More factors matter and factors matter more than you might think: The role of time variation in factor premia*

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

January 2024

Abstract

The literature has asserted that as few as four or five factor principal components (PCs) are sufficient to largely explain the cross-section of stock returns. By allowing for time variation in factor premia, we show that portfolios formed from factor PCs yield economically large out-of-sample Sharpe ratios that increase as up to forty PCs are employed. That is, non-latent time-varying factors have strong predictive power for the cross-section of stock returns, and to a substantial extent are not redundant of each other. Time variation in the number of economically relevant factors is related to changes in economic conditions and the diversity of firm characteristics, indicating roles for economic complexity and investor learning.

*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, Shrihari Santosh, and seminar participants at the 2023 American Finance Association Meetings, the 2023 Finance Down Under conference, the 2022 Paris Financial Management Conference, the University of Washington and the University of Oklahoma. Earlier drafts of this paper were titled "Time Variation in the Factor Zoo." Author emails are hb@asu.edu, aaronburt@ou.edu, and hrdlicka@uw.edu.

1. Introduction

The literature has identified hundreds of variables, including firm characteristics and "factors" constructed as returns to long-short portfolios, that appear to have significant ability to predict the cross-section of stock returns.[1] However, more recent research has asserted significant progress in taming the "factor zoo."[2] Hou, Xue, and Zhang (2020) report that the first four principal components (PCs) of 158 characteristics they study explain over half the sample variation, and that eight PCs can explain more than two thirds of the variation. Kelly, Pruitt, and Su (2019) use the method of instrumented principal components analysis (ICPA), and report that five ICPA factors explain the cross-section of returns better than existing factor models. Kozak, Nagel, and Santosh (KNS, 2018) study Sharpe ratios (SRs) for portfolios formed from factor PCs, and report that "while factors beyond the first few PCs contribute substantially to the maximum SR in sample, PCs beyond the first few no longer add to the SR out of sample." Collectively, this evidence suggests that the many characteristics and characteristic-based factors identified in the literature contain a substantial amount of redundant information, and that only a few sources of independent variation are required to explain the cross section of stock returns.

We show that allowing for time variation in factor return premia leads to notably different implications. In particular, we rely on rolling sixty-month in-sample estimation periods and show that out-of-sample (OOS) Sharpe ratios for portfolios formed from factor principal components are economically large (averaging over 2.6 on an annualized basis) and increase as portfolios are formed from as many as forty factor principal components. These conclusions hold in the data studied by KNS (2018) and are even stronger in a broader set of 205 factors compiled by Chen and Zimmerman (2021). Our results imply that empirically observable, non-latent factors contain substantial information relevant to

[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 method of factor analysis.

[2] Cochrane (2011) appears to have been the first to use the phrase "factor zoo" to refer to the body of evidence.

forecasting the cross-section of stock returns. Further, the fact that OOS Sharpe ratios continue to increase as portfolios are formed from larger numbers of PCs implies that the factors identified in the literature are, to a substantial extent, not redundant in linear combination.

The notion that only a few sources of common variation should be necessary to explain the cross section of returns is articulated by KNS (2018, page 1184): "the existence of a relatively small number of arbitrageurs should be sufficient to ensure that near-arbitrage opportunities—that is, trading strategies that earn extremely high Sharpe ratios (SRs)—do not exist." KNS (2018) further argue that the stochastic discount factor "can be represented to a good approximation" with a "few dominant factors." Consistent with this perspective, Kozak, Nagel, and Santosh (2020) employ an estimation procedure that penalizes deviations of Sharpe ratio estimates from zero.

Yet, we document high OOS Sharpe ratios, even in the same sample studied by KNS. These results do not necessarily imply unexploited arbitrage opportunities, but rather suggest that arbitrage activity faces larger barriers than may have been previously recognized. Time variation in factor premia implies that frequent trading would be required in any attempt to exploit the high Sharpe ratios, with attendant high transaction costs attributable in part to the fact that many factors involve positions in small and illiquid stocks. We also document that the return premia associated with individual factors display "spells" of significance and insignificance, to an extent that cannot be attributed to random variation around a constant premium. We further show that the numbers of factors that are economically relevant at various points in time are related to diversity in firm characteristics and to measures of economic complexity. In combination, these results suggest that the need for investors to learn or estimate the current state of time-varying factor premia comprises an additional barrier to arbitrage, and that this barrier is greater at times of greater economic complexity. In short, it is a challenge for active investors to identify the relevant factor premia in real time.

We believe that our findings are of inherent interest, independent of whether the evidence implies the existence of unexploited trading opportunities. For example, the evidence implies time variation in firms' costs of capital, with attendant implications for corporate finance decisions such as the timing of security issuances. Further, while parsimony is surely desirable, theoreticians should take account of the evidence indicating that the determinants of expected stock returns are more complex than many models would imply. The recent demand-based asset pricing literature (e.g., Koijen and Yogo, 2019; Gabaix and Koijen, 2021), which is premised on characteristic-based heterogeneity in asset demand across investors and implies that expected returns depend in part on demand shocks that originate with heterogeneous groups of investors comprises a promising example of an approach with potential to enhance understanding of an economy where many characteristic-based factors may be relevant at various times.

Accommodating time variation in factor premia may provide additional scope for specification searches or other sources of bias. We avoid amplifying any such bias in this study by focusing only on previously identified factors. Further, a requirement to link variation in estimated factor premia to variation in observable economic variables that were not directly employed to identify the factors imposes a degree of discipline. Here, we show that that the number of economically relevant factors is related to measures of the economic environment, including a recession indicator variable, interest rates, the percentage of firms that pay dividends, mean institutional ownership rates, and an economic complexity index. On balance, the results presented here indicate a significant role for time-varying economic complexity in asset pricing.

2. Placing our Study in the Literature

We are certainly not the first to consider the role of time variation in factor premia. However, the focus of most prior studies has mainly been on canonical factors such as the overall market, firm size or value (market-to-book ratio).[3] Haddad, Kozak, and Santosh (2020) expand on the prior research by showing that time variation in the return premia in the PCs of fifty “anomaly” portfolios can be identified based on the market-to-book ratios of the factors themselves. We go further by accommodating time variation in the set of empirically relevant factors rather than focusing on a constant set of PCs formed

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

from a fixed set of factors, and by relaxing the restriction that factor premia are functions of market-to-book ratios alone. We further demonstrate the empirical importance of doing so.

The evidence that time variation in factor premia is economically important also has implications for the literature on investor learning. Lewellen and Shanken (2002) note that rational investors' uncertainty regarding parameters such as the dividend growth rate allow for return predictability to arise in equilibrium. Martin and Nagel (2022) show how investors' incorrect priors regarding parameters of the return distribution in combination with complexity in the form of many 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 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 learning as a barrier to arbitrage, thereby providing scope for economically large factor premia that can persist for a period of time. On balance, our results highlight the desirability of additional research that considers investor learning in a dynamic environment.

Our findings that time variation in factor premia is economically important highlights the relevance of the considerable body of evidence regarding diversity across investors and firms. 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 each stock, can change over time.[4] Further, some investors hold positions for extended 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, or in response to rumors. 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,

[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).

Kovrijnykh, Li and Pavlova, 2021) or as a function of career horizons. The trades of professional investors can depend on considerations such as the funding liquidity of their employing firms, and return premia have also been shown to also depend also on the leverage of financial sector firms.[5] Consistent with the reasoning that return premia can depend on the identity of the marginal investor, 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. In addition, the economic characteristics of newly listed firms often differs 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.

Our finding of economically large OOS Sharpe ratios also relates to a substantial literature. The implementation costs associated with any attempt to capture these time-varying return premia through active trading could be substantial, particularly since the CZ factors (which use the weighting method of the original paper for each) are often constructed with equal weights in each component stock. Equal weighting implies the need for substantive positions even in small stocks and for monthly rebalancing trades. Assessing whether these results imply profitable trading opportunities after implementation costs could be of substantial interest. Such a careful investigation might support the conclusion that, in line with the conclusions drawn by Lewellen (2011), no excess returns can be captured. However, we believe it is important to understand the nature of cross-sectional variation in expected returns and time series variation therein, whether such trading opportunities exist or not. As one example, researchers are interested in knowing if illiquidity affects expected returns as implied, for example, by Amihud and Mendelson (1982), even if that illiquidity implies the absence of profit opportunities to an active trader.

Though we refer throughout this paper to time variation in factor premia, the economic relevance of factors can vary due to changes in either factors' return premium per unit of factor risk or in the

[5] See, for example, He, Kelly and Manela (2017), Tobias, Etula, and Muir (2014) and He and Krishnamurthy (2013).

quantity of factor risk, and we do not seek to decompose the factor premia into these components.[6] We study *t*-statistics estimated for factor alphas (which are proportional to factor-level information ratios), which capture both effects. The OOS Sharpe ratios we study capture both aspects as well. Further, we do not take a stand as to whether the return premia arise because of investor aversion to undesirable factor outcomes, mispricing in the face of barriers to arbitrage, or a combination thereof.

3. Time Variation in Factor Premia and the Cross Section of Stock Returns

In this paper we study 205 factors drawn from Chen and Zimmerman (2021). However, before reporting results based on this broad sample, we demonstrate that time variation in factor premia is empirically important even in narrower and arguably more familiar samples, including those employed by Kozak, Nagel and Santosh (2018).

a. Initial Evidence: Allowing for time variation in the KNS (2018) sample

Kozak, Nagel, and Santosh (2018) (KNS) comprise a prominent example of recent studies that conclude there are only a few sources of priced variation in the cross section of stock returns. They study a sample comprised of the long and short legs of fifteen anomaly-based factors, as well as a sample comprised of twenty-five size and book-to-market based portfolios. KNS employ a split sample approach whereby the first twenty-five years of their 1965 to 2015 sample period is used for in-sample estimation of factor principal components, and the second twenty-five years comprises the out-of-sample period. They report that as few as four factor principal components (PCs) capture the relevant information, as out-of-sample (OOS) Sharpe ratios do not meaningfully increase when portfolios are formed from more than four factor PCs.

The split-sample approach employed by KNS (2018) does not allow for potential time variation in factor premia within the twenty-five-year subperiods. We study the same data as KNS, while allowing for time variation in parameters. We study factor PCs in part to ensure that our outcomes differ from

[6] That return premia can vary over time relative to canonical factors' volatility is documented for example in Moreira and Muir (2017). We extend this insight to the hundreds of non-canonical factors.

prior studies primarily (or in the case of the KNS data entirely) because of our allowance for time variation in factor return premia. It should be recognized that PCs comprise a means, not an end, and that there are other statistical methods of extracting common variation. The assessment of OOS Sharpe ratios based on factor PCs allows the assessment of whether there is redundancy in the form of linear factor combinations.

Panels A and B of Figure 1 report in-sample and out-of-sample Sharpe ratios for portfolios constructed from the PCs of the long and short legs of fifteen anomaly-based factors, while Panels C and D report corresponding outcomes for portfolios formed from the PCs of twenty-five factors constructed based on firm size and book-to-market equity ratios. The red line displays outcomes obtained using the KNS split-sample approach. The red line replicates their results, showing that factors beyond the first few PCs contribute substantially to in-sample Sharpe ratios, but PCs beyond the first four do not substantially enhance OOS Sharpe ratios. KNS argue that this result is to be expected since even a relatively small number of arbitrageurs should "be sufficient to ensure that near-arbitrage opportunities—that is, trading strategies that earn extremely high Sharpe ratios do not exist."

We compile results that parallel those reported by KNS, while allowing for time variation in factor return premia by use of rolling 60-month estimation windows. For each month, t, we employ factor returns in months $t - 59$ to t to 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. We then consider out-of-sample outcomes based on returns in months $t+1$ to $t+36$. In each case, portfolio weights are chosen to optimize the portfolio's in-sample Sharpe ratio, i.e., to identify the tangency portfolio, while performance evaluation is based on OOS Sharpe ratios for portfolios with the same weights. Specifically, we multiply out-of-sample factor returns by the in-sample eigenvectors to create returns to OOS PCs [7]. We provide a detailed outline of the computation of the PCs and the construction of the portfolios in Appendix A.

[7] To maintain compatibility, we measure out-of-sample results from within the same 25 years that comprise the out-of-sample period for Kozak, Nagel and Santosh (2018). That is, when employing N months to define the in-

The blue lines on each Panel of Figure 1 display average (across months) Sharpe ratios for portfolios formed from varying numbers of the resulting factor PCs. Two points are noteworthy. First, Sharpe ratios, both in- and out-of-sample, are virtually all larger when time variation in factor premia is accommodated than when it is not. The divergence in Sharpe ratios is apparent in each Panel of Figure 1, but is greater for the factors formed based on size and book-to-market. Second, Sharpe ratios increase as more factor PCs are employed. As KNS note, this is a mechanical outcome for in-sample Sharpe ratios. However, it is not a mechanical result out-of-sample, and OOS Sharpe ratios increase notably as the number of PCs is increased when time variation is allowed for. That is, higher order factor PCs contain economically relevant information when time variation is allowed for. In contrast, but consistent with KNS, there is little or no increase in OOS Sharpe ratios beyond the first few PCs when split-sample estimation is employed.

The increases in OOS Sharpe ratios due to the allowance for time variation are large. In particular, focusing on portfolios constructed based on the widely-studied firm size and book-to-market characteristics, OOS Sharpe ratios formed from twenty to twenty-five factor principal components are more than twice as large when time variation is allowed for as compared to when it is not (Panel D of Figure 1). These empirical results, obtained in the same sample employed by KNS, support the conclusions that (1) allowing for time variation in factor premia enhances the ability to forecast the cross section of stock returns, and (2) more factor PCs are relevant when time variation is allowed for than when it is not.

4. Broader Evidence: 205 Factors from Chen and Zimmerman (2021)

KNS studied fifteen "anomaly" factors and twenty-five size and market-to-book factors for which return data was available during all months of their fifty-year sample period. Of course, the literature studies a much broader set of factors, some of which have available data for longer time intervals than

---

sample period, we begin the estimation with the final N months of the first half of the sample, so that the first out-of-sample month is identified starting at the sample midpoint. 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.

others. Our rolling estimation approach only requires factor data that factor return data is available during relatively short estimation windows, and can therefore be implemented for factors whose data is unavailable during periods outside the estimation window.

We implement our rolling estimation approach using 205 factors derived from firm characteristics previously studied in the literature, including the 161 "clear predictors" and 44 "likely predictors" identified by Chen and Zimmerman (2021).[8] Of course, some of these factors are similar in their construction, and the economic information contained in outcomes on similarly constructed factors could overlap substantially.

We report in Panel A of Figure 2 annualized OOS Sharpe ratios for portfolios formed from the PCs of the 205 CZ factors, when the end of the sixty-month estimation period ranges from June 1931 to December 2020. The yellow line displays Sharpe ratios for portfolios formed from the first five PCs (in line with the results of Hou, Xue, and Zhang, 2020 and Kozak, Nagel and Santosh, 2018) and the blue line shows the average Sharpe ratio obtained across all numbers of PCs considered (from one to fifty-nine). The orange line shows the Sharpe ratio obtained from 59 PCs, the maximum number that can be constructed when using a sixty-month estimation window. Table 1 reports time series averages of each of these Sharpe ratios, as well as differences in average Sharpe ratios and associated t-statistics.[9]

The data displayed on Panel A of Figure 2 and the summary statistics in Table 1 indicate that the CZ factors collectively have economically important out-of-sample predictive power for the cross-section of stock returns. The annualized OOS Sharpe ratios for portfolios formed from the first five factor PCs are mostly positive, average 1.4 for the full sample, and exceed 3.0 during some portions of the 1980s and 1990s. The blue line displays average Sharpe ratios (across one to fifty-nine PCs) and comprises our

[8] The authors graciously posted their data to https://www.openassetpricing.com/. Bessembinder, Burt and Hrdlicka (2023) report qualitatively and quantitatively similar results when using the factor data of Jensen, Kelly, and Pedersen (2021) instead. 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 (2023) report similar outcomes for a variety of in-sample and out-of-sample window lengths.

[9] Since outcomes are based on rolling sixty-month estimation, we employ Hansen-Hodrick standard errors with a bandwidth of sixty.

main focus. These Sharpe ratios are virtually always positive, average 2.64 across months, and particularly since about 1950 are consistently larger than those based on five PCs. The Sharpe ratio based on fifty-nine factor PCs is larger yet, averaging 2.98 across months. These results reaffirm that observable, i.e., non-latent, factors have significant forecast power for the cross-section of returns. The Sharpe ratios displayed on Figure 2 are generally larger than those displayed on Figure 1, implying that the larger set of CZ factors contain more information than the factors studied by KNS.

The difference between the time series average OOS Sharpe ratio obtained from the average across one to fifty-nine PCs and the time series average based on five PCs is 1.25, while the difference in the time series average Sharpe ratio based on the maximum number of PCs and that based on five PCs is 1.58. Each of these differences is highly statistically significant, with t-statistics of 7.22 and 3.55, respectively. These OOS Sharpe ratios imply that factor PCs beyond the first few contribute substantially to the ability to forecast the cross-section of stock returns.[10]

BNS note that in-sample Sharpe ratios increase mechanically when portfolios are formed from a larger number of factor PCs. This mechanical result should not be expected to, and does not, pertain to out-of-sample outcomes. To illustrate this point, we also report on Table 1 the time series average of the maximum Sharpe ratio obtained across any number of PCs. This maximum Sharpe ratio exceeds the Sharpe ratio obtained with the maximum number of PCs in every rolling sixty-month interval, and averages 4.24 across months. The average number of PCs for the portfolio that gives the maximum out-of-sample Sharpe ratio is forty-seven. Alternatively stated, the higher OOS Sharpe ratios obtained when portfolios are formed from larger numbers of factor PCs is an economic outcome, not a mechanical one.

The large maximum Sharpe ratios and the large numbers of PCs used to construct these portfolios imply that many factors contain economically relevant and non-redundant predictive

[10] We also assess the explanatory power of individual factor PCs, from the first to the fifty-ninth, by assessing Sharpe ratios for portfolios comprised of that PC factor alone (i.e., PC-transformed factor returns). In Appendix Figure C1, we report the average (across months) in-sample and OOS Sharpe ratios for each factor PC. Consistent with the reasoning expressed by KNS (2018), in-sample Sharpe ratios increase almost monotonically for the higher numbered PCs. In contrast, OOS Sharpe ratios are nearly constant, equal to approximately 0.20 to 0.25 across all PCs from the first to the fifty-ninth. That is, the OOS explanatory power of higher numbered PCs is neither greater or smaller than that of the first few PCs.

information regarding the cross-section of stock returns, particularly when time variation is allowed for. This outcome contrasts with the conclusions of recent papers such at Kelly, Pruitt, and Su (2019), Kozak, Nagel and Santosh (2018), and Kozak, Nagel and Santosh (2020) that conclude that only relatively few factor principal components are relevant. The different conclusion arises because our rolling estimation approach allows for flexible time variation in factor return premia in a way these studies do not.[11]

5. Are outcomes stronger when considering only those factors that are significant in sample?

For the results reported to this point, we employ all factors with available data during the relevant estimation window. While Chen and Zimmerman (2021) identified their set of factors based on the factors' statistical significance during the periods studied by authors of the original underlying studies, the evidence in support of some factors is stronger than others. Further, if factor premia vary across time, then considering only those factors with significant in-sample explanatory power should reduce noise and improve the ability to forecast the cross-section of returns. We identify the factors that have significant in-sample explanatory power by identifying those cases where the t-statistic on the alpha estimated in a regression of factor returns on excess market returns exceeds 3.0 (the level recommended by Harvey, Liu and Zhu, 2016). A significant alpha in this regression implies that a factor is economically relevant in the sense that it has explanatory power for the cross-section of non-market returns.[12]

---

[11] In Appendix B we report results obtained when we use the data and programs posted by Kozak, Nagel and Santosh (2020). OOS Sharpe ratios implied by their analysis and data continue to increase modestly as more factor PCs are used to form portfolios, even though their approach does not accommodate time variation in factor premia. Thus, while allowing for time variation is important, more than five factor PCs are of some use even without such allowance.

[12] It would of course be possible to assess the significance of a given factor by estimating alphas in regressions on multiple other factors. However, it is unclear which or how many additional factors should be employed, and with only sixty monthly observations, not all can be. We therefore choose the simple approach of assessing only if the factor explains the cross-section of non-market returns. If a given factor contains the same information as other factors, then OOS Sharpe ratios should not increase as the number of factor PCs increases.

Panel B of Figure 2 displays OOS Sharpe ratios that correspond to those in Panel A, but are obtained when PCs are formed from only those factors with significant in-sample alphas.[13] Table 1 also reports on means and standard deviations across months of these Sharpe ratios. The outcomes indicate that OOS Sharpe ratios for portfolios formed from the PCs of those factors with statistically significant in-sample alphas exceed those for portfolios formed from the PCs of all factors. In particular, the average annualized out-of-sample Sharpe ratio (across months) for portfolios formed from five PCs is 2.16 when the PCs are drawn from only significant factors versus 1.39 when the PCs are drawn from all factors. Corresponding outcomes for Sharpe ratios averaged across all potential numbers of PCs are 3.20 when PCs are constructed from only significant factors versus 2.64 when PCs are constructed from all available factors. In Section 4, we delve more deeply into the empirical determinants of times series variation in the numbers of economically relevant factors.

The annualized out-of-sample Sharpe ratios displayed in Figure 2 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. These substantial out-of-sample Sharpe ratios support the conclusion that the factors identified in the literature have economically important predictive power for the cross-section of stock returns. Importantly, the outcome that out-of-sample Sharpe ratios are larger when many factor PCs are employed would not be anticipated if researchers had systematically identified new factors that duplicated the information contained in existing factors.

6. Are the results attributable to extreme portfolio weights?

KNS (2018) express the concern that constructing mean-variance optimal portfolios from large numbers of factor PCs could result in portfolios with extreme weights. The apparent concern is that if higher order PCs did not lead to small Sharpe ratios, it would imply that the tangency portfolio would

[13] Excess market returns are obtained from Kenneth French's website. Some sections of the lines in Panel B are missing due to an insufficient number of significant factors (i.e., only 1 or 0) during those time periods.

involve unreasonably large positions. Accordingly, they assert that "ignoring the small-eigenvalue PC portfolios" is appropriate for OOS evaluation. We assess if the portfolios selected by our estimation methods, which are based on up to fifty-nine factor PCs, involve extreme weights. In particular, Figure 3 reports on the average (across months) of the maximum and the minimum weight on any PC in the mean-variance optimal portfolio, when the number of factor PCs varies from one to fifty-nine. While the exact definition of an "extreme" weight may be subjective, we note that our procedures lead to maximum weights very near 10% when the number of factor PCs exceeds thirty.

The preceding outcomes are based on the selection of mean-variance optimal portfolios, the details of which are somewhat complex. To assess whether our main conclusions regarding the role of time variation in factor premia and the number of sources of information regarding the cross-section of stock return are robust to simpler methods of estimation we compute out-of-sample Sharpe ratios for simple equal-weighted portfolios of factor PCs. Figure 4 displays monthly outcomes that parallel those in Panel A of Figure 2, when portfolios are equal-weighted and formed from the indicated number of factor principal components. Table 2 presents corresponding time series means of these OOS Sharpe ratios. These outcomes also support the conclusions that factors contain significant forecast power for the cross section of stock returns, as OOS Sharpe ratios are economically large, and are greater when portfolios are formed from more than a few factor principal components. In particular, the average OOS Sharpe ratio for equal-weighted portfolios formed from five factor PCs is 0.952, while that for equal weighted portfolios formed from all fifty-nine factor PCs is 2.957. Remarkably, the average Sharpe ratios for equal-weighted portfolios as reported on Table 2 are only modestly smaller than those for optimized portfolios as reported on Table 1. For example, considering the outcome averaged across all possible numbers of PCs, the average Sharpe ratio for optimized portfolios on Table 1 is 2.64, while that for equal-weighted portfolios on Table 2 is 2.11. Thus, our central conclusions, namely that time variation in factor premia is economically important and that factors contain more than four or five independent sources of information regarding the cross-section of stock returns, obtain even when portfolios are formed from factor PCs in the simplest possible manner.

7. Can OOS Sharpe ratios or the number of factor PCs that maximize OOS Sharpe ratios be forecast?

To this point, we have emphasized OOS Sharpe ratios that are based on portfolios formed from either the maximum possible number of factor PCs or based on average outcomes across all possible numbers of factor PCs. The number of factor PCs that lead to the highest OOS Sharpe ratio for any given thirty-six month evaluation period can be observed, ex post. We next assess whether evidence that more factors are relevant in-sample also implies higher OOS Sharpe ratios or that the highest OOS Sharpe ratios will be attributable to portfolios formed from a greater number of factor principal components.

We assess, for each rolling sixty-month in sample period, the number of factor PCs required to explain 95% of the time series variation across the available CZ factors. We term this the number of "relevant PCs" and plot the outcomes on Panel A of Figure 5 as a red line. To explain 95% of the in-sample variation in factor returns requires between twenty-nine and forty PCs for every rolling sixty-month window from the late 1950s through the end of the sample period.[14] In general more factors are relevant during those periods where more factors are available; the correlation between the number of CZ factors with available data in each month and the number of relevant PCs is 0.46, while the correlation between the number of statistically significant factors, (as previously described) and the number of relevant PCs is 0.90.

On Panel B of Figure 5 we display the number of PCs required to explain 50%, 60%, 75%, 90% of the variation in the available factors. 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 PCs, ranging at various times from three to eight. However, explaining a larger portion of the variation in the factors requires many more PCs. We show below that these additional PCs are

[14] The number of PCs 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 PCs is nearly equal to the number of statistically significant factors, suggesting that virtually all the factors contain significant non-redundant information for the cross-section of stock returns.

economically important to understanding the cross-section of returns, to an extent that they cannot simply dismissed as absorbing minor variation that can be safely ignored.

To assess whether the number of relevant factor PCs, defined as above as the number that explains 95% of the in-sample variation in factor returns, has predictive power for OOS portfolio performance, we estimate regressions where the dependent variables are, in turn, the maximum OOS Sharpe ratio and the number of PCs used to construct the portfolio with the maximum Sharpe ratio. Outcomes are reported in Table 3. For Panel A the explanatory variable is the number of factor PCs that are relevant (i.e., that explained 95% of the variation in factor returns) in sample, while for results in Panel B the explanatory variable is the number of factors with significant (t-statistic greater than 3.0) in-sample explanatory power. We assess results both when out-of-sample portfolios are formed from the PCs of all factors available in-sample (columns 1 and 2) and from the PCs of only the factors that are significant in-sample (columns 3 and 4).

The results reported in Table 3 support two conclusions. First, the number of PCs employed to form the portfolio with the highest out-of-sample SR is significantly greater when more factors are significant in-sample as well as when more PCs are required to explain in-sample variation in factor returns. This relation is particularly strong (R^2 statistic of 0.82) in Panel B, column 4, where the number of PCs used to form portfolios with the highest out-of-sample Sharpe ratios is predicted by the number of factors with significant in-sample explanatory power. That is, during those in-sample periods where a larger number of factors can explain the cross-section of non-market returns, OOS Sharpe ratios are maximized when a larger number of factor PCs are employed. Second, the maximum OOS Sharpe ratios are greater when there are a higher number of significant in-sample factors or when more PCs are required to explain time series variation in in-sample factors. This relation as well is strongest (R^2 statistic of 0.40) in Panel B, column 3, when the predictive variable is the number of factors with significant in-sample explanatory power.

The CZ factors were initially discovered by widely disparate authors, which allows the possibility that some of the factors might be largely duplicative of the information contained in others. However, the results described here show not only that the factors collectively contain substantial information regarding the out-of-sample cross-section of returns, but that larger numbers of in-sample relevant factors contain more information relevant to predicting the cross-section of stock returns. That is, the additional significant factors contain non-duplicative information, and larger numbers of significant factors imply both higher OOS Sharpe ratios and that the most profitable OOS portfolios will be formed from a larger number of factor PCs. These outcomes provide scope for the possibility that advanced methods such as machine learning approaches may allow more precise forecasts of the number of factor PCs to be employed to obtain high OOS Sharpe ratios. Of course, implementation costs would also be relevant.

8. Do the outcomes reported herein reflect data-snooping by original authors?

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 most of the factor-related findings can indeed be replicated, do not arise from specification searches, and survive adjustment for multiple testing.[15]

Our results indicating large OOS Sharpe ratios for portfolios formed from factor PCs do not fully resolve this debate, since in some cases intervals that are out-of-sample in terms of our rolling estimation methods nevertheless include calendar months that were studied by the authors who first documented the empirical relevance of the characteristic or factor involved. However, we contribute to the discussion in three ways.[16] 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 Lee (2021), who mainly studied factors on a univariate

[15] 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).

[16] Bessembinder, Burt, and Hrdlicka (2023) study data drawn exclusively from time periods subsequent to those studied by original authors in their Table 6, and report that the CZ factors retain significant forecast power for the cross-section of returns.

basis, by studying a larger set of factors on a multivariate basis.[17] Second, while the related studies have 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 beyond simply tabulating outcomes for varying time periods. In Section 4 we construct measures of economic complexity and demonstrate that the numbers of economically relevant factors are systematically related to these measures.

Figure 6 displays information regarding the significance of each of the 205 CZ factors during periods before, during, and after the sample periods studied by the original authors. The figure includes one row for each factor, and a column for each sample month. Factors are sorted based on their unconditional CAPM alpha t-statistics during the original sample period, from lowest to highest. A given row and column contains a blue dot if the *t*-statistic on the factor's alpha estimated in a CAPM regression 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 data is 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 in Figure 6. 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. In Panel A of Table 4 we report data on the frequency of such occurrences. If significance is assessed by a t-statistic on the alpha estimate of 1.96 or greater, over three quarters (77%) of the factors have significant explanatory power during at least one 60-month interval prior to the range of dates used in the original studies, and 93% have significant explanatory power during at least one

[17] 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.

sixty-month interval after the range of dates used in the original studies. 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, 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. 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 6 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. In Panel B of Table 4 we report evidence 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 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.

9. Is the evidence of time variation in factor premia statistically significant?

The pattern displayed in Figure 6 and summarized in Table 4 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 quantity 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 4 that assesses the extent to which variation in the number of significant factors is 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 rely on a simulation, 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.[18] 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. Having done so we 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. We compile 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 under the null hypothesis of constant factor premia.

Panel A of Figure 7 displays the simulated distributions for the cross-factor average of the average spell lengths, while Panel B displays corresponding cross-factor average spell counts. The red dashed lines denote corresponding sample outcomes. The information displayed on Panel A of Figure 7 shows that the statistics pertaining to the actual sample and reported on Table 4 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 varied through time without immediate reversion to their long run means.

[18] We assume here that residual volatility is uncorrelated across factors. This simplifying assumption is consistent with the empirical outcomes reported in Section 2 indicating that the various factors contain essentially non-redundant information.

The information displayed on Panel B of Figure 7 shows that the cross-factor average number of significant spells also diverges from the distribution obtained under the null of constant premia, except when the t-statistic cutoff is 2.5. 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 that were obtained under the null hypothesis are more sensitive to the t-statistic cutoff employed as compared to actual sample outcomes. Empirical estimates of economically modest and time-invariant return premia are likely to be recategorized as insignificant rather than significant as the t-statistic hurdle increases. In contrast, the t-statistic employed is of less relevance for factor premia that are economically large during some periods (as they are then more likely to remain significant even as the t-statistic threshold is increased) and close to zero during other periods (when they are likely to be insignificant even with low t-statistic thresholds). That is, the simulation outcomes displayed on Figure 7 imply that it is exceptionally unlikely that the sample data reported on Table 4 could be observed under the null hypothesis of constant factor premia. Rather, the sample outcomes are consistent with those that would be anticipated if return premia associated with individual factors were economically large for some extended periods and near zero at other times. This finding contrasts with a perspective that has appeared in the prior literature, by which a return premium deemed to be anomalous should be permanently eliminated by arbitrage activity once the existence of the premium becomes broadly known.

10. The Role of Economic Complexity

We next assess whether the number of economically relevant factors is related to the state of the economy or to 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 Appendix Table C4.

11. The role of the number of listed firms.

We first consider the potential role of the number of firms traded in the U.S. markets. We reason that large changes in the number of publicly traded firms attributable for example to surges in IPO activity, firm failures, or merger waves, 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 over time. Multiple and varied risk factors may be necessary to explain patterns in the returns of varying firm types.

Column (1) of Table 5 reports outcomes obtained from regression of the number of statistically significant factors (Panel A) or the number of relevant factor PCs (Panel B) during months t *-59* to t on the number of firms publicly listed in month t. Each coefficient estimate is positive and statistically significant, and in Panel A the regression R^2 statistic is equal to 0.50, implying that the number of listed firms explains half the variation in the number of factors with significant explanatory power.[19] This finding supports the reasoning that a larger number of factors and PCs thereof are required to explain cross-sectional variation in mean returns when a greater number of firms are listed.[20] 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.[21]

[19] In Internet Appendix Table IA-2 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.

[20] Note that this result need not arise mechanically. If, for example, the CAPM determined expected returns for all stocks then the addition of new stocks with unique characteristics would only require estimation of their potentially distinct market betas, not the use of additional factors.

[21] 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. However, we show in Appendix Table C3 that we continue to estimate positive coefficients on the number of firms

12. Economic Complexity and Diversity in Firm Characteristics

We next assess the extent to which the number of significant factors and the number of relevant PCs thereof is related to measures 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 standard deviation. Thus, regression coefficients are interpreted as a response to a one-standard deviation change in that variable.

We conjecture that the business cycle may 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 one 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 ninety-day Treasury-bill rate and the spread between the 10-year treasury note yield and the ninety-day rate. Interest rates potentially capture the effects of monetary policy and funding conditions. The unemployment rate, the 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. To capture this aspect of 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.[22] To potentially

even when mean standard error of the alpha estimates is included as a regressor. We also show there 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.

[22] 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.

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 complexity, we use the Economic Complexity Index constructed by Simoes and Hidalgo (2011), which they describe as 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.[23] 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 like 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, in contrast, indicates that the underlying firms themselves are becoming increasingly differentiated.

Appendix Figure C3 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,

[23] 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.

approximately one hundred characteristics can be computed. This rapid growth reflects the fact that many characteristics require prior accounting statement data (often sparsely collected at the beginning of a firm's public life) and 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.

We report in columns (2) to (10) of Table 5 the results of univariate regressions of the number of statistically significant factors (Panel A) and the number of relevant PCs thereof (Panel B) on each of these measures of economic complexity in turn. The results indicate that the number of statistically significant factors is related to macroeconomic conditions, decreasing during recessions, increasing during periods of higher interest rates, but decreasing with the interest term spread. The unemployment rate, in contrast, does not have significant explanatory power. However, none of these macroeconomic variables is significant in explaining the number of relevant PCs. Further, 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 5 indicate that the number of statistically significant factors as well as the number of relevant PCs thereof are negatively related to the percentage of firms that pay dividends. 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. The coefficient estimates reported in column (7) of Table 5 indicate that the number of statistically significant factors (though not the number of PCs thereof) is 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 imply that increased institutional ownership reduces the numbers of significant factors, potentially because it effectively reduces variation in the identity of the marginal investor. The coefficient estimates on the economic complexity index (column 8) and the Amihud illiquidity measure (column 9) are not statistically significant in either Panel of Table 5.

The coefficient estimate for the diversity of firm characteristics (column 10) is positive and statistically significant in both Panels A and B of Table 5, with relatively large R-squared statistic of 0.38 in Panel A (number of significant factors) and 0.51 in Panel B (number of relevant PCs). This result implies that more factors and relevant PCs thereof are significant during those periods when there is greater cross-sectional variability in firm characteristics.

Columns 11 and 12 present results for multivariate specifications. We omit mean institutional ownership from the results in column 11, because data is available only from 1980 onward. The unemployment rate remains significant in each specification. The interest rate variables remain significant in Panel A. The coefficient on the cross-sectional mean Amihud illiquidity measure is positive and significant in Column 11 of each panel; 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 becomes insignificant in the multivariate setting.

Notably, the coefficient estimate on the diversity of firm characteristics remains positive and statistically significant in each specification, both when explaining the number of significant factors or the number of relevant PCs thereof. That is, even after allowing for the explanatory power of macroeconomic variables, cross-sectional variation in firm characteristics has explanatory power for the number of significant factors and relevant PCs. 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 in column 11 are large, equal to 0.72 in Panel A and 0.59 in Panel

B, implying that observable measures of economic complexity and firm diversity together explain most of the variation in the numbers of factors with economically significant explanatory power for the cross-section of stock returns. That is, time variation in the number of significant factors and the number of relevant PCs thereof is not random, but rather is linked to variation in macroeconomic conditions and observable diversity in firms' characteristics.

13. Conclusions

The reasoning that only a few factors or factor principal components 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 a substantial number of observable factors have explanatory power for the cross-section of stock returns be viewed as a collective failure?

We think not. Financial markets and the broader economy are complex and dynamic. The characteristics of the firms that are available for investment can change over 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. Further, investors may need time to learn about relevant firm characteristics and their associated expected returns. 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.

Cochrane (2011) observes that most variation in price-to-dividend ratios is attributable to changes in discount rates, i.e., expected returns. Prices are determined in 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 the same standard models, and can vary as market conditions, the economic environment, and other motivations for trading 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 empirical findings relevant to these issues, showing that non-latent, i.e., observable, factors identified in the literature have substantial forecast power for the cross-section of stock returns. Out-of-sample Sharpe ratios for portfolios formed from the PCs of in-sample factors are larger when time variation in factor premia is accommodated by means of rolling window estimation and when focusing on those factors with statistically significant in-sample explanatory power. Out-of-sample Sharpe ratios are larger when more factor PCs are used to form portfolios, implying that the factors identified by prior researchers are to a substantial extent not redundant of each other. Further, the conclusions of existing studies that assert that only a few factor PCs are relevant are altered when the only change in research design is to allow for time variation in factor premia.

We use simulation methods to show that neither the average number of periods where a factor is significant, nor the average duration of 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 is particularly strongly related to cross-sectional variation in observable firm characteristics. Our findings support 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.

Our findings imply that a time-varying number of non-redundant factors are required to price the cross-section of returns as the economy evolves and diverse firms are listed and delisted. In such a dynamic economy a factor can be significant in explaining returns during some periods but not others. This suggests that insignificant post-publication outcomes need not imply that a factor was unpriced in the original sample period nor that it will necessarily remain unpriced. However, accommodating such time variation may also provide additional scope for specification searches or other sources of bias. We avoid amplifying any such bias in this study by focusing only on previously-identified factors. Further, a requirement to link variation in estimated factor premia to variation in observable economic variables that were not directly employed to identify the factors imposes a degree of discipline. 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 firms with differing characteristics are listed and delisted.

References

Adrian, Tobias, Erkko 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.

Barillas, Francisco, Raymond Kan, Cesare Robotti and Jay Shanken. 2020. *Journal of Finanial and Quantitative Analysis,* 55(6), pp. 1840-74.

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.

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.

Gabaix, Xavier, and Ralph SJ Koijen. *In search of the origins of financial fluctuations: The inelastic markets hypothesis*. No. w28967. National Bureau of Economic Research, 2021.

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.

Hansen, Lars Peter, and Ravi Jagannathan., Implications of security market data for models of dynamic economicis. 1991. *Journal of Political* Economy. 99(2), pp 225-262.

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.

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

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.

Ilmanen, 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.

Lewellen, Jonathan, 2011, Institutional investors and the limits of arbitrage, *Journal of Financial Economics*, 102, 62-80.

Lewellen, Jonathan, and Jay Shaken, 2002, Learning, asset-pricing tests, and market efficiency, *Journal of Finance*, 57, 1113-1145.

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.

Moreira, A. and Muir, T., 2017. Volatility-managed portfolios. Journal of Finance, 72(4), pp.1611-1644.

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.

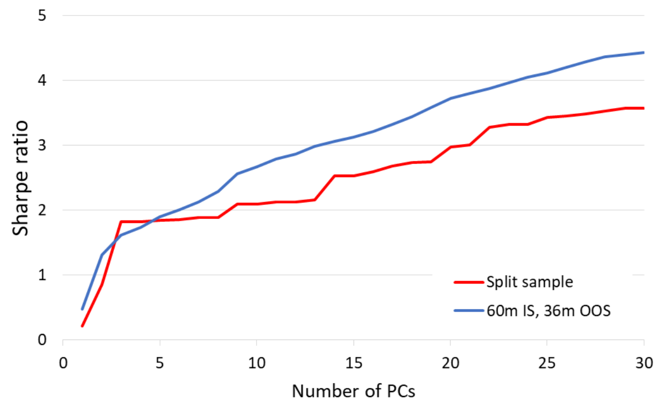
Stulz, Rene M, 2018, The shrinking universe of public firms: Facts, causes, and consequences, *NBER Reporter* 12-15.

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:*

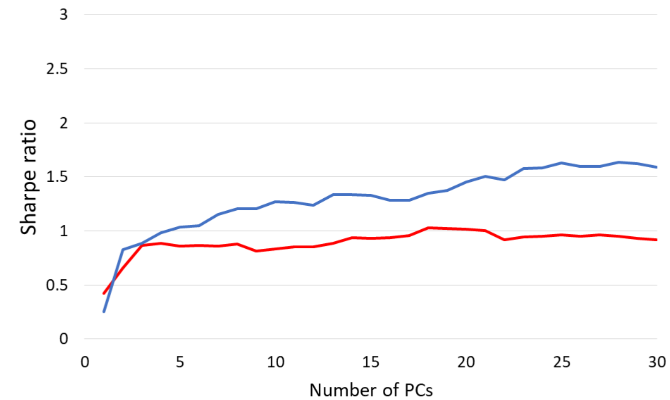
Jackson Hole Symposium, 19-65.

Figure 1. Effect of PCs' estimation and forecasting horizons on out-of-sample Sharpe ratios. This figure shows the average annualized Sharpe ratios of portfolios formed from increasing numbers of PCs for different estimation windows of two sets of test assets from 1965-2015. Panel A shows the averages of rolling in-sample Sharpe ratios obtained from forming optimal portfolios consisting of different numbers of PCs. 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 PCs 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 blue line shows the Sharpe ratio where the in-sample winodw is 60 months and the out-of-sample window is 36 months. For consistency with the split sample analysis, out-of-sample Sharpe ratios are computed only using the second half of the split sample.

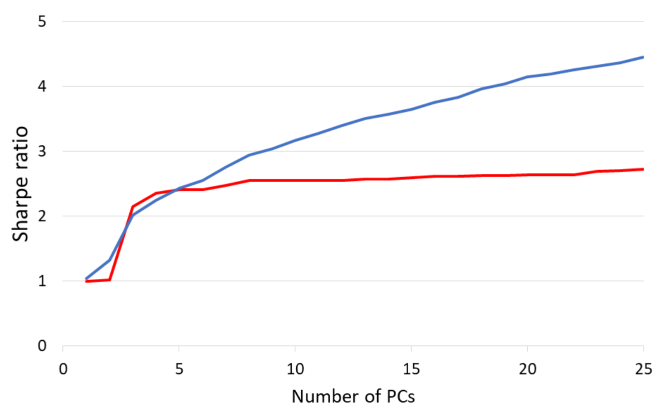
Panel A: In-sample Sharpe ratios – short/long legs of 15 anomalies



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

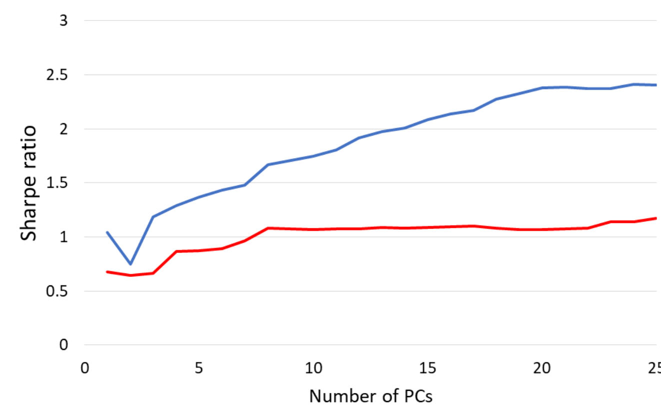
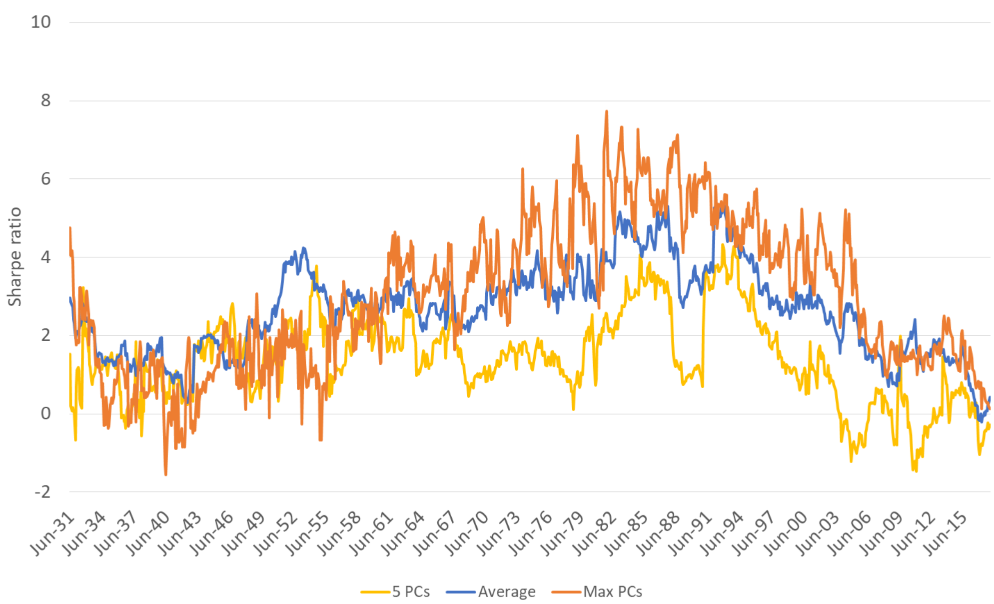


Figure 2. **Out-of-sample Sharpe ratios of portfolios constructed from principal components**. These figures show the annualized out-of-sample Sharpe ratios of portfolios formed from varying number of principal components. For each month t, we compute the factor loadings for the in-sample principal components for two sets of factor returns from *t – 59* to *t*. We then use the factor loadings to form out-of-sample principal components from *t + 1* to *t + 36*. We present 3 different Sharpe ratios formed from portfolios of PCs: 1) the SR from a portfolio of the first 5 PCs, 2) the average SR across all portfolios formed by increasing numbers of PCs, and 3) the SR from a portfolio of the maximum number of principal components. Panel A shows results for PCs calculated from the set of all factors in our sample. Panels B shows results using PCs calculated for the subset of significant factors. See section 2.B for the full methodology. The missing points in Panel B are due to the lack of significant factors in the in-sample period.

Panel A: Out-of-sample Sharpe ratios of PC portfolios computed for all factors.



Panel B: Out-of-sample Sharpe ratios of PC portfolios computed from significant factors.

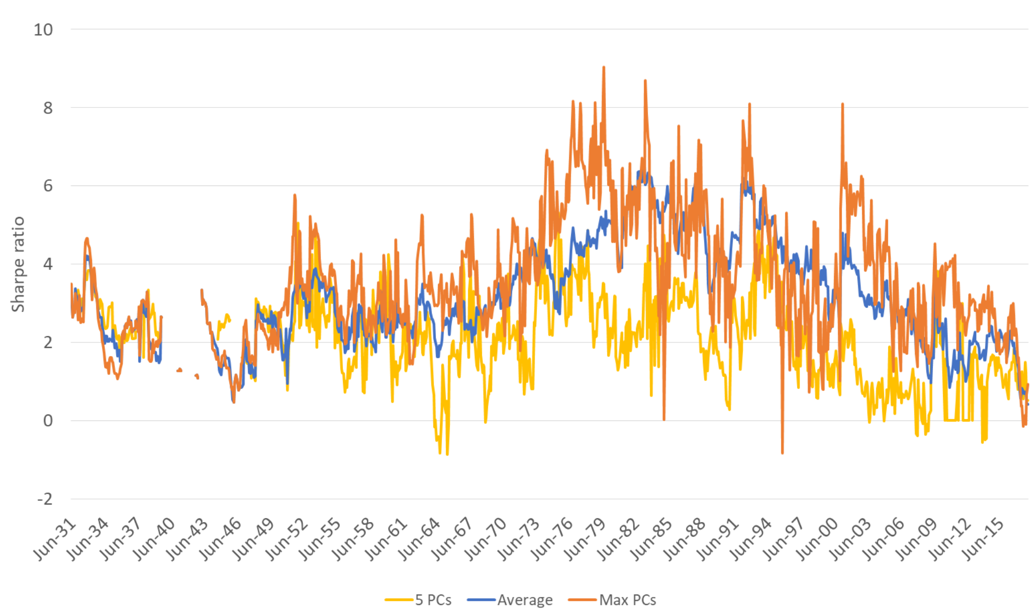


Table 1. Out-of-sample Sharpe ratios for portfolios consisting of varying numbers of principal components. This table provides descriptive statistics regarding the Sharpe ratios shown in Figure 2. We tabulate the mean and standard deviation of the time series of 4 different annualized Sharpe ratios formed from portfolios of PCs: 1) the SR from a portfolio of the first 5 PCs, 2) the average SR across all portfolios formed by increasing numbers of PCs, 3) the SR from a portfolio of the maximum number of principal components, and 4) the maximum SR attainable from any number of PCs. 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 2.B for the full methodology. The statistics tabulated under the heading "PCs from all factors" corresponds to Panel A of Figure 2, while those under "PCs from sig factors only" correspond to Panel B. The t-statistic on the difference of SRs from the 5-PC portfolio is calculated based on Hansen-Hodrick standard errors with a bandwidth of 60 are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

	PCs from all factors				**PCs from sig factors only**			
	5 PCs	Average	Max PCs	Max SR	5 PCs	Average	Max PCs	Max SR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average	1.394	2.640	2.977	4.242	2.164	3.200	3.570	4.676
Std dev	1.103	1.131	1.874	1.622	1.097	1.318	1.609	2.067
N	1039	1039	1039	1039	935	1005	1005	1005
Difference from SR of 5 PCs		1.246***	1.583***	2.848***		1.036***	1.406***	2.512***
t-statistic		7.22	3.55	9353		3.68	4.56	4.97
% difference		89%	114%	204%		74%	101%	180%

Figure 3. Average minimum and maximum weights within a portfolio of PCs. This figure shows the average minimum and maximum weights found within a portfolio of a given number of PCs. For each month t, we compute the in-sample principal components for the full sample of factor returns from $t - 59$ to t. We use the factor loadings to form out-of-sample principal components from $t + 1$ to $t + 36$ and construct portfolios comprised of increasing numbers of these out-of-sample PCs. The portfolio weights are the mean-variance optimal weights computed using the in-sample PCs. The red dots below plot the average maximum weight in a portfolio consisting of the given number of PCs shown on the x-axis. The blue dots plot the average minimum weight in the same portfolio. The sample of factors comes from the set of 205 "clear" and "likely" predictors provided by Chen and Zimmermann (2021) from 1931 to 2020.

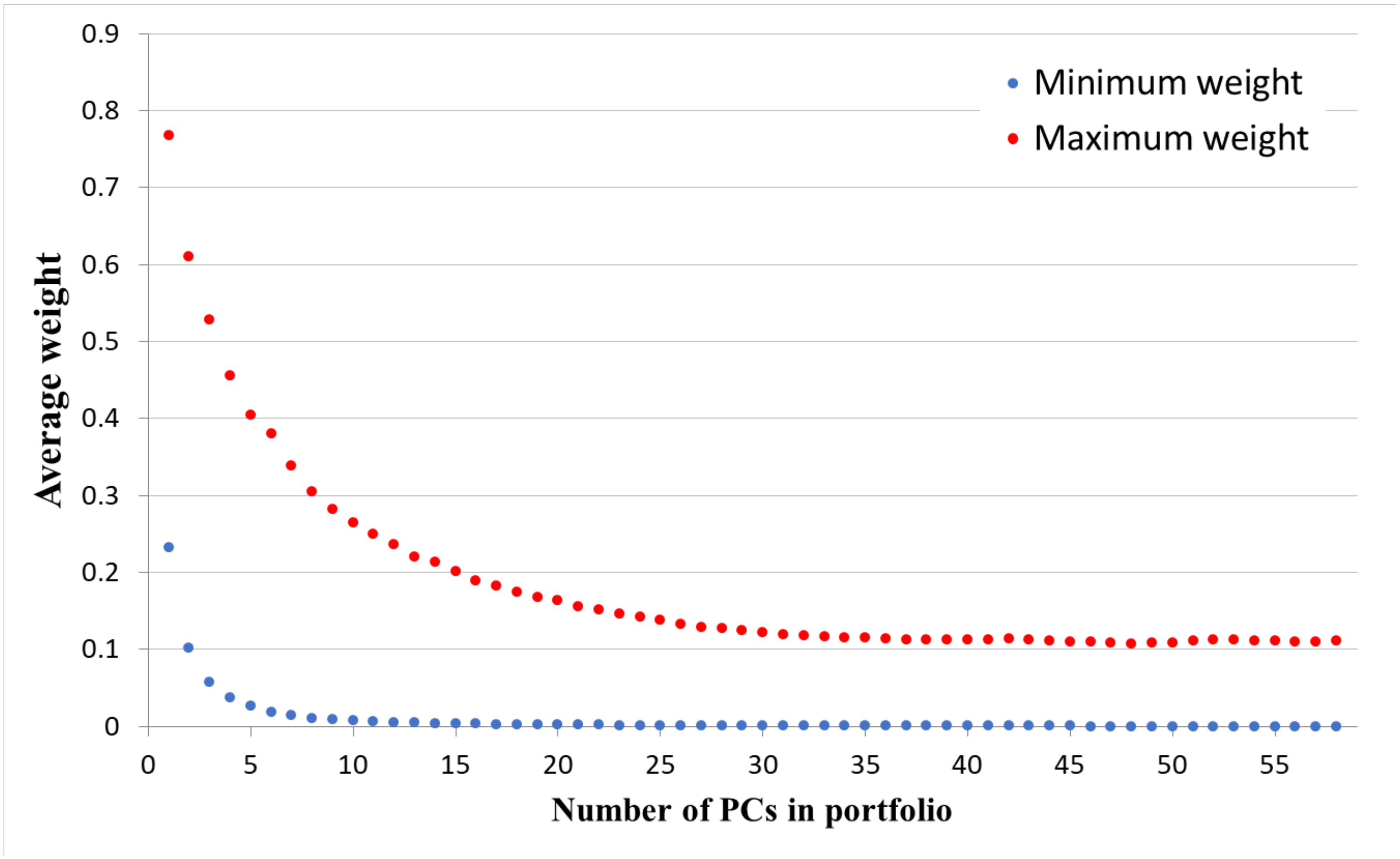


Figure 4. **Out-of-sample Sharpe ratios of equal-weighted portfolios constructed from principal components**. These figures show the annualized out-of-sample Sharpe ratios of portfolios formed from varying number of principal components for the set of all factors in our sample. For each month t, we compute the factor loadings for the in-sample principal components for two sets of factor returns from *t – 59* to *t*. We then use the factor loadings to form out-of-sample principal components from *t + 1* to *t + 36*. We present 4 different Sharpe ratios formed from equal-weighted portfolios of PCs: 1) the SR from a portfolio of the first 5 PCs, 2) the average SR across all portfolios formed by increasing numbers of PCs, and 3) the SR from a portfolio of the maximum number of principal components. See section 2.B for the full methodology. The missing points in Panel B are due to the lack of significant factors in the in-sample period.

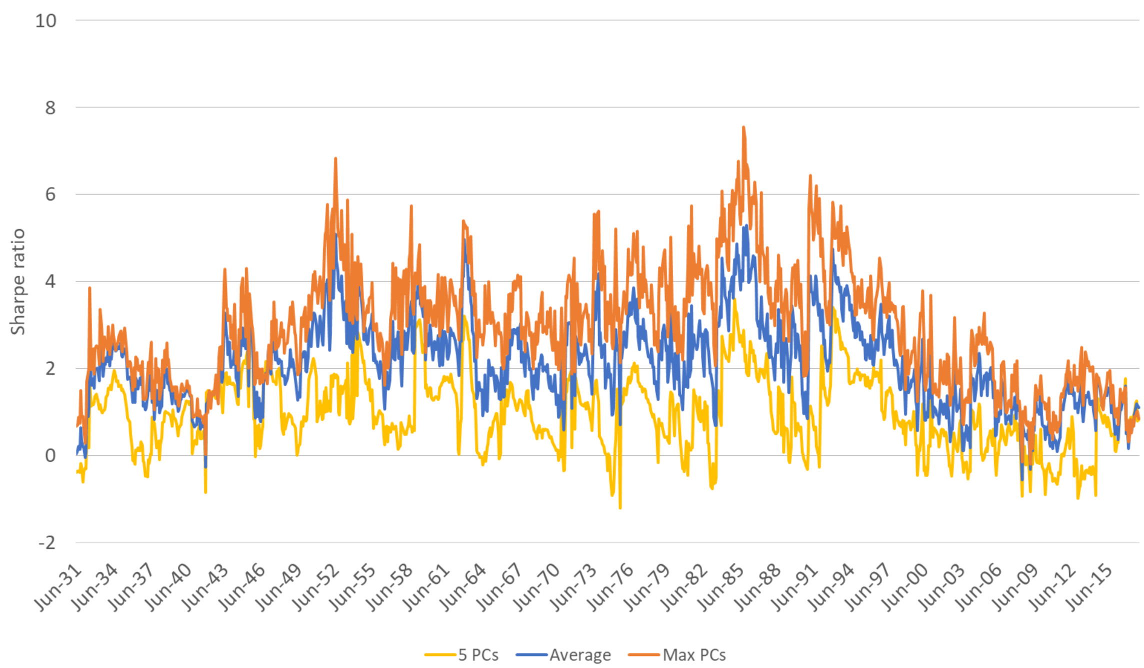


Table 2. Out-of-sample Sharpe ratios for equal-weighted portfolios consisting of varying numbers of principal components. This table provides descriptive statistics regarding the Sharpe ratios shown in Figure 4. We tabulate the mean and standard deviation of the time series of 4 different annualized Sharpe ratios formed from portfolios of PCs: 1) the SR from a portfolio of the first 5 PCs, 2) the average SR across all portfolios formed by increasing numbers of PCs, 3) the SR from a portfolio of the maximum number of principal components, and 4) the maximum SR attainable from any number of PCs. 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 2.B for the full methodology. The statistics tabulated under the heading "PCs from all factors" corresponds to Panel A of Figure 2, while those under "PCs from sig factors only" correspond to Panel B. The t-statistic on the difference of SRs from the 5-PC portfolio is calculated based on Hansen-Hodrick standard errors with a bandwidth of 60 are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

	Portfolio SR		
	5 PCs	Average	Max PCs
Average	0.952	2.110	2.957
Std dev	0.858	1.017	1.340
N	1039	1039	1039
Difference of SRs from 5 PCs	0.000	1.157***	2.005***
t-statistic		10.08	8.54
% difference		22%	111%

Table 3. **Relation between the number of relevant principal components, significant in-sample factors and out-of-sample Sharpe ratios**. For each month t, we compute the factor loadings for the in-sample principal components for two sets of factor returns from $t - 59$ to t. We then use the factor loadings to form out-of-sample principal components from $t + 1$ to $t + 36$. The number of relevant principal components is the number of PCs required to explain 95% of the cumulative variation in the in-sample factor returns from $t - 59$ to t. We also compute the number of significant factors by counting the total number of factors with a CAPM alpha that has a t-statistic greater than 3.0 over the same in-sample period. The maximum out-of-sample Sharpe ratio is obtained from the optimal portfolio of out-of-sample PCs. All Sharpe ratios are annualized. Panel A shows the results of regressing the maximum out-of-sample Sharpe ratios and the number of PCs that make up the maximum out-of-sample SR portfolio on the number of relevant principal components. Panel B shows the analogous results to Panel A but replaces the independent variable with the number of significant factors. Hansen-Hodrick standard errors with a bandwidth of 60 are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

Panel A. Out-of-sample PC portfolio Sharpe ratios and number of in-sample relevant PCs

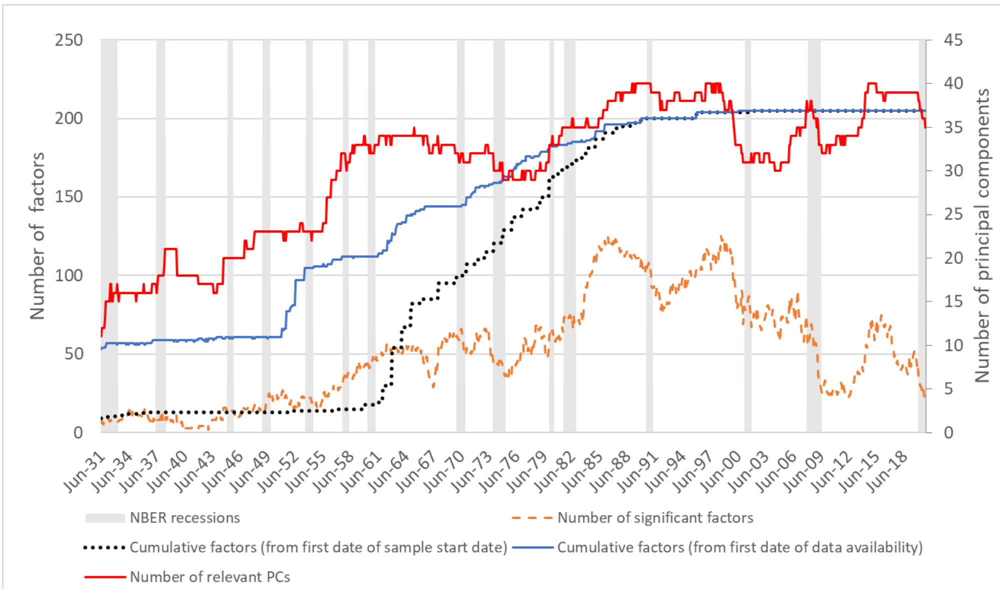
	Portfolio of principal components obtained from:			
	All factors		Significant factors	
	Max SR	Num PCs	Max SR	Num PCs
	(1)	(2)	(3)	(4)
Number of relevant PCs	0.07	1.22***	0.14***	1.36***
	(0.05)	(0.14)	(0.05)	(0.28)
Intercept	2.10	3.22	0.38	-24.97***
	(1.28)	(3.73)	(1.20)	(6.87)
R-squared	0.11	0.35	0.24	0.51
N	1039	1039	1005	1005

Panel B. Out-of-sample PC portfolio Sharpe ratios and number of in-sample significant factors

	Portfolio of principal components obtained from:			
	All factors		Significant factors	
	Max SR	Num PCs	Max SR	Num PCs
	(1)	(2)	(3)	(4)
Number of significant factors	0.04***	0.35***	0.06***	0.54***
	(0.01)	(0.08)	(0.01)	(0.03)
Intercept	3.34***	31.28***	3.19***	2.56**
	(0.47)	(3.41)	(0.48)	(1.17)
R-squared	0.26	0.27	0.40	0.82
N	1039	1039	1005	1005

Figure 5. Time series variation in the number of relevant principal components. This figure shows the number of relevant principal components that explain a large number of factors documented in the finance literature. For each month *t*, we compute the in-sample principal components for a set of factor returns from *t – 59* to *t*. We also compute the number of factors that have a significant (t-statistic > 3.00) CAPM alpha over the same in-sample period. The sample of factors comes from the set of 205 "clear" and "likely" predictors provided by Chen and Zimmermann (2021) from 1931 to 2020. The dotted black line in Panel A is the cumulative number of factors 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. The solid blue line in Panel A is the cumulative number of factors incremented at the date of each factor's first available return given the data available today. The solid red line of Panel A shows the number of relevant principal components, where the number of relevant principal components is the number of PCs required to explain 95% of the cumulative variation in the factor returns from *t – 59* to *t*. The dashed orange line shows the number of significant factors over time. Panel B shows the amount of variation in the factor returns explained by a given number of principal components. The grey vertical bars represent periods of NBER-defined recessions.

Panel A: Time series variation in relevant principal components and significant factors



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

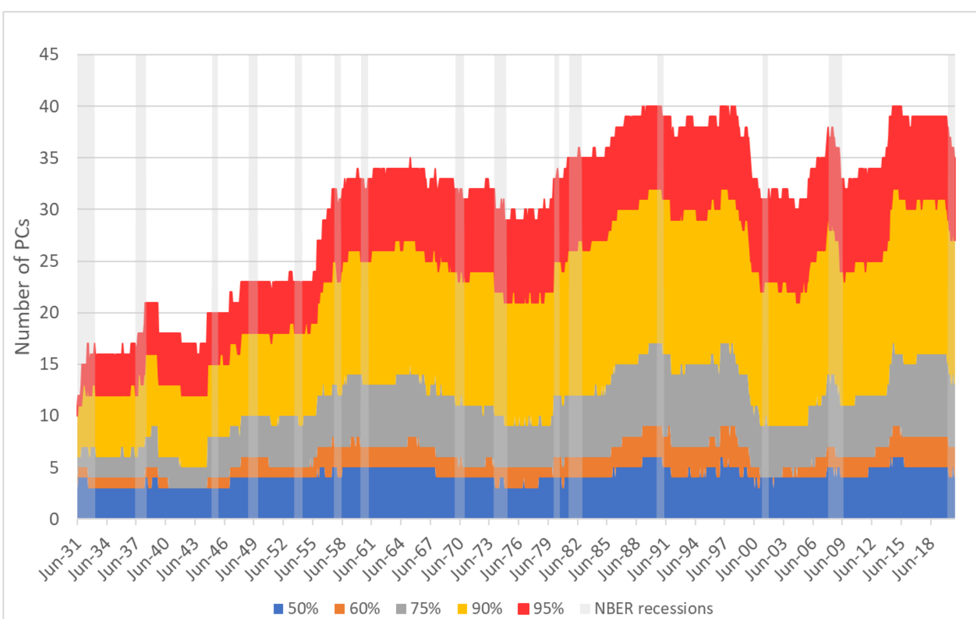


Figure 6. Time series variation in individual factors' significance. This figure shows whether a factor at a specific date has a statistically significant CAPM alpha over the preceding 60 months for the full sample of factors from 1931-2020. For each month *t*, we regress each factor's monthly returns from *t* – *59* to *t* on the market's monthly excess returns to obtain 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 dots signifying a month in which the factor is significant over the previous 60 months. 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 factors are sorted in increasing order of the unconditional t-statistic of their unconditional CAPM alpha during the original sample period. The grey vertical bars represent periods of NBER-defined recessions.

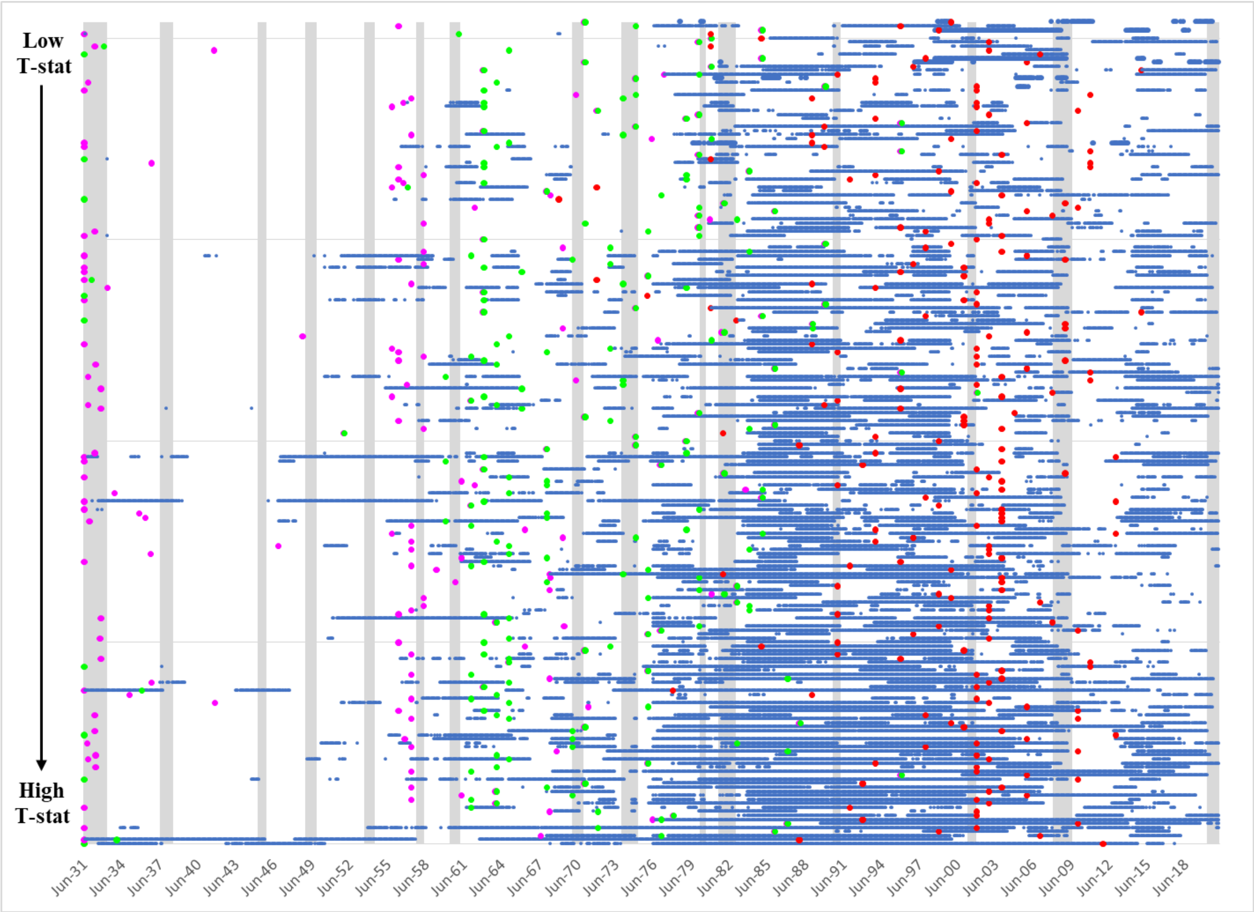


Figure 7. Simulated distributions for average significance spell lengths and counts. This figure shows simulated distributions of the cross-factor average of individual factors' average spell lengths and spell counts, when each factor has a constant return premium. We estimate for each factor a constant 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.

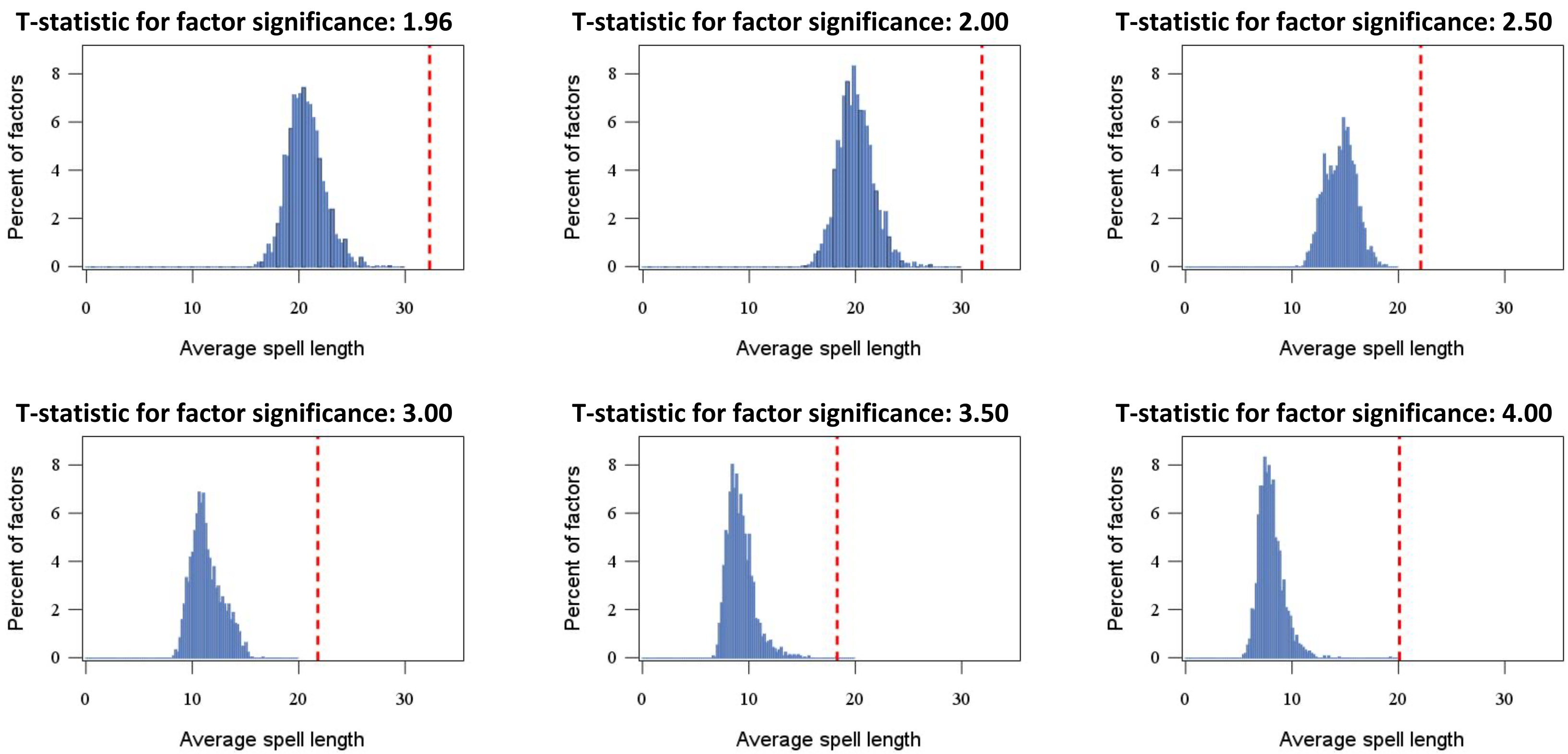


Figure 7 continued.

Panel B: Simulated distribution of average spell count.

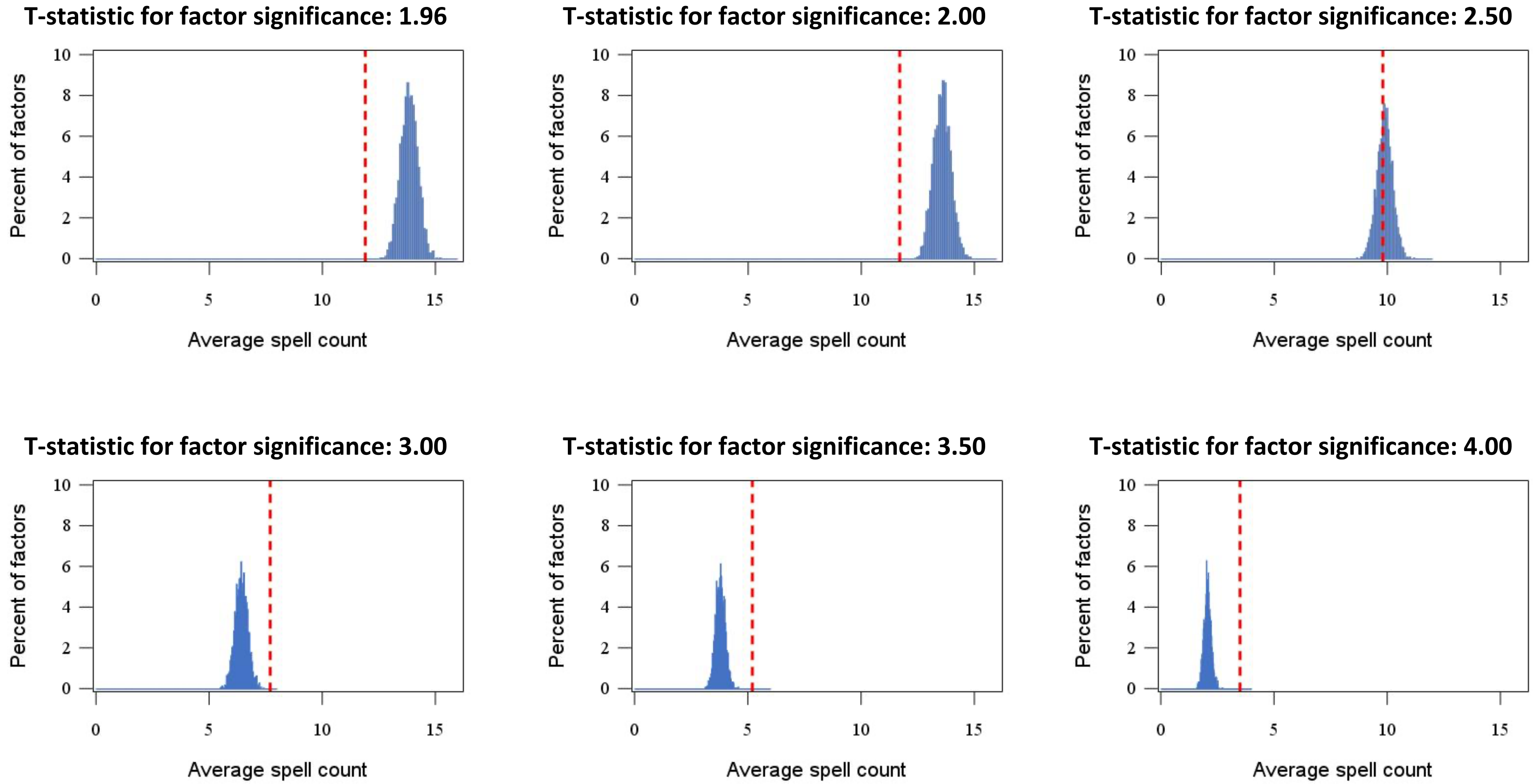


Table 4. **Summary statistics of factor significance spells across various thresholds of significance.** For each month *t*, we regress each factor's monthly returns from *t – 59* to *t* on the market's monthly excess returns to obtain each factor's CAPM alpha. A factor is significant at month *t* if 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

		% significant:	
t-statistic	p-value	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

		Cross-sectional statistics of factors' spell counts				
t-statistic	p-value	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

		Cross-sectional statistics of factors' average spells						
t-statistic	p-value	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. Comovement of the number of significant factors and relevant PCs with economy and firm characteristics. This table shows the results of regressing the number of significant factors or relevant principal components 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 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. We also compute the number of relevant principal components at each date *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*. Panel A uses the number of significant factors at each month t as the dependent variable, while Panel B uses the number of relevant factors at each month *t*. The independent variables are the same in both panels. 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 90-day T-bill rate is the 3-month Treasury Bill Secondary Market Rate and the 10-year treasury note yield spread is the difference of the market yield on U.S. treasury securities at a 10-year constant maturity and the 90-day T-bill rate. 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 C1 for a complete description of the measures. See Appendix Table C4 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.

Panel A: Comovement of the number of significant factors with economy and firm characteristics

	Dep var: Number of significant factors											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number of public firms	26.74***										6.56	-2.90
	(7.97)										(4.89)	(19.08)
NBER recession indicator		-11.90**									-7.44**	-2.59
		(5.58)									(3.16)	(3.97)
Unemployment rate			-5.30								-11.63***	-10.80***
			(4.22)								(2.45)	(2.00)
90-day T-Bill yield				8.08							12.29***	-1.08
				(5.38)							(3.67)	(4.91)
10-year T-Note yield spread					-0.02						8.41***	2.24
					(3.64)						(1.33)	(2.52)
% dividend-paying firms						-12.09*					-2.48	-13.32
						(6.21)					(8.05)	(16.60)
Mean institutional ownership							-16.81***					-19.51
							(4.46)					(13.32)
Economic complexity index								6.79			0.50	-4.29
								(4.20)			(2.36)	(2.74)
Mean Amihud illiquidity									8.25		9.07***	0.54
									(11.77)		(3.31)	(4.71)
Diversity of firm characteristics										17.23***	10.85**	18.41***
										(5.71)	(4.75)	(6.94)
R-squared	0.50	0.03	0.05	0.14	0.00	0.11	0.46	0.09	0.01	0.38	0.72	0.80
N	636	636	636	636	636	636	483	600	636	600	600	456

Table 5 (continued).

Panel B: Comovement of the number of relevant PCs with economy and firm characteristics

	Dep var: Number of relevant PCs											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number of public firms	2.02*										-1.72*	3.09
	(1.14)										(0.97)	(3.16)
NBER recession indicator		-1.14									-0.06	-0.27
		(1.06)									(0.56)	(0.52)
Unemployment rate			-0.62								-1.06***	-1.18***
			(0.46)								(0.39)	(0.42)
Fed funds rate				-0.41							-0.44	-1.88**
				(0.61)							(0.43)	(0.84)
10-year T-Bond yield					-0.40						0.90*	0.44
					(0.73)						(0.48)	(0.95)
% dividend-paying firms						-2.96**					-0.68	4.91
						(1.18)					(1.98)	(3.26)
Mean institutional ownership							-0.29					0.29
							(0.87)					(1.43)
Economic complexity index								0.46			-0.26	-1.20
								(0.80)			(0.38)	(0.74)
Mean Amihud illiquidity									0.83		2.46***	2.58***
									(2.00)		(0.87)	(0.90)
Diversity of firm characteristics										2.93***	3.42***	3.05***
										(0.62)	(0.96)	(1.06)
R-squared	0.13	0.01	0.03	0.02	0.01	0.30	0.01	0.02	0.01	0.51	0.59	0.54
N	636	636	636	636	636	636	483	600	636	600	600	456

1. Appendix A: PC estimation and portfolio construction

We estimate the principal components using 60-month rolling windows. For each month, *t*, we consider all factors that have non-missing returns for the in-sample months $t - 59$ to t.[24] It is important to note that the factor returns are neither centered nor standardized. For the subset of factor returns, $F_{t-59:t}$, from t – 59 to t, we compute the eigendecomposition of the factor covariance matrix

$$\sum = Q\Lambda Q^{-1} \text{ with } \Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

where Q is the square n x n matrix of eigenvectors (at each column i) and Λ is the diagonal matrix of eigenvalues. The eigenvectors are ordered based on the decreasing magnitudes of their corresponding eigenvalues. We construct the principal component factors (PCF) by right multiplying the factor returns from t – 59 to t + 36 with Q:

$$PCF_{t-59:t+36} = F_{t-59:t+36}Q$$

We classify the subset of principal components from $t - 59$ to t as the in-sample (IS) principal components, PCF_t^{IS} and those from $t + 1$ to $t + 36$ as the out-of-sample principal components, PCF_t^{OOS}.

With the principal component factors so defined, we proceed to construct out-of-sample portfolios of an increasing number of principal component factors as follows. First, for a given number of the first *n* principal component factors, we compute the mean-variance optimal weights from the in-sample principal component factors of a given number, where the weight matrix is

$$W = \frac{\sum^{-1}\mu}{1^T(\sum^{-1}\mu)^{abs}}$$

where $\sum$ is the covariance matrix of those *n* in-sample principal component factors, μ is the mean of the same, and 1 is the 1 x *n*-column ones vector. The superscript *abs* in the denominator denotes the absolute value which is taken element by element of the vector. As opposed to the traditional mean-variance weights formula, which scales the weights such that they sum to 1, we scale the weights such

[24] Generally, once a factor is non-missing, it continues to be non-missing through the end of the sample. There are 6 factors that become sparse after 2007.

that the absolute value of the weights sums to 1. Applying the absolute value operation to the denominator is the key difference between the weights formula we use and the traditional mean-variance formula.

Scaling by the sum of absolute values of the unscaled weights is required because the principal component factors consist of long-short portfolios. This constraint guarantees (in-sample) that the optimal portfolio falls on the upper part of the mean variance frontier. The economics of this constraint is that it takes seriously the issue that one cannot raise capital to go further long in one zero net cost portfolio by shorting another zero net cost portfolio, because shorting said portfolio raises no capital. Another interpretation of this constraint is that it takes seriously positions being scaled by wealth. A further benefit of this constraint is that weights in the optimal portfolio are well behaved and never extreme. This constraint imposes a maximum on a given weight is 1.

In contrast, applying the traditional scaling of simply dividing by the sum of the unscaled weights can give very wild weights. This can also lead to negative Sharpe ratios (in-sample) because the optimal portfolio so formed may fall on the lower part of the mean variance frontier. Though the absolute value of said Sharpe ratio will match the Sharpe ratio obtained using the absolute sum constraint, since the scaling parameters cancel in the numerator and denominator of the Sharpe ratio up to a sign. The prior literature deals with this inconvenience of negative Sharpe ratios by squaring the Sharpe ratio to give positive values, but that solution still often gives very extreme weights for the portfolio constituents especially with larger numbers of factors, which is a common criticism of using large numbers of factors.

Having obtained the optimal weights from the in-sample principal component factor returns, we construct the out-of-sample principal component portfolios by applying the in-sample optimal weights to the out-of-sample principal component factor returns from $t + 1$ to $t + 36$ for the first n principal components to create an out-of-sample portfolio consisting of 36 monthly returns. The Sharpe ratio is computed from this set of returns.

2. Appendix A: Reconciliation with Kozak, Nagel, and Santosh (2020)

Kozak, Nagel and Santosh (2020) report that, while a large number of factors are required to explain the cross section of returns to the fifty-anomaly based factors they study, a relatively sparse stochastic discount factor formed from only four PCs 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.

We investigate further. The computer code employed by Kozak, Nagel, and Santosh (2020) computes not only the out-of-sample R^2 statistic, but also out-of-sample Sharpe ratios.[25] 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, out-of-sample Sharpe ratios estimated in their sample by their program increase moderately from 0.75 with 4 factor PCs to 0.90 with ten factor PCs, and to 1.11 with 48 factor PCs.

However, these Sharpe ratio estimates may be biased due to the fact that their program computes PC eigenvectors over the full sample period.[26] We therefore modify their program to construct factor PCs 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 B1 displays the outcomes. Panel A displays the out-of-sample R^2 statistic, and corresponds to Figure 3A in Kozak, Nagel and Santosh (2020). 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 PCs, 0.94 with ten factor PCs, and 1.09 with 48 factor PCs. We conclude that the data and programs employed by Kozak, Nagel and Santosh (2020) also support that employing factor PCs beyond the first few leads to somewhat greater explanatory power for the cross-section of out-of-sample stock returns, even without any allowance for time variation in factor premia.

[25] More specifically, their program computes the square root of the expected squared Sharpe ratio.

[26] We are grateful to Stefan Nagel for identifying this bias and suggesting the solution to eliminate it.

Our results may also help to understand why Kozak, Nagel and Santosh (2020) report that many factors (as opposed to factor PCs), some with small SDF weightings, are necessary for good out-of-sample performance. In particular, we posit that time variation in factor premia contributes 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 the SDF as reported by Kozak, Nagel and Santosh (2020), and the percentage of months where the portfolio has a significant alpha.[27]

The results, displayed on Figure B2, 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 complied 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.

[27] 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.

Figure B1. 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 PCs 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 PCs. The '●' also reflects the maximum achievable out-of-sample r-squared in Panel A.

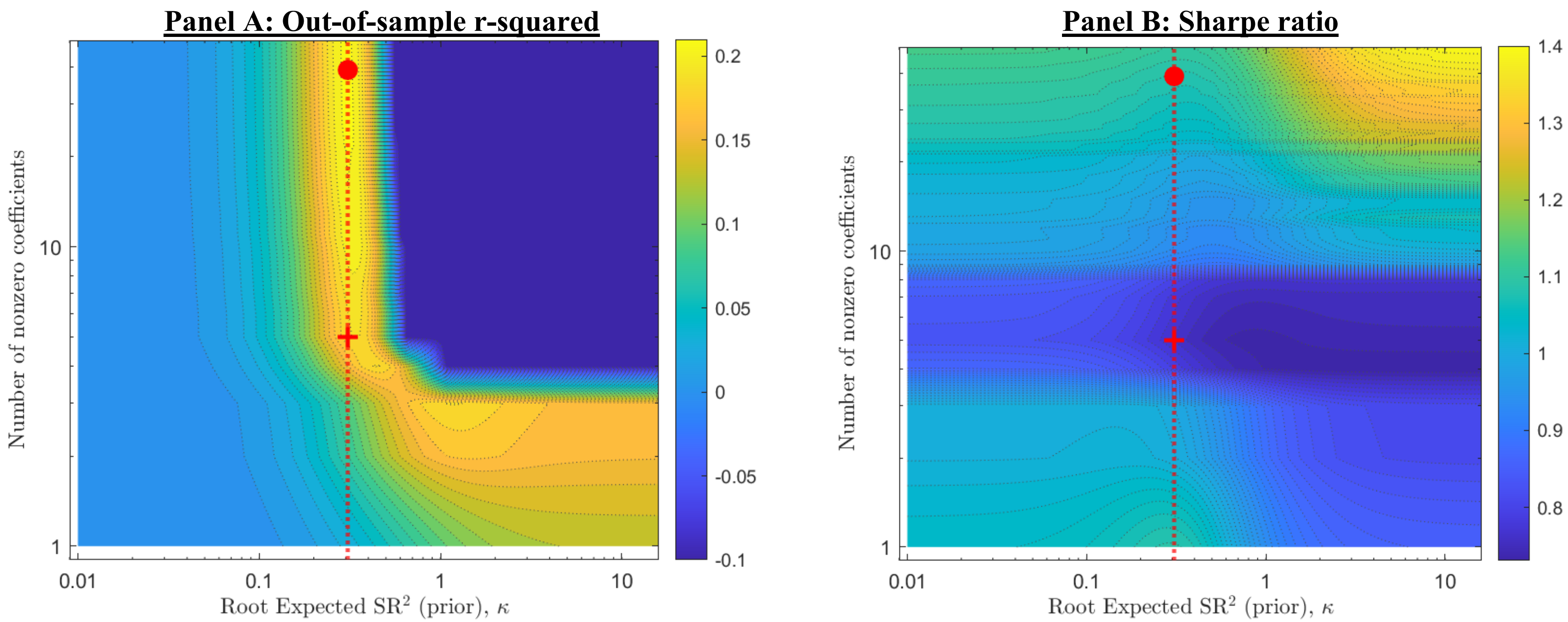
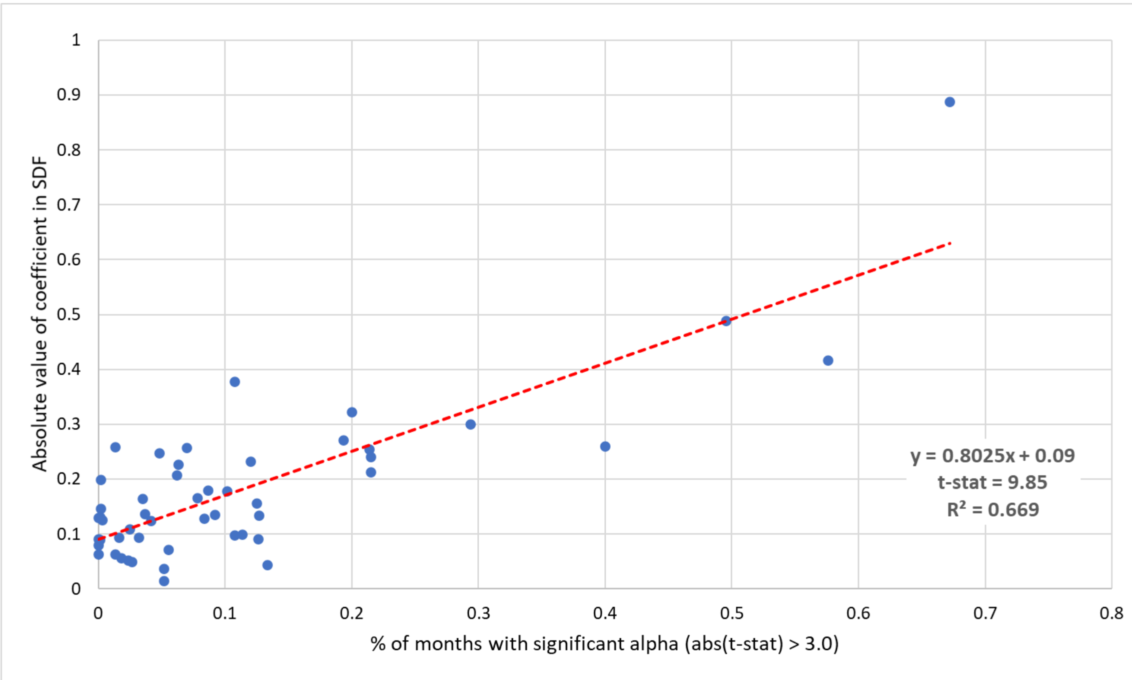


Figure B2. Factors fraction of time significant related to the size of its coefficient in optimal unconditional SDF. This figure shows that the coefficients in the optimal unconditional SDF of Kozak, Nagel and Santosh (2020) are positively related to the fraction of months that a given factor is significant 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.



Appendix C: Additional Figures and Tables

Figure C1. Regression slopes of in-sample and out-of-sample SRs by principal component. This figure shows the time series average of the rolling in-sample (IS) and out-of-sample (OOS) Sharpe ratios as well as the slope coefficient from a regression of the OOS SR on the IS SR. For each month *t*, we compute the in-sample principal components for the full sample of factor returns from *t – 59* to *t*. We also use the in-sample factor loadings to form out-of-sample principal component factors from *t + 1* to *t + 36*. We then calculate the in-sample SR for each PC individually and the corresponding out-of-sample SR for the same PC. The internet appendix includes a table and scatterplots of all SRs in the time series by principal component number.

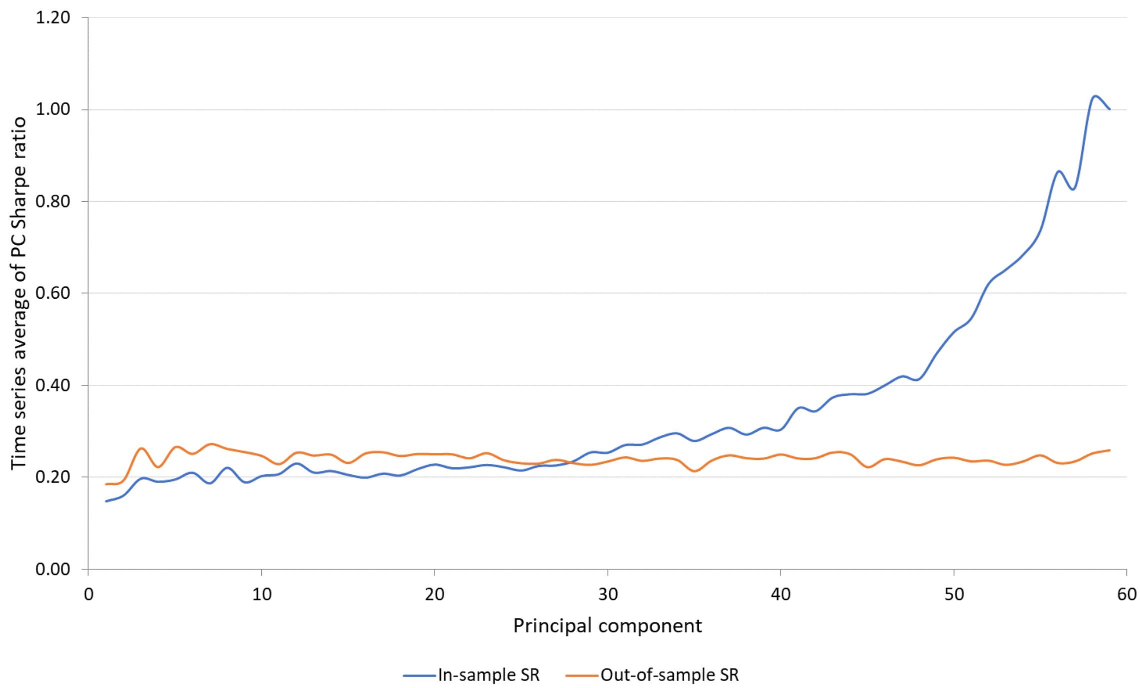
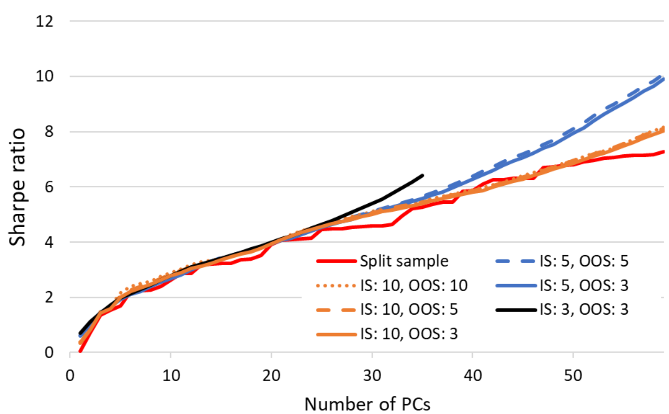


Figure C2. Effect of PCs' estimation and forecasting horizons on out-of-sample Sharpe ratios – alternative factor data. This figure replicates Figure 1 but uses the Chen and Zimmerman factors rather than the data provided by Kozak, Nagel and Santosh (2018). This figure shows the average annualized Sharpe ratios of portfolios formed from increasing numbers of PCs for different estimation windows. Panel A shows the averages of rolling in-sample Sharpe ratios obtained from forming optimal portfolios consisting of different numbers of PCs. Panel B shows the averages of rolling out-of-sample Sharpe ratios. The PCs 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 following 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. For consistency with the split sample analysis, the 10, 5, and 3-year out-of-sample Sharpe ratios are computed only using the second half of the split sample.

Panel A: In-sample Sharpe ratios – CZ factors



Panel B: Out-of-sample Sharpe ratios – CZ factors

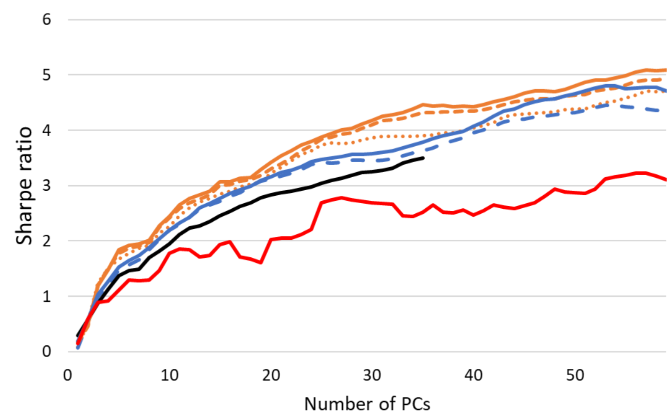


Figure C3. Cumulative number of non-missing cross-sectional characteristics over a firm's lifecycle. This figure shows the average number of cross-sectional characteristics available for each firm in a given month since the firm first appears in the cross-sectional characteristics dataset of Chen and 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.

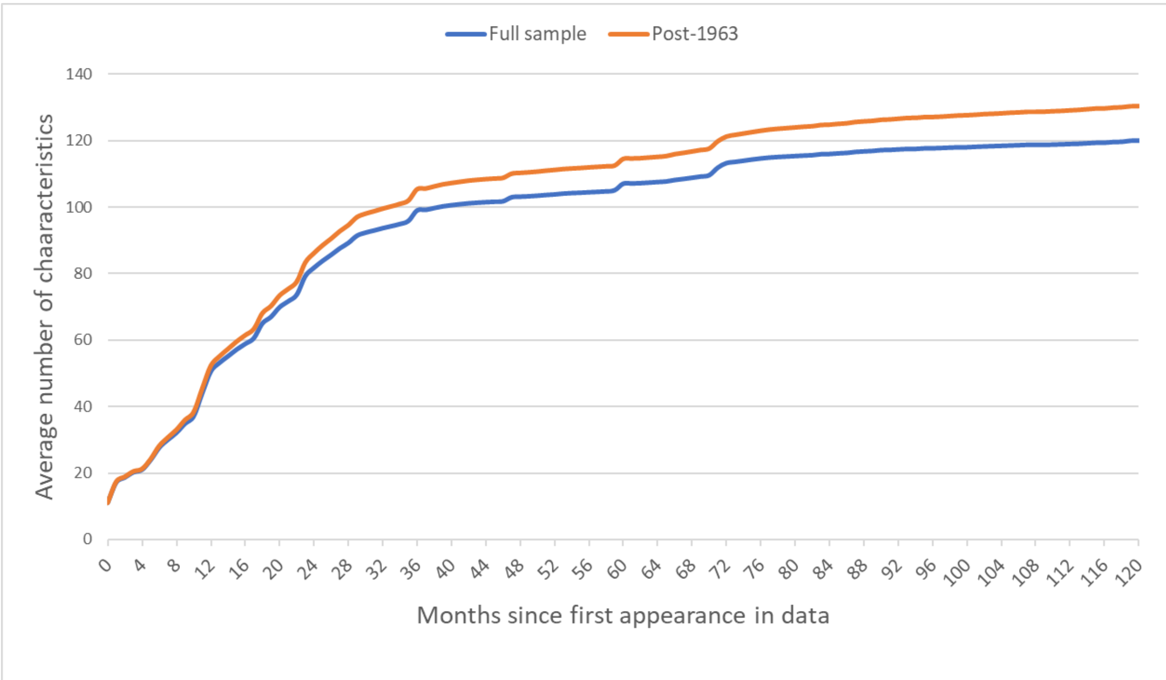


Table C1. Variable definitions This table summarizes the various variables we use throughout the analysis. The variables are listed in order of appearance in the paper.

Variable name	Description
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-59* to *t* 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-59* to *t* 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 relevant PCs	The number of principal components at month *t* required to explain 95% of the variation in factor returns from *t-59* to *t*. 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-60*. 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-59* to *t*). 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.

Table C2. Difference in out-of-sample Sharpe ratios of 5 PC and maximum PC portfolios. This table tabulates statistics for Figure 1 panels B and D. It shows the increase in the out-of-sample Sharpe ratio between a portfolio consisting of the first 5 PCs and the portfolio of the maximum number of PCs 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 PCs 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 years and out-of-sample window of 25 years replicates Kozak, Nagel and Santosh (2018) and matches the red solid lines in Figure C2.

Rolling estimation window (years)		Test assets	
In-sample	Out-of-sample	Anomalies	BTM
25	25	0.061	0.303
10	10	0.024	1.095
10	5	0.507	0.647
10	3	0.499	0.729
5	5	0.678	0.842
5	3	0.654	0.895
3	3	0.414	0.724

Table C3. 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**	-2.69	-30.01***	-19.39***	-17.01	-17.61**	-37.64***	-34.61***
	(3.48)	(1.98)	(6.98)	(6.72)	(13.42)	(8.22)	(4.03)	(4.13)
Number of public firms		17.31***		11.16***		26.91***		7.97***
		(2.90)		(1.67)		(7.26)		(1.67)
Mean factor absolute alpha			27.08***	18.51***			37.26***	31.47***
			(7.75)	(6.63)			(3.80)	(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

Table C4. Comovement of the number of significant factors and relevant PCs with economy and firm characteristics. This table replicates Table 5 of the main text for the sample years of 1931-2017. The table shows the results of regressing the number of significant factors or relevant principal components 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 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. We also compute the number of relevant principal components at each date *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*. Panel A uses the number of significant factors at each month t as the dependent variable, while Panel B uses the number of relevant factors at each month *t*. The independent variables are the same in both panels. 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 90-day T-bill rate is the 3-month Treasury Bill Secondary Market Rate and the 10-year treasury note yield spread is the difference of the market yield on U.S. treasury securities at a 10-year constant maturity and the 90-day T-bill rate. 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 C1 for a complete description of the measures. Hansen-Hodrick standard errors are in parentheses. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels.

Panel A: Comovement of the number of significant factors with economy and firm characteristics

	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***										5.85	5.40	-4.97
	(3.08)										(6.41)	(4.73)	(19.75)
NBER recession indicator		-9.41**									-9.90**	-8.92**	-2.27
		(4.66)									(4.17)	(4.10)	(3.70)
Unemployment rate			0.11								-11.16***	-9.76***	-9.81***
			(4.38)								(1.77)	(2.33)	(1.73)
Fed funds rate				8.95*							-5.62**	-4.20*	-2.90
				(5.18)							(2.46)	(2.18)	(4.63)
10-year T-Bond yield					9.64*						15.46***	15.04***	1.83
					(5.23)						(4.48)	(3.77)	(5.94)
% dividend-paying firms						-12.12***					-1.32	-2.65	-16.19
						(4.47)					(8.98)	(7.94)	(17.84)
Mean institutional ownership							-16.81***						-20.14
							(4.46)						(13.18)
Economic complexity index								5.59				1.83	-3.91
								(4.72)				(2.30)	(2.54)
Mean Amihud illiquidity									-5.30***		12.97***	11.84***	0.70
									(1.90)		(4.33)	(4.22)	(4.55)
Diversity of firm characteristics										14.76***	9.32*	10.13**	18.51**
										(4.76)	(5.29)	(4.97)	(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

Table C4 (continued).

Panel B: Comovement of the number of relevant PCs with economy and firm characteristics

	Dep var: Number of relevant PCs											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number of public firms	5.70***										-1.92	3.09
	(1.33)										(1.20)	(3.16)
NBER recession indicator		-2.61									-0.35	-0.27
		(1.65)									(0.62)	(0.52)
Unemployment rate			0.86								-1.03***	-1.18***
			(1.01)								(0.39)	(0.42)
Fed funds rate				0.05							-0.43	-1.88**
				(0.76)							(0.39)	(0.84)
10-year T-Bond yield					-0.32						0.77	0.44
					(0.75)						(0.51)	(0.95)
% dividend-paying firms						-3.74***					-0.16	4.91
						(1.40)					(1.98)	(3.26)
Mean institutional ownership							-0.29					0.29
							(0.87)					(1.43)
Economic complexity index								0.38			-0.05	-1.20
								(0.78)			(0.49)	(0.74)
Mean Amihud illiquidity									-3.73***		2.94***	2.58***
									(0.62)		(0.93)	(0.90)
Diversity of firm characteristics										3.32**	3.48***	3.05***
										(1.50)	(0.89)	(1.06)
R-squared	0.58	0.02	0.03	0.00	0.01	0.25	0.01	0.01	0.25	0.21	0.56	0.54
N	1075	1075	876	798	708	1075	483	648	1075	1039	648	456