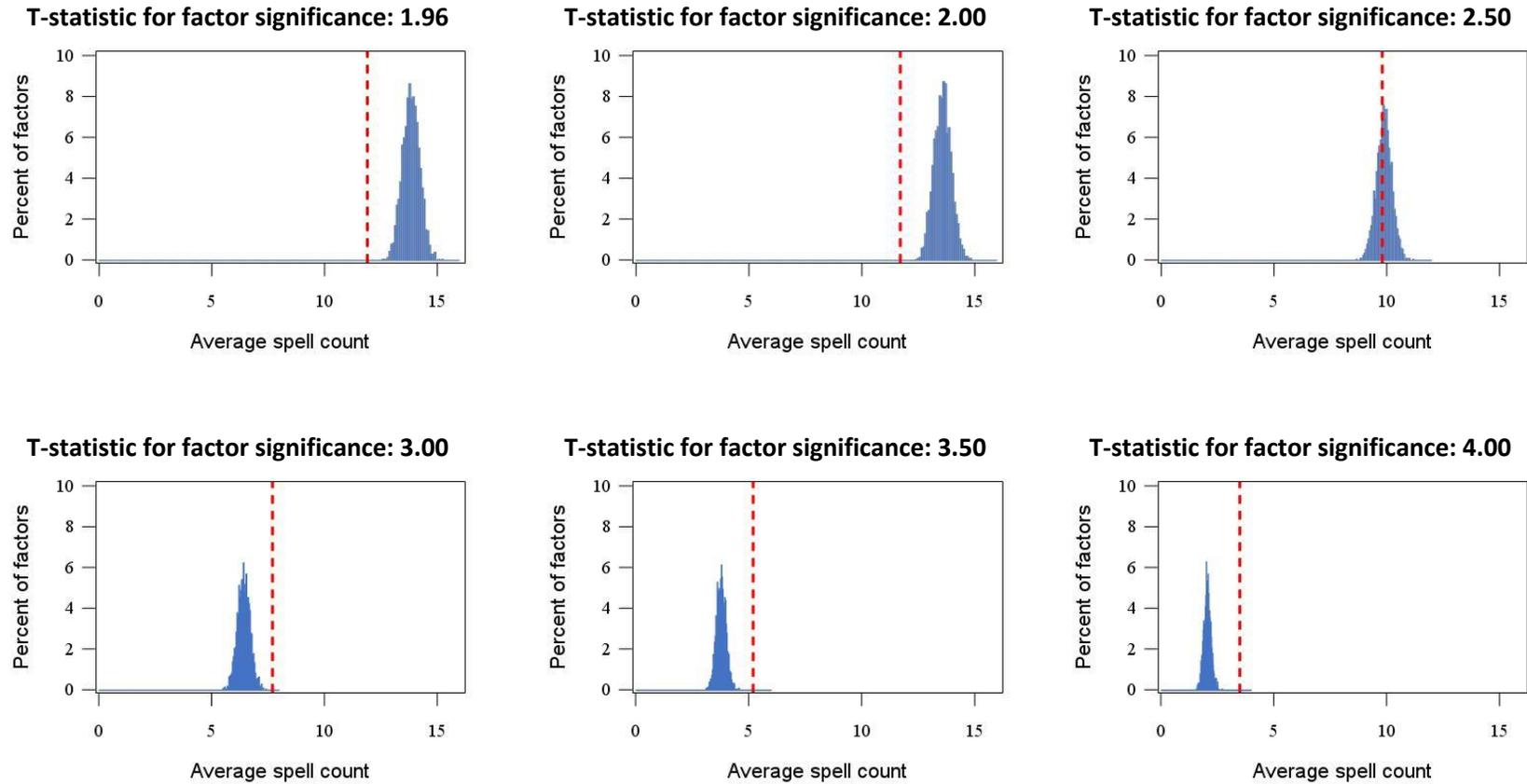


Table 1. Relations between number of significant factors and significant principal components This table shows the results from regressing the number of significant factors on the number of principal components obtained using those factors. For each month t , we regress each factor's monthly returns from $t - 59$ to t on the market's monthly excess returns to obtain each factor's CAPM alpha. A factor is significant at month t if the t -statistic of its CAPM alpha exceeds 3.00. Factors must have 60 non-missing returns over the alpha estimation period. To compute the number of significant principal components at each date t , we count the number of principal components required to explain 95% of the cumulative variation of a set of factor returns from $t - 59$ to t . We compute the principal components from four samples of factor returns: 1) monthly returns of all factors, 2) monthly returns of significant factors, 3) daily returns of all factors, and 4) daily returns of significant factors. We standardize each independent variable by subtracting the mean of that variable over the time series and dividing that difference by the variable's standard deviation over the time series. Hansen-Hodrick standard errors with a bandwidth of 60 are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

	Dep var: Number of significant factors			
	(1)	(2)	(3)	(4)
PCs monthly factors	3.62*** (0.63)			
PCs monthly sign. factors		3.98*** (0.17)		
PCs daily factors			1.04*** (0.13)	
PCs daily sign. factors				1.55*** (0.14)
constant	-58.51*** (14.89)	0.15 (2.48)	-16.34*** (5.24)	17.36*** (4.91)
R-squared	0.65	0.94	0.75	0.83
N	1075	1075	1075	1075

Table 2. Relation between the number of significant factors in-sample and out-of-sample Sharpe ratios. This table shows the results of a regression of the maximum Sharpe ratios and number of principal components that generate those Sharpe ratios on the number of factors that are statistically significant in sample. The maximum Sharpe ratio is obtained from the optimal portfolio of out-of-sample PCs and the number of PCs that make up the maximum Sharpe ratio portfolio. The PCs are calculated from the factor returns from $t-59$ to t . The portfolios are formed using the out-of-sample data from $t+1$ to $t+36$. See section 3.C for the full methodology. The regression results correspond to Figure 9. Hansen-Hodrick standard errors with a bandwidth of 60 are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

	Principal components obtained from sample of:			
	All factors		Significant factors	
	Max SR	Num PCs	Max SR	Num PCs
Number of significant factors	(1)	(2)	(3)	(4)
	0.01***	0.35***	0.02***	0.54***
	(0.00)	(0.08)	(0.00)	(0.03)
Intercept	0.97***	31.28***	0.93***	2.56**
	(0.13)	(3.41)	(0.14)	(1.17)
R-squared	0.26	0.27	0.40	0.82
N	1039	1039	1005	1005

Table 3. Out-of-sample Sharpe ratios for portfolios consisting of varying numbers of principal components.

This table provides descriptive statistics regarding the Sharpe ratios shown in Figures 6 and 7. Panel A tabulates statistics for Figure 6 Panels B and D. It shows the out-of-sample Sharpe ratios of portfolios formed from 4 different sets of PCs: 1) the first 5 PCs, 2) the average Sharpe ratio across all portfolios formed by increasing numbers of PCs, 3) the maximum number of principal components, and 4) the number of PCs that form the portfolio with the maximum Sharpe ratio. The “Difference from SR of 5 PCs” shows the increase in the Sharpe ratio of a given portfolio relative to the portfolio formed by the first 5 PCs. The PCs are calculated from the factor returns from t-59 to t. The portfolios are formed using the out-of-sample data from t+1 to t+36. See section 3.C for the full methodology. Panel B tabulates statistics for Figure 7 panels B and D. It shows the increase in the out-of-sample Sharpe ratio between a portfolio consisting of the first 5 principal components and the portfolio of the maximum number of principal components for two different sets of test assets as used in Kozak, Nagel and Santosh (2018). The two sets of test assets are the long and short legs of 30 anomalies and the 25 size-B/M portfolios. The principal components are computed on a rolling monthly basis for different in-sample and out-of-sample estimation windows. The out-of-sample portfolios are constructed using the in-sample optimal weights. The in-sample window of 25 and out-of-sample window of 25 replicates Kozak, Nagel and Santosh (2018) and matches the red solid lines in Figure 7.

Panel A: Out-of-sample Sharpe ratios for portfolios of varying number of PCs

	PCs from all factors				PCs from sig factors only			
	5 PCs	Average	Max PCs	Max SR	5 PCs	Average	Max PCs	Max SR
Mean	0.382	0.840	1.122	1.350	0.729	1.074	1.185	1.641
Std dev	0.369	0.352	0.496	0.510	0.359	0.399	0.493	0.575
N	600	600	600	600	586	600	600	600
Difference from SR of 5 PCs		0.458	0.740	0.968		0.344	0.456	0.911
% difference		120%	194%	253%		90%	119%	238%

Panel B: Difference in out-of-sample Sharpe ratios of 5 PC and maximum PC portfolios

Rolling estimation window (years)		Test assets	
In-sample	Out-of-sample	Anomalies	Size/BM
25	25	0.061	0.303
10	10	0.499	0.896
10	5	0.623	1.110
10	3	0.603	1.174
5	5	0.490	0.939
5	3	0.447	1.035
3	3	0.556	0.842

Table 4. Summary statistics of factor significance spells across various thresholds of significance. For each month t , we regress each factor's monthly returns from $t - 59$ to t on the market's monthly excess returns to obtain each factor's CAPM alpha. A factor is significant at month t if the t -statistic of its CAPM alpha exceeds one of the various thresholds listed in the table. Factors must have 60 non-missing returns over the alpha estimation period. A significance spell for a given factor is the number of months (i.e., spell length) the factor is continuously significant. Panel A shows the proportion of factors that exhibit at least one significance spell before (after) the sample period of the original paper to identify the factor. Panel B provides summary statistics on the number of significance spells for the cross-section of factors. Panel C computes each factor's average length of a spell and shows summary statistics of this measure for the cross-section of factors conditional on having at least one significance spell. The exceptions are that "Abs min" and "Abs max" show the absolute minimum and maximum spell length of all factors. ***, **, * represent significance at the 10%, 5% and 1% level relative to the simulated distribution in Figure 11.

Panel A: Proportion of factors with at least one significance spell

	t-statistic	p-value	% significant:	
			before original sample	after original sample
	1.96	0.050	77.2	92.6
	2.00	0.046	77.2	92.1
	2.50	0.012	66.9	82.3
	3.00	0.003	54.3	68.5
	3.50	0.001	44.1	50.7
	4.00	0.000	23.6	36.5

Panel B: Number of significance spells per factor

t-statistic	p-value	Cross-sectional statistics of factors' spell counts				
		Mean	SD	Median	Min	Max
1.96	0.050	11.9***	6.9	11	0	35
2.00	0.046	11.7***	6.9	11	0	37
2.50	0.012	9.8	6.5	9	0	33
3.00	0.003	7.7***	5.8	6	0	23
3.50	0.001	5.2***	4.6	4	0	21
4.00	0.000	3.5***	4.2	2	0	19

Panel C: Average length of significance spell

t-statistic	p-value	Cross-sectional statistics of factors' average spells						
		Mean	SD	Median	Min	Max	Abs Min	Abs Max
1.96	0.050	32.3***	52.9	20.0	1.4	535	1	624
2.00	0.046	32.0***	52.6	20.0	1.2	535	1	624
2.50	0.012	22.1***	25.7	15.1	1	233	1	572
3.00	0.003	21.8***	45.5	12.7	1	523	1	523
3.50	0.001	18.3***	30.3	11.4	1	260	1	427
4.00	0.000	20.1***	29.6	12.2	1	233	1	415

Table 5. Relations between the number of significant factors and the standard deviation of stock-level alphas This table shows the results of regressing the number of significant factors in each period on the standard deviation of CAPM alphas for all common stocks in the CRSP universe. For each month t , we regress each factor's monthly returns from $t - 59$ to t on the market's monthly excess returns to obtain each factor's CAPM alpha. A factor is significant at month t if the t -statistic of its CAPM alpha is greater than 3.00. The dependent variable is a count of the number of significant factors at each month t . Stock-level CAPM alphas are obtained by regressing a stock's returns from $t - 59$ to t on the market return. The standard deviation is either equal-weighted or value-weighted (using each stock's market capitalization at $t - 60$). We control for the mean standard error and the mean residual volatility of the stock-level CAPM alphas. We standardize each independent variable by subtracting the mean of that variable over the time series and dividing that difference by the variable's standard deviation over the time series. To be included, stocks and factors must have 60 non-missing returns over the alpha estimation period. Hansen-Hodrick standard errors using a bandwidth of 60 are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

	Dep var: Number of significant factors					
	(1)	(2)	(3)	(4)	(5)	(6)
Equal-weighted (%):						
Standard deviation of alphas	9.80**	22.69*	22.22*			
	(4.69)	(11.79)	(11.37)			
Mean standard error of alphas		-14.47				
		(10.44)				
Mean residual volatility			-14.14			
			(9.93)			
Value-weighted (%):						
Standard deviation of alphas				9.39**	12.57**	12.61**
				(4.68)	(5.62)	(5.54)
Mean standard error of alphas					-4.67	
					(4.05)	
Mean residual volatility						-4.85
						(3.92)
R-squared	0.18	0.26	0.27	0.17	0.19	0.19
N	1075	1075	1075	1075	1075	1075

Table 6. Comovement of the number of significant factors and economy and firm characteristics This table shows the results of regressing the number of significant factors in each period on various economic measures at each month for the sample of factors from 1968-2020. For each month t , we regress each factor's monthly returns from $t - 59$ to t on the market's monthly excess returns to obtain the factor's CAPM alpha and its corresponding standard error. A factor is significant at month t if the t -statistic of its CAPM alpha is greater than 3.00. The dependent variable is a count of the number of significant factors at each month t . The number of public firms is a count of all common stocks at t traded on the NYSE, NASDAQ or Amex at month t . The NBER recession indicator is an indicator equal to one if the month is classified as an NBER recession and zero otherwise. The unemployment rate is the number of unemployed as a percentage of the labor force as provided by the U.S. bureau of labor statistics. The federal funds rate is the established rate by the Federal Reserve at month t . The 10-year treasury bond yield is the market yield on U.S. treasury securities at a 10-year constant maturity. The percent of dividend-paying firms is the total number of common stocks which have paid a dividend in the previous 12 months divided by the number of firms at month t . The mean institutional ownership is the fraction of a firm's shares outstanding held by 13-f firms. The economic complexity index is a measure of economic complexity used from Simoes and Hidalgo (2011). Diversity of firm characteristics is a measure of diversity in the cross-sectional characteristics across firms. See Appendix Table A1 for a complete description of the measures. See the appendix for a similar analysis using the sample of factors from 1931-2020. Hansen-Hodrick standard errors are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

	Dep var: Number of significant factors											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number of public firms	26.74*** (7.97)										6.03 (4.79)	-4.97 (19.75)
NBER recession indicator		-11.90** (5.58)									-7.29** (3.50)	-2.27 (3.70)
Unemployment rate			-5.30 (4.22)								-10.66*** (2.31)	-9.81*** (1.73)
Fed funds rate				7.58 (5.36)							-5.28*** (1.77)	-2.90 (4.63)
10-year T-Bond yield					9.22* (5.31)						16.51*** (3.44)	1.83 (5.94)
% dividend-paying firms						-12.09* (6.21)					-3.81 (7.80)	-16.19 (17.84)
Mean institutional ownership							-16.81*** (4.46)					-20.14 (13.18)
Economic complexity index								6.79 (4.20)			0.75 (2.24)	-3.91 (2.54)
Mean Amihud illiquidity									8.25 (11.77)		9.43*** (3.52)	0.70 (4.55)
Diversity of firm characteristics										17.23*** (5.71)	10.39** (4.66)	18.51** (7.46)
R-squared	0.50	0.03	0.05	0.12	0.17	0.11	0.46	0.09	0.01	0.38	0.71	0.80
N	636	636	636	636	636	636	483	600	636	600	600	456

Appendix:

I. The Potential Role of Idiosyncratic Volatility: Theory and Intuition

Consider an APT-type return process with the market factor, f , and a vector g of additional non-market factors. The exposure vectors are β and γ .

$$R_{i,t} = \beta_{i,t}f_t + \gamma'_{i,t}g_t + \epsilon_{i,t} \quad (1)$$

Parameter Definitions. We assume for simplicity the market and non-market factors are orthogonal, but this is not necessary for the main points. The idiosyncratic returns ϵ are also orthogonal to all the factors. There are two sets of parameters: those for the factors and those for the distribution of stocks in the cross-section. Though formally we will allow them to be time varying, we will assume for a given sample window that the parameters are fixed to simplify the notation. Further, we will assume the cross-sectional stock level exposures are constant over the sample windows. Also, we will assume all the factors and idiosyncratic shocks are independent across time.

We will then compare the key statistics of interest as a function of these parameters, which one can think of as changing across sample windows. We can generalize to the parameters changing within the sample windows. Doing so leaves the main conclusions, except for in perverse situations.

All parameters are assumed to be normally distributed with mean denoted by μ and variances denoted by Σ (unless a scalar as in the case of idiosyncratic shocks, in which case we will denote the standard deviation as σ). The idiosyncratic shocks are mean zero by construction. The cross-section distribution of the idiosyncratic shock variance must be non-normal. We take no stand on that distribution beyond it having a defined mean: $\bar{\sigma}_t$. There are a finite number of stocks N in the cross-section.

Statistics of interest. In this section we derive how the key statistics used in the empirical section depend upon the parameters of the return generating process. For all statistics we focus on the expected value of

these statistics given the parameters. This expectation is across both time series draws in a particular sample period and across cross-sectional draws of stocks.

Consider the regression of individual stocks over a 60-month sample window on the market factor f :

$$R_{i,t} = \alpha_{i,T} + \beta_{i,T}f_t + \theta_{i,t} \quad \text{for } t \in \{T - 59, \dots, T\} \quad (2)$$

We compute the cross-sectional variance of the alpha estimates and the average of the standard variances. (We focus on variances rather than standard deviations for tractability.) For these calculations, we ignore the variation in the market factor that is not removed by the regressions, as this is not the main focus of the paper.

We then focus on regressions of non-market factor returns on the market factor:

$$g[k]_t = a + b f_t + \eta_{k,t} \quad \text{for } t \in \{T - 59, \dots, T\}, \quad (3)$$

where k denotes the k^{th} non-market factor. We compute the expected probability of the factor having an intercept greater than a statistical cutoff t^* . Because the market and non-market factors are orthogonal, we again will ignore the limited variation introduced by random variation in the market factor.

A. Cross-sectional variance of stock level alphas

The estimated alpha from the stock-level regression, Equation (2), (ignoring the market factor contribution) is

$$\hat{\alpha}_{i,T} = \frac{1}{T} \sum_{\tau=T-59}^T (\epsilon_{i,\tau} + \gamma'_{i,\tau}g_{\tau}) \quad (4)$$

We suppress the time subscripts on the parameters since we are assuming the parameters are constant through the estimation window. The cross-sectional variance of these alphas is

$$\sigma_{\alpha,T,CS}^2 = \frac{1}{N-1} \sum_{i \in N} (\hat{\alpha}_{i,T} - \bar{\alpha}_T)^2 \quad (5)$$

where

$$\bar{\hat{\alpha}}_T = \frac{1}{N} \sum_{i \in N} \hat{\alpha}_{i,T} \quad (6)$$

Plugging Equations (6) and (4) into Equation (5) and taking its expectation gives

$$E[\sigma_{\alpha,T,CS}^2] = \left(\frac{N-1}{N}\right) \left[\frac{\bar{\sigma}_\epsilon^2}{60} + \mu'_g \Sigma_\gamma \mu_g + \frac{1}{60} h(\Sigma_\gamma, \Sigma_g) \right] \quad (7)$$

where h is a symmetric increasing function of the two variances.

Thus, the expected cross-sectional variance is determined primarily by the non-market factor premia μ_g and the cross-sectional exposure to the factors Σ_γ . It is less so determined by the average idiosyncratic risk and the combined variation in the exposure to the factors and factor volatility.

B. Average standard variance of the alphas

The standard variance of the stock-level alphas from Equation (2), ignoring the contribution from the modeled factors realizations, is

$$[SE(\alpha_{i,T})]^2 = \left(\frac{1}{60-2}\right) \left(\frac{1}{60}\right) \sum_{\tau=T-59}^T \left(\epsilon_{i,T} + \gamma'_{i,\tau} - \frac{1}{60} \sum_{s=T-59}^T \{ \epsilon_{i,s} + \gamma'_{i,s} g_s \} \right)^2 \quad (8)$$

Taking expectations and then averaging across firms in the cross-section yields

$$Ave \left(E [SE(\alpha_{i,T})^2] \right) = \left(\frac{1}{60-2}\right) \left(\frac{59}{60}\right) \left[\bar{\sigma}_\epsilon^2 + \mu'_\gamma \Sigma_g \mu_\gamma + h(\Sigma_\gamma, \Sigma_g) \right] \quad (9)$$

Thus, this average standard variance is more heavily dependent upon the idiosyncratic risk and the combined volatility of the factor and factor exposure than the cross-sectional standard deviation, and comprises a good control variable for these components. It also depends upon the interaction of the average factor exposure and the volatility of the factor, which under the assumption that average factor exposures outside the market are close to zero, will contribute a negligible amount. Hence, this makes a good control to remove the parts of the cross-sectional standard deviation of alpha that are of less interest.

Comparing equations (7) and (9) it can be observed that the cross-sectional variance of alpha estimates contains variation attributable to return premia associated with non-market factors while the

average standard error of the alpha estimates does not. Further, the cross-sectional mean exposures to non-market factors, which appear in expression (9), are empirically small (perhaps in part because factor returns are based on long-short portfolios, where the long and short legs have offsetting exposures). The cross-sectional mean non-market factor exposure is .006, while the median is -.015, implying that the $\mu'_\gamma \Sigma_g \mu_\gamma$ term in (9) is relatively unimportant. As a consequence, idiosyncratic return volatility is relatively more important in explaining the average standard error of the alpha estimates, while return premia associated with non-market factors are relatively more important in explaining cross-sectional variation in market-model alpha estimates. Thus, when both the volatility and the average standard error of alpha estimates are included as explanatory variables in the same regression, the coefficient on the former primarily reflects the effect of non-market return factor premia, while the coefficient on the latter primarily reflects average idiosyncratic risk.

C. Probability that unmodeled factors are significant

In the empirical section we measure the statistical (and economic significance) of the unmodeled factors by looking at the t-values of the intercept of the unmodeled factors on the modeled factors. In particular, we calculate the number of significant factors as those with t-values above a cutoff t^* . To understand how the likelihood of a factor being significant is a function of the parameters, we compute the expected power. We begin by considering the case where the factor is directly observable and then generalize to the case where the factor is formed from a long-short portfolio from characteristic-based sorts.

C.1 Power of observable factor

This case corresponds directly to the regression in Equation (3). Again, we ignore the contribution of the random realization of the market factor. The expected power for factor k is a standard calculation:

$$1 - \Phi \left(t^* - \frac{\sqrt{60} \mu_g[k]}{\sigma_{\eta_k}} \right) \quad (10)$$

where Φ is the CDF of a standard normal distribution. As would be anticipated, power increases in the factor premium and decreases in the volatility of the factor, σ_{η_k} . Thus, other things equal there will be a positive correlation between the cross-sectional variance of alphas from individual stock regressions and the number of significant factors.

C.2. Power of unobservable factor

The positive relation between the number of listed firms and the number of significant factors is potentially attributable in part to the increased precision of the factor measurement with more firms. To assess this issue, let the non-market factors be unobserved, but let there be an observable firm level characteristic $C[k]_{i,t}$ that is jointly normal with firm i 's factor exposure $\gamma[k]_{i,t}$. Let the variance-covariance matrix for this joint distribution be $\Sigma_{C,t}$. For conciseness, we assume that the characteristics and exposures are fixed over a sample window. Let us form the factor $G[k]_{i,t}$ during $t \in \{T - 59, \dots, T\}$ from a long-short portfolio. The long and short ends are formed by sorting all stocks at $T - 60$ into Q quintiles based on the characteristic $C[k]_{i,T-60}$.

The factor $G[k]$ so formed will be a function of the number of firms and the number of quintiles. Our interest lies in the expected mean and the expected variance of the factor, as these affect the power to detect its significance. For a finite number of firms, the breakpoints and hence, average exposure, to $g[k]$ will vary from sample to sample based on the random realizations of the characteristics and factor exposures. The amount of variation in these will decrease in the number of firms, however these exposures will not vary in expectation with the number of firms. Hence the contribution to the expected mean and variance of the factor $G[k]$ will not vary from the exposure to $g[k]$ as the number of firms change.

Nevertheless, the total expected variance of $G[k]$ will vary with the number firms due to amount of idiosyncratic risk that diversifies away in the long and short end of the portfolio. The number of firms in each long and short portfolio is $\frac{N}{Q}$. Under the assumption of idiosyncratic risk being independent of everything else, the expected variance of the idiosyncratic risk in each of these portfolios is $\frac{Q}{N} \bar{\sigma}_\epsilon^2$. Because

this risk is additive across the long and short portfolio the total variance of factor $G[k]$ due to this diversified idiosyncratic risk is

$$\frac{2Q}{N} \bar{\sigma}_\epsilon^2. \quad (11)$$

Thus, this component decreases in the number of firms. Since the variance of the factor decreases as we increase the number of firms, the power to detect a factor as significant increases with the number of firms, as shown in (10). Thus, either increasing the factor premium of g or the number of firms (via reduced volatility) can drive the number of significant factors detected. To distinguish whether the relation we find between the number of firms and the number of significant factors is driven by only the later or both parameters, we can take advantage of the standard error of the intercept estimates in the regression

$$G[k]_t = a + b'f_t + \eta_{k,t} \text{ for } t \in \{T - 59, \dots, T\} \quad (12)$$

which is also driven by this idiosyncratic risk in the long-short portfolio (Equation (11)). Thus, we can regress the number of significant factors on the average intercept estimates (alphas) and the average intercept standard errors. The prediction is that we will observe a positive coefficient on the former and a negative coefficient on the later. If both are true, then the relation between the number of significant factors and the number of firms is not simply attributable to increased precision in the observability of the unmodeled factors from an increasing number of firms.

II. Simulation

We simulate a time series of returns for a total of 205 factors for 96 periods $t = 1$ to $t = 96$. All factor returns are assumed to follow a normal distribution. The number of priced factors varies from $n = 1$ to $n = 100$ (as in the actual data), with the remaining factors being unpriced. Priced factors have a positive mean return, μ_p , and standard deviation, σ_p . Unpriced factors have a zero mean return and standard deviation, σ_u . We begin our analysis by assuming all factors to be independent, that is the cross-correlations among factors is 0.

After constructing the simulated factor returns for 96 time periods and 205 factors, we compute the principal component factor loadings for the set of simulated factor returns for the in-sample period of the first 60 months, $t = 1$ to $t = 60$. We construct the out-of-sample principal components by applying the in-sample factor loadings to the out-of-sample simulated data for the final 36 months, i.e., from $t = 61$ to $t = 96$. We then construct in-sample and out-of-sample portfolios with an increasing number of principal components included in each portfolio and measure the Sharpe ratios. More specifically, we construct optimized portfolios based on the first, first two, first three, etc., up to the first 59 principal components. In the in-sample case, portfolio weights are chosen to optimize the portfolio Sharpe ratio. For the out-of-sample portfolios, we use the optimal portfolio weights from the in-sample portfolios. We repeat this simulation 10,000 times resulting in 59 portfolios from 10,000 draws of simulated data. We average the Sharpe ratio within portfolios of the same number of principal components across all simulations, resulting in 59 average Sharpe ratios. We report the maximum Sharpe ratio of the 59 portfolios and the number of principal components that are used to form this Sharpe ratio.

To capture the time variation in the number of significant factors and the Sharpe ratios of those factors, we calibrate the mean of the priced factors, μ_p , and the standard deviations, σ_p , of the priced factors by using the values measured in the actual data during a high Sharpe ratio regime (e.g., around 1986) or during a low Sharpe ratio regime (e.g., around 1967 or 2009 – See figure 6 Panels B and D). We also vary the standard deviation of the unpriced factors, σ_u , between the high and low values observed in the data for each of these two regimes. Table A5 presents the numerical values of these parameters. More

specifically, μ_p varies between 0.4% and 1.6%, σ_p is calibrated to be either 1.8% or 3.2%, and σ_u is calibrated to be either 2.5% or 3.6%. All are monthly values.

The baseline scenario is calibrated to match the parameters measured in the data during a high SR regime with a low standard deviation of unpriced factors. The blue solid line in Figure A6 Panel A presents the maximum portfolio Sharpe ratios of the baseline scenario as a function of the number of priced factors. We find that increasing the number of priced factors while holding other parameters constant increases the out-of-sample Sharpe ratios. This increase occurs at a decreasing rate as more priced factors are likely to be partially redundant with earlier factors. Importantly, at 50 priced factors, the simulation matches the Sharpe ratios of approximately 2.5 and 0.7 measured in the data and shown in Figure 9 Panel B for the high and low regime periods. As expected, the change in the unpriced standard deviation (as shown by the dotted lines), has little effect on the Sharpe ratios, suggesting the out-of-sample Sharpe ratios are a good measure of redundancy.

Panel B of Figure A6 shows the maximum out-of-sample portfolio Sharpe ratios as a function of μ_p for 1, 50 and 100 priced factors. For each case, we vary σ_p between 1.8% and 3.2%, and σ_u between 2.5% and 3.6%. We find that an increase in the mean of the priced factors results in monotonically increasing Sharpe ratios, again confirming that the out-of-sample Sharpe ratios is a reliable measure reflecting the number of priced factors. The figure also reveals that as the unpriced factor standard deviations increase, the out-of-sample Sharpe ratios decrease. This variation in the parameters allows us to match the out-of-sample Sharpe ratios found in the actual data.

Thus far, we have assumed independence between all factors, both unpriced and priced. We now assess the extent to which this assumption has on the out-of-sample Sharpe ratios. We focus on correlations between same-type (i.e., priced/priced or unpriced/unpriced) factors and cross-correlations between different-type (i.e., priced/unpriced) factors. In Figure A6 Panel C, we plot the maximum Sharpe ratios for all correlations varying from 0.00 to 0.20 and as a function of the number of priced factors. We present results for a calibration with the cross-correlations zeroed out and the cross-correlation positive

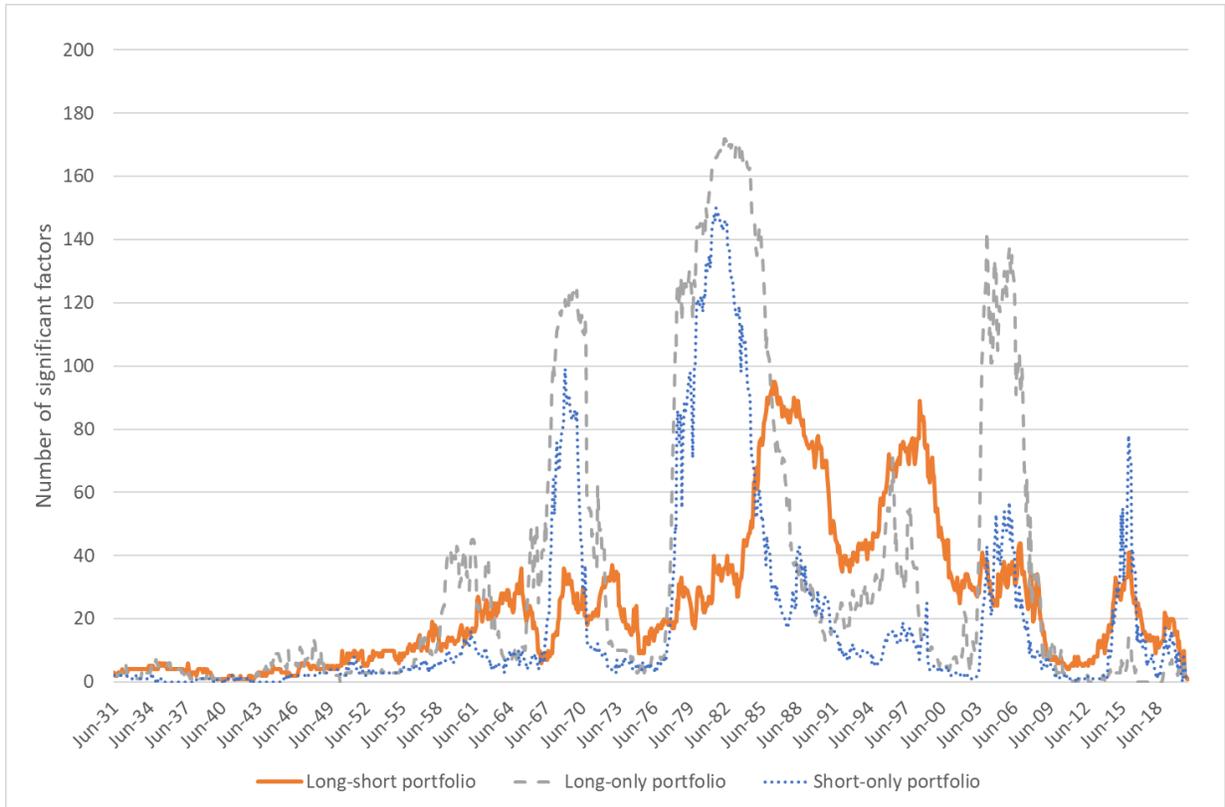
and equal to the same-type correlations. Our first finding is that the maximum out-of-sample Sharpe ratio is decreasing as the mean correlation among factors is increasing. As the correlation increases, the priced factors become increasingly redundant resulting in lower Sharpe ratios. This suggests that the data must have a high number of priced factors as we require very low correlations among priced factors to match the data. We also observe that for correlations greater than 0.05, positive cross correlation among priced and unpriced factors results in increased out-of-sample Sharpe ratios. Essentially, positive cross-correlations allow for free hedges from the unpriced factors. This is not economically sensible. Therefore we focus on the case where these cross-correlations are zero. If the correlations are sufficiently low, the free hedges become small and contribute very little to the diversification effects in the portfolio and, thereby, the out-of-sample Sharpe ratio.

Finally, we explore how the correlation affects the number of principal components that yields the portfolio with the maximum Sharpe ratio as a function of the number of priced factors and mean correlation between same-type factors. Panel D of Figure A6 shows that in the case where the factors are independence, the number of priced factors is positively correlated with the number of principal components generating the portfolio with the maximum Sharpe ratio. This positive correlation is consistent with the data (Panels A and C in Figure 6 and Table 2). This result is also the analog of the hump-shaped pattern of the out-of-sample Sharpe ratios found in Panel B of Figure 7.

In contrast, as the correlation among same-type factors increases, however, we see a strong negative correlation with the number of principal components that make up the out-of-sample portfolio with the highest Sharpe ratio. In essence, the same-type correlations increase the redundancy among priced factors, thereby limiting the number of principal components that provide marginal increases to the Sharpe ratio. That this correlation negative correlation is counter to the data is further evidence against the redundancy of the significant factors we observe in the data. Overall, the simulation provides evidence of limited redundancy in factor data and confirms the magnitudes of out-of-sample Sharpe ratios obtained from the data.

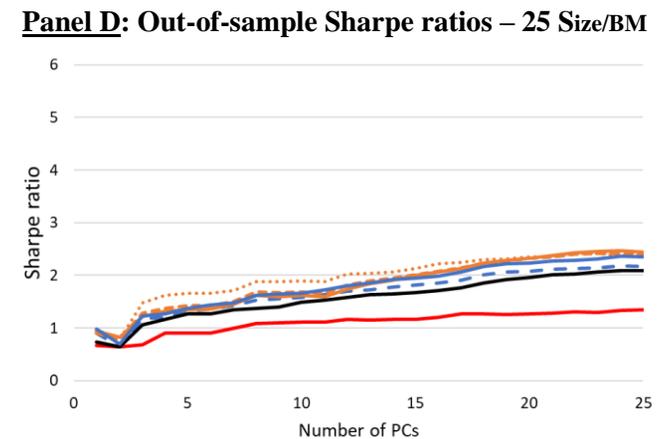
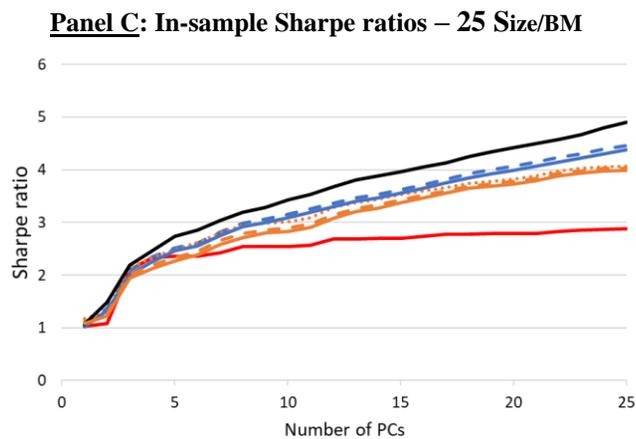
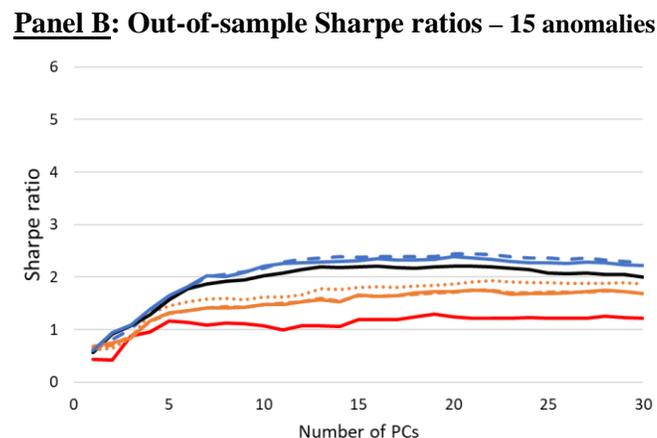
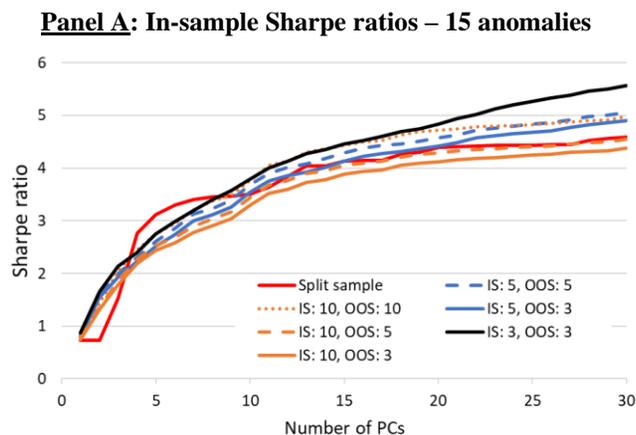
Appendix Figures and Tables

Appendix Figure A1. Time series variation in the number of significant factor portfolios. The time series variation in the number of significant factors based on the alphas of 1) the long-short portfolio of the factor (solid orange line), 2) the long-only portfolio of the factor (dashed grey line), and 3) the short-only portfolio of the factor (dotted blue line). For each factor at each month t , we regress each of the three portfolios' monthly returns from $t - 59$ to t on the market's monthly excess returns to obtain each portfolio's CAPM alpha. To be included, portfolios must have 60 non-missing returns over the alpha estimation period. For each month t , we count the number of significant factors based on each of the three portfolios' alphas. A factor is significant at month t if the t -statistic of its CAPM alpha on a given portfolio exceeds 3.0.



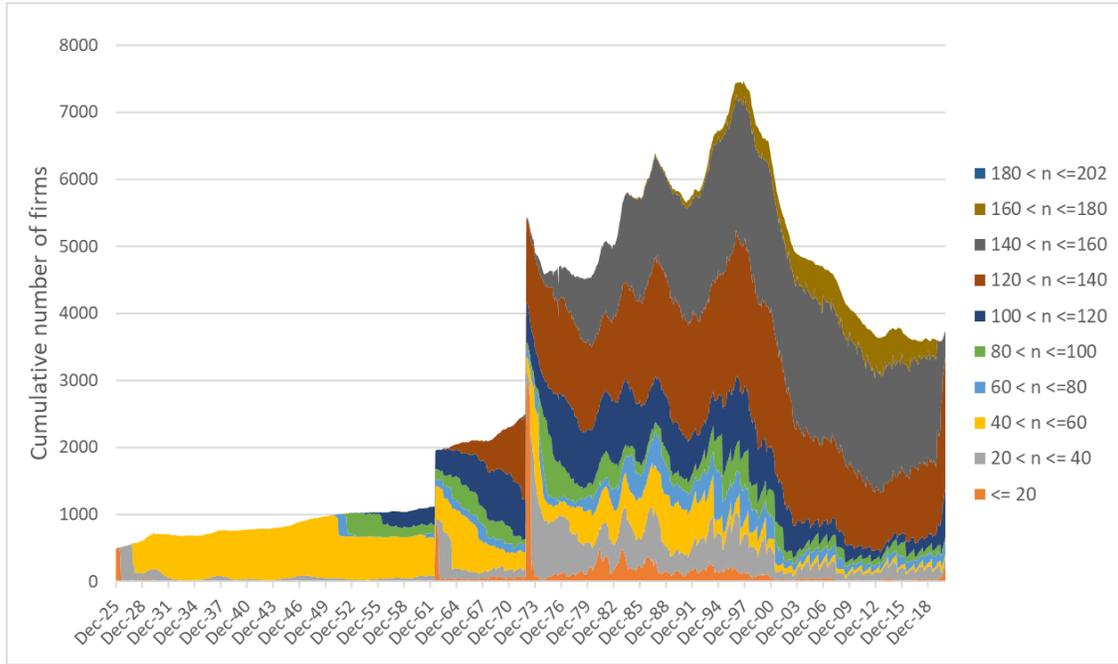
Appendix Figure A2. Effect of principal components’ estimation and forecasting horizons on out-of-sample Sharpe ratios – alternative factor data.

This figure replicates Figure 10 in the main text but uses the Chen and Zimmerman factors rather than the data provided by Kozak, Nagel and Santosh (2018). This figure shows the average Sharpe ratios of portfolios formed from increasing numbers of principal components for different estimation windows. The data sample consists of the short and long legs of 15 anomalies from 1972 to 2015 as provided by Chen and Zimmerman (2021). The 15 anomalies are used in Kozak, Nagel and Santosh (2018). Panel A shows the averages of rolling in-sample Sharpe ratios obtained from forming optimal portfolios consisting of different numbers of principal components. Panel B shows the averages of rolling out-of-sample Sharpe ratios. Panel C shows the averages of rolling out-of-sample Sharpe ratios for the 25 size and book-to-market portfolios. Panel D shows the averages of the corresponding out-of-sample Sharpe ratios for the same assets. The principal components are computed on a rolling monthly basis for different estimation windows. The out-of-sample portfolios are constructed using the in-sample optimal weights. The solid red line shows the Sharpe ratios where the sample is split in half and replicates Kozak, Nagel and Santosh (2018). The other lines present in-sample and out-of-sample windows from $t - k$ to t , where k can be 120, 60 or 36 months of daily returns.

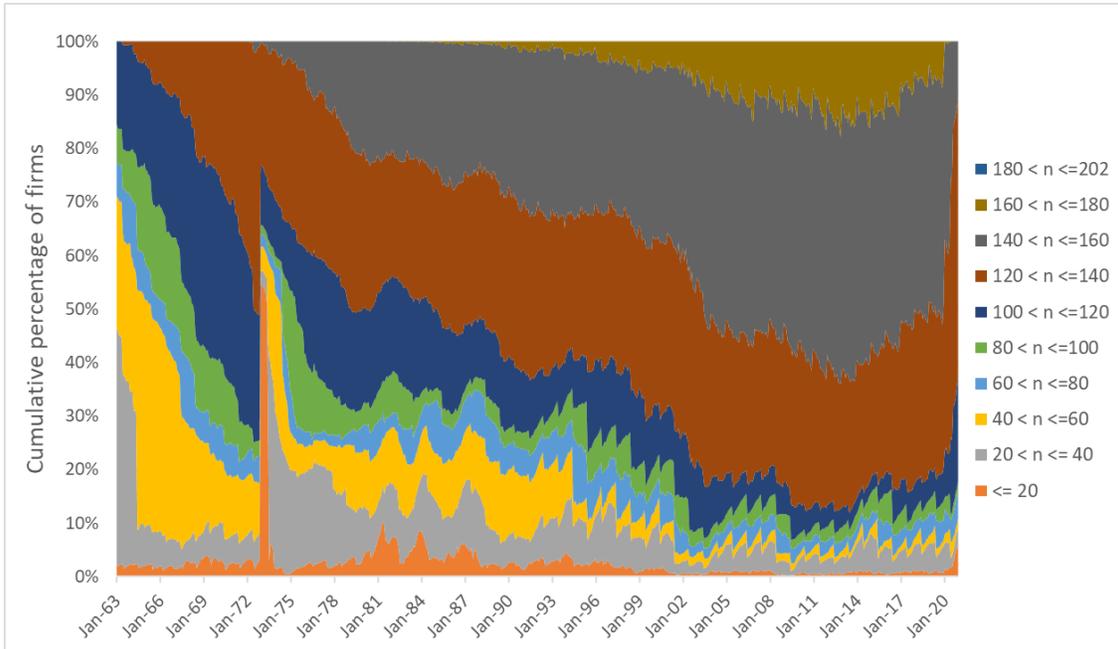


Appendix Figure A3. Number of firm-level non-missing characteristics over time. This figure shows the number of firms grouped by the total number of non-missing characteristics for each firm at each month in the sample. For each firm at each date, we compute the number of non-missing characteristics in the original sample of cross-sectional characteristics provided by Chen and Zimmerman (2021). Panel A shows the cumulative number of firms as the number of characteristics increases across groupings for the full sample from 1925-2020. Panel B shows the cumulative percentage of firms at each date that fall in each grouping for the subsample of years 1963-2020.

Panel A: Cumulative number of firms grouped by number of non-missing characteristics (1925-2020)

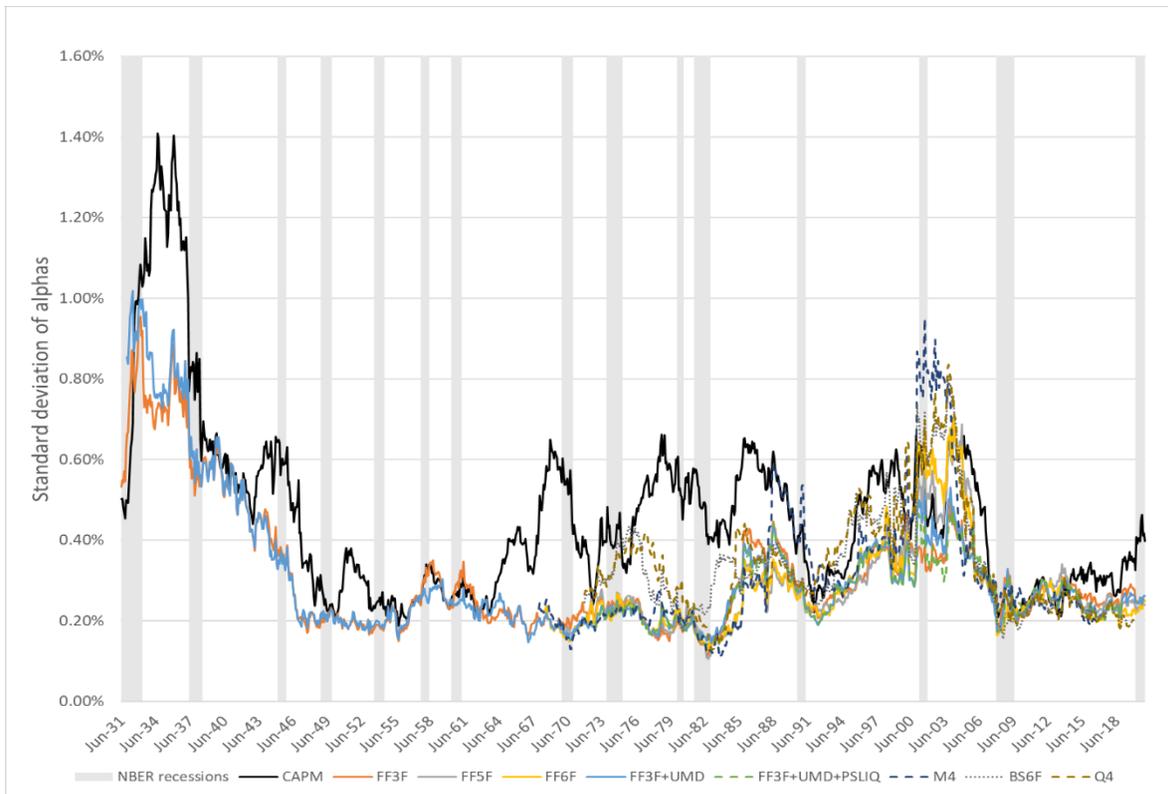


Panel B: Percentage of firms grouped by number of non-missing characteristics (1963-2020)



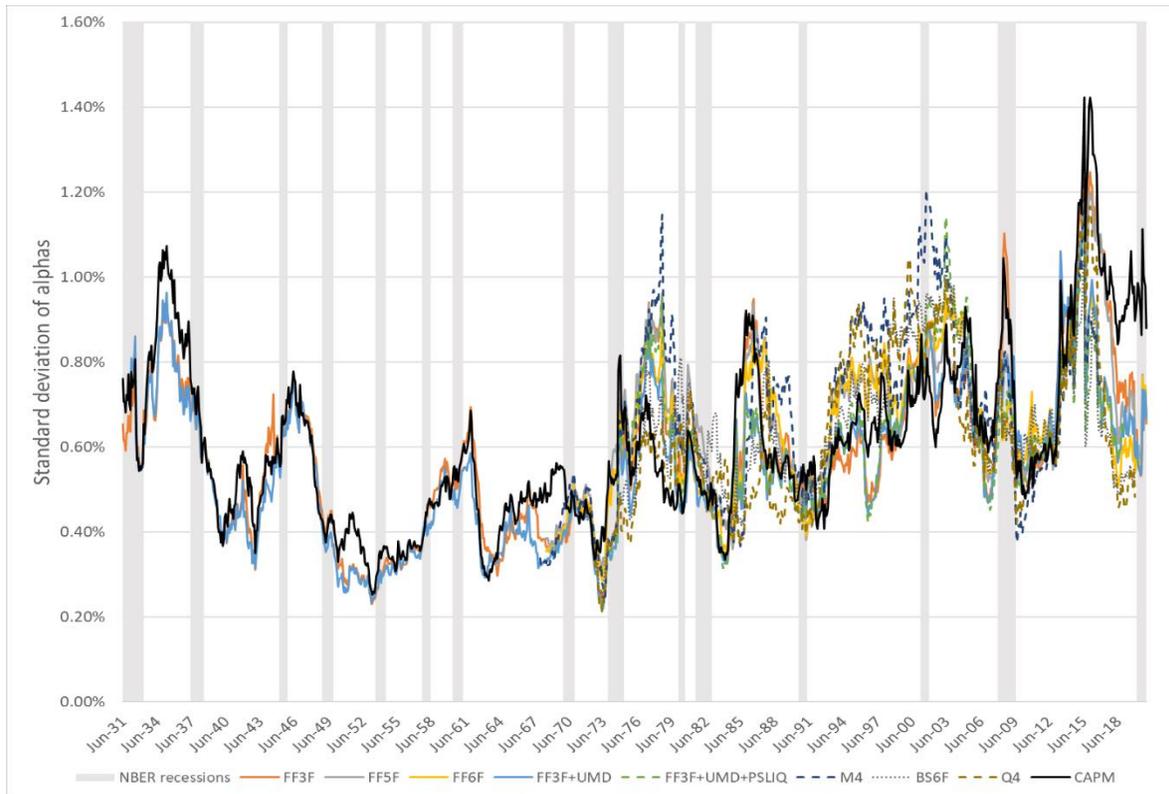
Appendix Figure A4. Time series variation in alphas of various asset pricing models This figure shows the time series variation in the standard deviation of alphas obtained from various asset pricing models of different sets of test assets. For each month t , a test asset's excess monthly returns from $t - 59$ to t are regressed on factors of various asset pricing models to obtain an alpha relative to that asset pricing model. To be included, a test asset is required to have 60 non-missing returns over the estimation period. At each month t , we plot the standard deviation of all alphas obtained from a specific asset pricing. We use four samples of test assets when computing the alphas. Figure 3 in the main text shows the analog for all stocks in the CRSP universe. Panel A below consists of the 25 size and book-to-market portfolios. Panel B consists of the Fama-French 30 industry portfolios. The asset pricing models include the Capital Asset Pricing Model (CAPM), the Fama and French (1993) 3-factor model (FF3F), the Fama and French (2015) 5-factor model (FF5F), the Fama and French (2018) 6-factor model (FF6F), the FF3F model augmented with Carhart (1997) momentum factor (FF3F+UMD), the FF3F+UMD model augmented with momentum and Pastor and Stambaugh (2003) liquidity factor (FF3F+UMD+PSLIQ), the Stambaugh and Yuan (2017) factor model (M4), the Barillas and Shanken (2018) 6-factor model (BS6F) and the Hou et al. (2015) q-factor model (Q4). The standard deviation of alphas is in percent per month. The grey vertical bars represent periods of NBER-defined recessions.

Panel A: 25 size and book-to-market portfolios

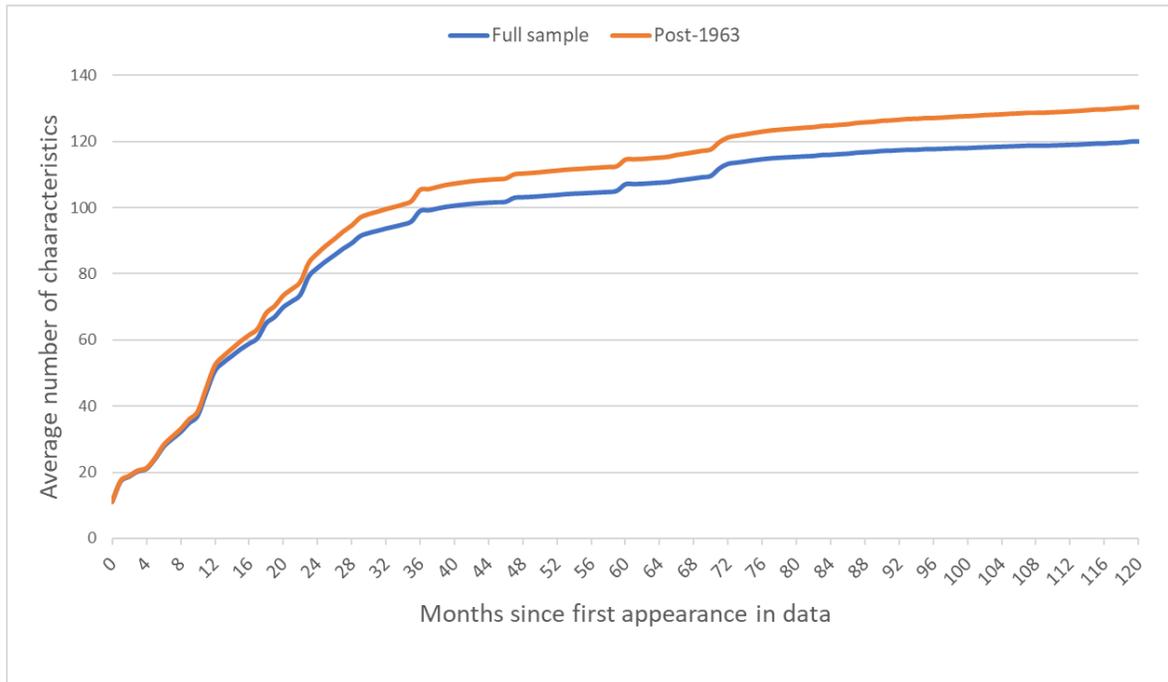


Appendix Figure A4 continued.

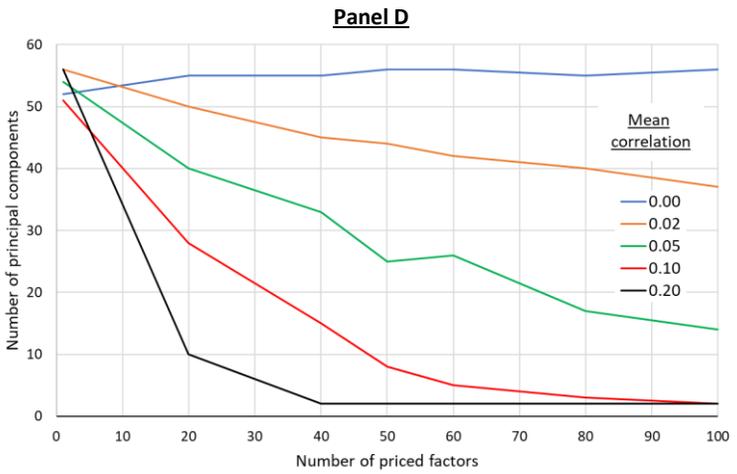
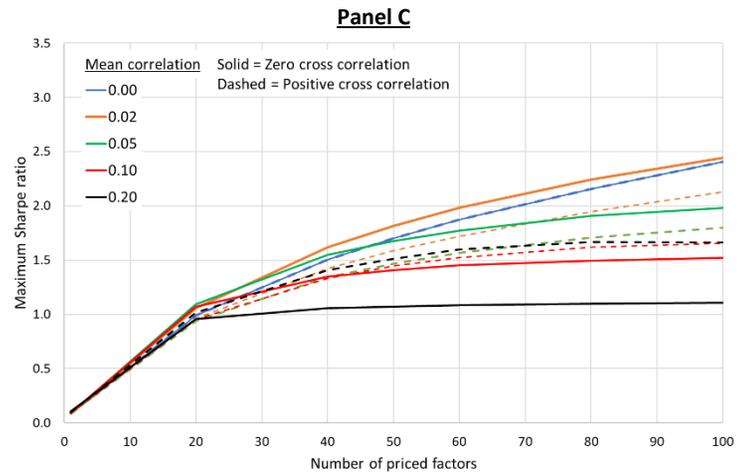
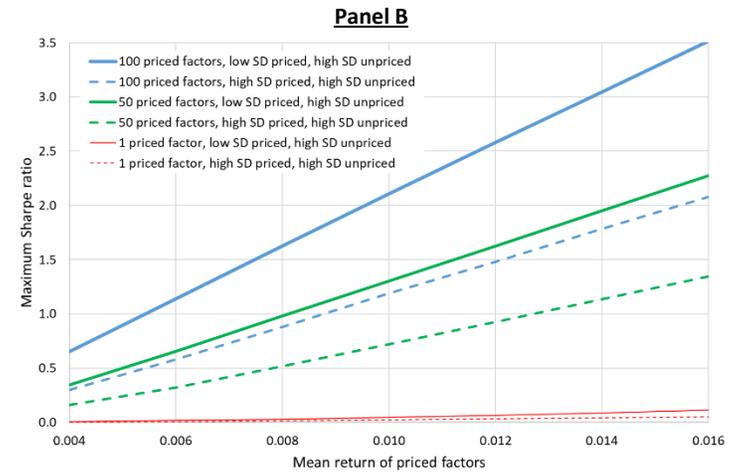
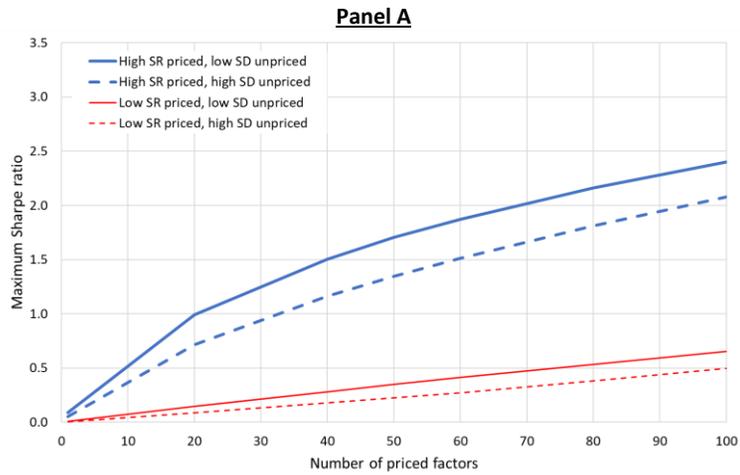
Panel B: 30 Fama-French industry portfolios



Appendix Figure A5. Cumulative number of non-missing cross-sectional characteristics over a firm's lifecycle. This figure shows the average number of cross-sectional characteristics available for each firm in a given month since the firm first appears in the cross-sectional characteristics dataset of Chen and Zimmerman (2021). The first month the firm appears is indexed at zero. The blue line is for firms that first appeared at any time during the sample. The orange line is the set of firms that first appeared after January 1963.



Appendix Figure A6 – Simulation of Sharpe ratios obtained from portfolios of out-of-sample principal components. These figures show the maximum Sharpe ratios of portfolios consisting of out-of-sample principal components constructed using simulated data. The simulation assumes 205 factors. Panel A shows how the maximum Sharpe ratios vary with the number of priced factors under parameters obtained from the factor returns during periods of high and low Sharpe ratios, and high and low standard deviations of unpriced factors. Panel B shows how the maximum Sharpe ratios vary with the mean return of the priced factors, while also varying the number of priced factors and the standard deviations of the priced and unpriced factors. Panel C shows how the Sharpe ratios vary with the number of priced factors and different correlations of the factors. The solid lines in Panel C represent the case when zero cross-correlation exists between the priced and unpriced factors, while the dashed lines have those correlations set to the same mean as the correlations. Panel D shows the number of principal components in the portfolio that yields the maximum Sharpe ratio for the case of zero cross-correlations between priced and unpriced factors.



Appendix Table A1. Variable definitions This table summarizes the various variables we use throughout the analysis. The variables are listed in order of appearance in the paper.

<u>Variable name</u>	<u>Description</u>
Number of significant factors	The total number of significant factors at each month t . At each month t , we regress each factor's returns from $t-60$ to $t-1$ on the market's excess returns over the same period to obtain the factor's CAPM alpha. A factor is considered significant if the t -statistic of its CAPM alpha is greater than 3.00. To be included, the factor must have zero non-missing returns over the 60-month period. Factors come from Chen and Zimmerman (2021) and are categorized as "clear" or "likely" predictors.
Standard deviation of stocks' alphas	The equal-weighted (value-weighted) cross-sectional standard deviation of stocks' alphas at month t . At each month t , we regress each factor's returns from $t-60$ to $t-1$ on the market's excess returns to obtain the stock's CAPM alpha. To be included, the stock must have zero non-missing returns over the 60-month period. Stocks are all common stocks (CRSP share codes 10 or 11) listed on the NYSE, AMEX and NASDAQ. The value-weighted cross-sectional standard deviation is weighted by each stock's market capitalization at $t-61$. Units are expressed as percentage points. This measure is standardized across the sample for ease of interpretability.
NBER recession	A recession indicator equal to 1 if the economy at month t was in a recession as defined by the National Bureau of Economic Research. Data can be obtained here: https://fred.stlouisfed.org/series/USREC
25 Size and Book-to-Market portfolios	Monthly equal-weighted returns from 25 size and book-to-market portfolios provided on Ken French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
30 Fama-French industry portfolios	Monthly equal-weighted returns from Fama-French 30 industry portfolios provided on Ken French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
Number of significant principal components	The number of significant components at month t required to explain 95% of the variation in factor returns from $t-60$ to $t-1$. The set of factors may be either all factors or only significant factors during the time period. Returns may be at either the monthly or daily frequency.
Significance spell of factor	The number of consecutive months for which a factor remains significant.

Mean standard error of alphas	The equal-weighted average standard error of CAPM alphas for all stocks at month t . The value-weighted mean standard error is weighted by each stock's market capitalization at $t-61$. This measure is standardized across the sample for ease of interpretability.
Mean residual volatility	The average residual volatility across stocks obtained from the 60-month stock-level CAPM regressions at each month t . This measure is standardized across the sample for ease of interpretability.
Number of public firms	The total number of CRSP common stocks listed on the NYSE, AMEX or NASDAQ at time t . This measure is standard across the sample for ease of interpretability.
Number of public firms at beginning of estimation period	The total number of CRSP common stocks listed on the NYSE, AMEX or NASDAQ at time $t-60$. This measure is standard across the sample for ease of interpretability.
Number of public firms with alpha at t	The total number of CRSP common stocks at month t listed on the NYSE, AMEX or NASDAQ for which an alpha can be calculated (i.e., stock has zero non-missing returns from $t-60$ to $t-1$). This measure is standard across the sample for ease of interpretability.
Mean absolute alpha	The equal-weighted average of the absolute value of alpha for all factors at month t . This measure is standardized across the sample for ease of interpretability.
Mean standard error of alphas	The equal-weighted average standard error of alphas for all factors at month t . This measure is standardized across the sample for ease of interpretability.
Unemployment rate	The percentage of the labor force unemployed at t as determined by the US Bureau of Labor Statistics. This measure is standardized across the sample for ease of interpretability. Data can be obtained here: https://fred.stlouisfed.org/series/UNRATE
Fed funds rate	The federal funds rate at the end of each month t . This measure is standardized across the sample for ease of interpretability. Data can be obtained here: https://fred.stlouisfed.org/series/FEDFUNDS

10-year Treasury Bond yield	The 10-year Treasury bond yield at the end of each month t . This measure is standardized across the sample for ease of interpretability. Data can be obtained here: https://fred.stlouisfed.org/series/DGS10
% of dividend-paying firms	The total number of common stocks which pay a dividend divided by the total number of common stocks at each month t . A stock is defined as paying a dividend if at least one dividend was paid over the previous year. This measure is standardized across the sample for ease of interpretability.
Mean institutional ownership	The average institutional ownership across stocks at month t . For each stock at month t , the percentage of institutional ownership is determined by the total number of shares held by institutions divided by the total number of shares outstanding. Institutional shareholdings are obtained from Thomson-Reuters 13-F database. This measure is standardized across the sample for ease of interpretability.
Economic complexity index	An annualized measure of economic complexity based on the complexity of trade activities within the United States. Each month t uses the measure from December of the most previous year. This measure is standardized across the sample for ease of interpretability. Data can be obtained here: https://oec.world/en/rankings/legacy/eci
Mean Amihud Illiquidity	For each stock in each month, we compute the Amihud (2002) illiquidity measure using daily data. We require at least 10 trading days in a month. We then average this measure across all stocks in that month.
Diversity of firm characteristics	We compute the standard deviation for each of the 205 cross-sectional characteristics across firms in each month t . We then standardize each of these characteristic standard deviations based on the entire time series for that characteristic. We then sum all the standardized characteristics available at each month. Finally, we move the measure 36 months back in time to account for the delayed introduction of characteristics during the first 3 years from which a firm first appears in the data. The final measure used in the regression is standardized across the sample for ease of interpretability.

Appendix Table A2. Comovement of the number of significant factors with the number of public firms This table shows the results of regressing the number of significant factors in each period on the number of public firms at each date. For each month t , we regress each factor's monthly returns from $t - 59$ to t on the market's monthly excess returns to obtain each factor's CAPM alpha. A factor is significant at month t if the t -statistic of its CAPM alpha is greater than 3.00. The dependent variable is a count of the number of significant factors at each month t . The number of public firms consists of all common stocks trading on the NYSE, NASDAQ or Amex and is computed in three ways: 1) all public firms at month t , 2) all public firms at month $t - 59$, and 3) all public firms with a non-missing alpha over the period $t - 59$ to t . To be included, stocks and factors must have 60 non-missing returns over the alpha estimation period. Panel A covers the sample of factors from 1931-2020. Panel B covers the sample of factors from 1968-2020. We standardize each independent variable by subtracting the mean of that variable over the time series and dividing that difference by the variable's standard deviation over the time series. Hansen-Hodrick standard errors are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

	Dep var: Number of significant factors							
	1931-2020				1968-2020			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total number of firms:								
At time t	18.24*** (3.08)			29.89** (13.76)	26.74*** (7.97)			32.46** (12.99)
At time t-60		15.42*** (3.92)		5.98 (9.28)		13.86*** (4.88)		2.85 (9.51)
with 60-month alpha			15.69*** (3.70)	-18.23 (11.37)			17.01*** (5.67)	-11.10 (11.58)
R-squared	0.63	0.45	0.47	0.67	0.50	0.17	0.19	0.52
N	1075	1075	1075	1075	636	636	636	636

Appendix Table A3. Disentangling noise from power This table shows the results of regressing the number of significant factors in each period on the cross-sectional mean of the absolute value of all factors' CAPM alphas, the mean standard error of those alphas and the number of public firms. For each month t , we regress each factor's monthly returns from $t - 59$ to t on the market's monthly excess returns to obtain the factor's CAPM alpha and its corresponding standard error. The dependent variable is a count of the number of significant factors at each month t . A factor is significant at month t if the t -statistic of its CAPM alpha is greater than 3.00. To be included, factors must have 60 non-missing returns over the alpha estimation period. The number of public firms is a count of all common stocks outstanding at t . We standardize each independent variable by subtracting the mean of that variable over the full time series and dividing that difference by the variable's standard deviation over the time series. Hansen-Hodrick standard errors with a bandwidth of 60 are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

	Dependent var: Number of significant factors							
	1931-2020				1968-2020			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean standard error of factor alphas	-8.69** (3.48)	-2.69 (1.98)	-30.01*** (6.98)	-19.39*** (6.72)	-17.01 (13.42)	-17.61** (8.22)	-37.64*** (4.03)	-34.61*** (4.13)
Number of public firms		17.31*** (2.90)		11.16*** (1.67)		26.91*** (7.26)		7.97*** (1.67)
Mean factor absolute alpha			27.08*** (7.75)	18.51*** (6.63)			37.26*** (3.80)	31.47*** (4.35)
R-squared	0.14	0.64	0.67	0.83	0.09	0.60	0.90	0.92
N	1075	1075	1075	1075	636	636	636	636

Appendix Table A4. Comovement of the number of significant factors and economy and firm characteristics This table shows the results of regressing the number of significant factors in each period on various economic measures at each month. For each month t , we regress each factor's monthly returns from $t - 59$ to t on the market's monthly excess returns to obtain the factor's CAPM alpha and its corresponding standard error. A factor is significant at month t if the t -statistic of its CAPM alpha is greater than 3.00. The dependent variable is a count of the number of significant factors at each month t . The number of public firms is a count of all common stocks at t traded on the NYSE, NASDAQ or Amex at month t . The NBER recession indicator is an indicator equal to one if the month is classified as an NBER recession and zero otherwise. The unemployment rate is the number of unemployed as a percentage of the labor force as provided by the U.S. bureau of labor statistics. The federal funds rate is the established rate by the Federal Reserve at month t . The 10-year treasury bond yield is the market yield on U.S. treasury securities at a 10-year constant maturity. The percent of dividend-paying firms is the total number of common stocks which have paid a dividend in the previous 12 months divided by the number of firms at month t . The mean institutional ownership is the fraction of a firm's shares outstanding held by 13-f firms. The economic complexity index is a measure of economic complexity used from Simoes and Hidalgo (2011). Diversity of firm characteristics is a measure of diversity in the cross-sectional characteristics across firms. See Appendix Table A1 for a complete description of the measures. The sample covers factors from 1931-2020. Hansen-Hodrick standard errors are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

Sample of factors from 1931-2020

	Dep var: Number of significant factors												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Number of public firms	18.24*** (3.08)										5.85 (6.41)	5.40 (4.73)	-4.97 (19.75)
NBER recession indicator		-9.41** (4.66)									-9.90** (4.17)	-8.92** (4.10)	-2.27 (3.70)
Unemployment rate			0.11 (4.38)								-11.16*** (1.77)	-9.76*** (2.33)	-9.81*** (1.73)
Fed funds rate				8.95* (5.18)							-5.62** (2.46)	-4.20* (2.18)	-2.90 (4.63)
10-year T-Bond yield					9.64* (5.23)						15.46*** (4.48)	15.04*** (3.77)	1.83 (5.94)
% dividend-paying firms						-12.12*** (4.47)					-1.32 (8.98)	-2.65 (7.94)	-16.19 (17.84)
Mean institutional ownership							-16.81*** (4.46)						-20.14 (13.18)
Economic complexity index								5.59 (4.72)				1.83 (2.30)	-3.91 (2.54)
Mean Amihud illiquidity									-5.30*** (1.90)		12.97*** (4.33)	11.84*** (4.22)	0.70 (4.55)
Diversity of firm characteristics										14.76*** (4.76)	9.32* (5.29)	10.13** (4.97)	18.51** (7.46)
R-squared	0.63	0.02	0.00	0.16	0.18	0.28	0.46	0.06	0.05	0.42	0.67	0.70	0.80
N	1075	1075	876	798	708	1075	483	648	1075	1039	672	648	456

Appendix Table A5. Simulation calibrations This table provides initial calibrations for the results found in Figure A1 Panel A.

Scenario	μ_p	σ_p	σ_u	ρ_{pu}
Baseline: High SR priced, low unpriced SD	0.016	0.032	0.025	0
High SR priced, high unpriced SD	0.016	0.032	0.036	0
Low SR priced, low unpriced SD	0.004	0.018	0.025	0
Low SR priced, high unpriced SD	0.004	0.018	0.036	0