Decreasing Returns to Scale has Eroded Hedge Fund Performance Persistence

Nicolas P.B. Bollen[*], Juha Joenväärä[†], and Mikko Kauppila[‡]

This version: May 15, 2024

ABSTRACT

This paper successfully replicates Kosowski, Naik, and Teo (2007) and Jagannathan, Malakhov, and Novikov (2010), two seminal studies of hedge fund performance persistence. We show that top funds continue to persist in a more recent sample, even when using novel "real-time" data that approximates an investor's actual information set. The persistence available to investors has substantially weakened, however, and is only observed when using Kosowski et al.'s Bayesian alpha to predict performance. We identify the econometric source of the superiority of Kosowski et al.'s methodology and show that the decline in performance persistence is associated with decreasing returns to scale for superior funds.

JEL Classifications: G11, G23

Keywords: performance persistence, decreasing returns to scale, replication, hedge funds

[*]Nicolas P.B. Bollen (corresponding author), Owen Graduate School of Management, Vanderbilt University, nick.bollen@vanderbilt.edu; [†]Juha Joenväärä, Aalto University School of Business, juha.joenvaara@aalto.fi; [‡]Mikko Kauppila, Aalto University School of Business, mikko.j.kauppila@aalto.fi. Nicolas Bollen thanks the Hans Stoll Financial Markets Research Center at Vanderbilt University for research support. We thank seminar participants at the 10th Annual Paris Hedge Fund Conference, the 2022 Luxemburg Alternative Investments Workshop, the 2023 Eastern Finance Association Meeting, Philippe Jorion, Roman Kräussl, Alexey Malakhov, and two anonymous referees for their helpful comments and suggestions.

Aggregate hedge fund performance declined substantially following the 2008–2009 global financial crisis, as documented by Bollen et al. (2021), McCahery and de Roode (2021), Sullivan (2021), and Ardia et al. (2024). Furthermore, Dichev and Yu (2011) and Bali et al. (2013) argue that the benefit accruing to hedge fund investors was smaller than commonly believed in the years prior. Despite these findings, assets under management in the hedge fund industry were around $6 trillion at year-end 2019 (Barth et al., 2023), an all-time high, reflecting great demand from institutional investors. University endowments, for example, allocated 20% of their capital to hedge funds in 2020 (NACUBO, 2021). In an era of weaker industry-wide performance, a continued allocation would be justified if one could select individual hedge funds that subsequently deliver abnormal performance.

A rich literature explores a wide variety of characteristics of funds and their managers to discover the most informative predictor of future fund performance. Some of these are not related to past performance, such as the type of car purchased by a manager (Brown et al., 2018) and the average SAT score of a fund manager's undergraduate institution (Li et al., 2011). However, most prior studies focus on past performance following the intuition that a manager's skill is long-lasting and could give rise to performance persistence (Grinblatt and Titman, 1992). Performance measures include a fund's distinctiveness (Sun et al., 2012), the manager's market timing ability (Chen and Liang, 2007), and various measures of alpha, defined as the average return unexplained by one or more proxies for sources of systematic risk (Fung and Hsieh, 2004; Kosowski et al., 2007; and Jagannathan et al., 2010). Bollen et al. (2021) review 26 predictor variables and find that the alpha measures of Kosowski et al. (2007), hereafter KNT, and Jagannathan et al. (2010), hereafter JMN, perform the best. Consequently, we critically review and replicate these two seminal studies of hedge fund performance persistence and determine whether the techniques they develop continue to provide useful guidance in the years following their publication.

The two papers we replicate analyze hedge fund data spanning roughly the same period, 1994–2002 in KNT and 1996–2005 in JMN. So perhaps it is no surprise that their conclusions are qualitatively identical: both find evidence of persistence in hedge fund performance. However, the two sets of authors use different techniques for measuring abnormal performance. In addition, the hedge fund data commonly available for academic research requires many subjective treatments to mitigate potential biases, and KNT and JMN address these concerns differently. For these reasons, a side-by-side replication provides insight for future research along both econometric and data management dimensions.

Recent replications of important studies of mutual fund performance persistence include Choi and Zhao (2021), who replicate Carhart (1997), and Riley (2021), who replicates Kosowski et al. (2006). Like most studies of mutual fund performance, these papers focus on actively managed equity funds and use the CRSP Survivor-Bias-Free U.S. Mutual Fund Database. The comprehensive coverage of mutual funds in the CRSP database is made possible by the reporting requirements imposed by the SEC Investment Company Act. In contrast, hedge fund managers have not until recently been required to disclose returns. Those who opt to report returns have their choice of multiple database vendors. The voluntary nature of hedge fund reporting creates two sets of challenges for researchers: first, deciding which commercial databases to use and how best to consolidate them, and second, deciding how to apply treatments to address any potential biases in the data. From a replication perspective, researchers must attempt to construct a sample following the same set of decisions made by the original authors, which requires a sufficiently explicit description in the prior work.

Our first set of analyses involves replicating the procedures in KNT and JMN for assessing hedge fund performance persistence. We construct samples that match as closely as possible the samples used in KNT and JMN, which, as described above, is complicated by the nature of voluntarily reported hedge fund returns. Despite the challenges created by the hedge fund data, we can replicate summary statistics and evidence of persistence in the two papers quite well. KNT, for example, report that funds in their sample generated an OLS alpha from the Fung and Hsieh (2004) seven factor model, hereafter FH7, with an average *t*-statistic of 1.43 compared to our replication's 1.42. Using their Bayesian *t*-statistic of alpha as a sorting variable, KNT find the top decile generates an annual holding period FH7 alpha of 6.56% versus 6.34% in our replication.[1] JMN compute a cross-sectional average monthly excess return of funds in their sample year-by-year, and our replication's average excess return is within 1 basis point for six of the nine years in the sample. Using their relative alpha as a sorting variable, JMN find the top tercile generates an average monthly holding period FH7 alpha that is significant in three of their four sub-samples, with magnitudes of 33bp, 27bp, and 18bp, respectively. Our replication finds significance in the same three sub-samples with monthly average holding period alphas of 29bp, 21bp, and 18bp.

Our second set of analyses moves beyond replication and explores whether an investor can exploit hedge fund performance persistence in the years following the publication of KNT and JMN. To provide as meaningful an assessment as possible, we use a union of six commercial databases with data through June 2020 and account for real-time data availability. We show that KNT's Bayesian alpha *t*-statistic can predict future alphas over its post-publication 2003–2020 period for top decile funds, whereas JMN's relative alpha cannot. As discussed later in the paper,

[1] Both KNT and JMN measure holding performance by the FH7 alpha. Performance measurement is naturally affected by benchmark choice. We use FH7 alpha throughout to facilitate comparison with the results of KNT and JMN, and other prior work for which FH7 is the most widely used benchmark.

we find that the information content of the KNT model is preserved when dispensing with the longer time series offered by the non-benchmark assets in the underlying Pàstor and Stambaugh (2002) framework. We conclude that the source of KNT's advantage is that it avoids over-fitting in-sample. However, even the Bayesian alpha provides only modest improvements over the simple OLS alpha as a performance predictor.

The top decile of funds as sorted by the Bayesian alpha *t*-statistic generates a holding period FH7 alpha over the full period that is less than half as large as our estimate when replicating the original KNT sample. These results are consistent with existing evidence of a decline in aggregate hedge fund performance in the years following the financial crisis, as documented in Sullivan (2021) and elsewhere. Bollen et al. (2021) study several possible explanations for a decline in hedge fund performance and point to decreasing returns to scale as a likely candidate. Existing studies are inconclusive. Agarwal et al. (2009) and Yin (2016), among others, find a significant coefficient on lagged size as a determinant of fund performance. Rzakhanov et al. (2019), Cao et al. (2021), and Forsberg et al. (2022), present evidence of decreasing returns to scale in the aggregate, but not at the fund level. We argue that decreasing returns to scale is an observable mechanism only for those managers who have alpha-generating skill. By pooling observations of size and performance across all funds regardless of their alpha, standard tests have low power. We document decreasing returns to scale at the fund-level by including an interaction of lagged size and lagged performance. We also find that decreasing returns to scale itself has declined over time, consistent with the notion that the impact of scale at the industry level has depleted alpha opportunities at the fund level in more recent years.

Our paper contributes to several research streams. First, we add to the relatively nascent literature that replicates important prior studies of professional asset management and tests for temporal robustness. We replicate the results reported in KNT and JMN, two foundational papers of hedge fund performance persistence, and compare the information content of their respective predictors in a unified testing framework. We present evidence that KNT's Bayesian alpha can sort funds into subsets that generate more robust performance differences than JMN's relative alpha in a post-publication sample. We perform a battery of diagnostics and conclude that the superiority of the KNT methodology is largely attributable to a better balance between in-sample fit and out-of-sample predictability. The performance of KNT's top funds has declined over time, however. These results complement recent studies that examine performance persistence in mutual funds (Choi and Zhao, 2021, and Riley, 2021) and private equity funds (Harris et al., 2022). Taken together, this body of work indicates that though persistence has weakened for all active fund types, those in the alternative space continue to feature some degree of performance persistence.

Second, we provide new insights regarding the use of commercial hedge fund databases typically studied in academic research. Most studies acknowledge the presence of backfill bias resulting from a manager's discretion in uploading a fund's historical performance prior to the "listing date," i.e., the initiation of fund reporting. Jorion and Schwarz (2019) argue it is important to drop all observations prior to the listing date to eliminate the upward bias in historical performance. In our setting, this "exact" backfill approach is also critical as we are modeling an investor who relies on the databases to identify available funds, hence the investor would not be aware of a fund prior to the listing date. After applying a backfill correction, the standard operating procedure is to assume a return observation for any prior date was available to an investor at that time. This procedure ignores possible revisions to archived return observations (Patton et al., 2015), delays in reporting to the database (Aragon and Nanda, 2017), and fund eliminations when the entire track record is removed from the vendor database (Bhardwaj et al., 2014).[2] We compare two approaches to dealing with these issues. The first uses one recent download of the databases and addresses reporting delays by imposing a three-month lag between the last month of a prior performance assessment for the purpose of fund selection and the first month of an investor's holding period.[3] We use the term "real-time data" to refer to the combined application of the exact backfill correction and the reporting delay, as we are estimating the information available to an investor at a prior date. The second uses an archive of historical database downloads to ensure that our simulated hedge fund selection procedure only uses data that were available at a given date. We find that the two approaches yield similar results, indicating that researchers can rely on the most recent downloads of the databases, so long as they apply the exact backfill correction and impose a realistic gap between an investor's ranking period and holding period.

Third, in our analysis of decreasing returns to scale, we employ for the first time an interaction of lagged size and lagged performance. This interaction addresses both the noise introduced by including observations of funds with insignificant performance, as well as the unspecified relation between size and lagged performance discussed in Cao et al. (2021). We find that decreasing returns to scale is more pronounced in those funds with prior evidence of skill, and that this relationship has weakened over time, likely the result of the impact of industry size documented in prior studies.

Despite evidence of a reduction in the aggregate performance of hedge funds over time, and the related ability of investors to exploit persistence in forming hedge fund portfolios, we find that the top decile of funds as identified by the KNT Bayesian alpha *t*-statistic outperforms even

[2] Return revisions are not confined to the hedge fund databases. Akey et al. (2021) document changes to the Fama–French factors and show they can affect mutual fund alpha estimates.

[3] We test for robustness to variation in the length of the lag as discussed later in the paper.

in realistic three-year holding periods. Our real-time persistence results are only slightly weaker for fund-of-funds, indicating that investors may be able to exploit hedge fund persistence. To the extent that the best funds tend not to report to commercial databases, as found by Barth et al. (2023), this result likely puts a lower bound on the potential benefit of hedge fund investment when investors employ a selection methodology similar to the one described next.

1. Replication of Kosowski, Naik, and Teo (2007)

1.1. Data

KNT construct their dataset by consolidating three commercial databases: Lipper TASS (TASS), Hedge Fund Research (HFR), and CISDM (now part of Morningstar). Though the databases include observations of returns as early as January 1990, KNT only use data from 1994 onwards to mitigate survivorship bias (Ackermann et al., 1999; Brown et al., 1999; Fung and Hsieh, 2000, 2009; and Liang, 2000). KNT retain only those funds that report returns net of all fees, and only those funds with AUM ≥ 20 million USD, because smaller funds do not have the capacity to attract institutional investors. KNT divide funds into five broad styles in a manner like Agarwal et al. (2009): Directional traders, Multiprocess, Relative value, Security selection, and Fund of funds. Given their prominence, Long/short equity funds are sometimes analyzed separately, although they are a subset of Security selection funds.

To replicate KNT and extend the analysis through June 2020, we aggregate October 2020 snapshots of six commercial databases as in Joenväärä et al. (2021): BarclayHedge, Eurekahedge, Hedge Fund Management (HFM), HFR, Morningstar, and TASS. We use KNT's exclusionary criteria and style classification as described above. Our union of six hedge fund databases likely results in a replicating sample that differs somewhat from the KNT data set for at least two reasons. First, Morningstar's database does not include "graveyard" funds that stopped reporting prior to

2005. Second, Aggarwal and Jorion (2010a) document that Tremont's purchase of TASS induced a spurious survivorship bias because the Tremont database was not absorbed directly into the TASS database.

Summary statistics of the samples are displayed in Table 1. Panel A shows results reported by KNT whereas Panel B shows statistics of our replication over the same 1994–2002 period. Though we use three additional databases, the number of funds in our sample is 289 fewer than the number in KNT. A possible explanation for this difference is the way in which we aggregate share classes to the fund level. In addition, some funds present in the database snapshots used by KNT may have been dropped by the vendors prior to our downloads or are not in the sample because we do not have access to some Morningstar funds that stopped reporting. Our inability to exactly match the KNT sample illustrates how the nature of archival hedge fund data by itself complicates replication. Despite the apparent differences in our samples, the summary statistics are almost identical. We find a monthly average alpha of 0.46%, for example, versus 0.42% in KNT. Both samples also feature a great deal of non-normality: roughly 40% of the funds in our sample reject the Jarque–Bera test at the 10% level versus about 39% in KNT. Panel C reports summary statistics for our extended data over the period 2003 through June 2020. The degree of non-normality is about the same as in the earlier sample, but the average alpha drops to −0.01%, consistent with findings reported in Sullivan (2021) and elsewhere. The decline in aggregate performance in commercial database funds motivates our interest in persistence at the fund level.

1.2. Methodology

1.2.1. Bootstrap

KNT measure the alpha of fund i as the intercept $\hat{\alpha}^i$ (or its t-statistic $\hat{t}_{\hat{\alpha}^i}$) from the FH7 model

$$r_t^i = \hat{\alpha}^i + \sum_{k=1}^{7} \hat{\beta}_k^i F_{k,t} + \hat{\epsilon}_t^i,$$

where r_t^i is a fund's monthly excess returns, $F_{k,t}$ are the seven Fung and Hsieh (2004) factors, $\hat{\beta}_k^i$ are the estimated fund-level loadings on these factors, and $\hat{\epsilon}_t^i$ is the residual. However, due to short samples, these measures of alpha are prone to substantial estimation error. Furthermore, the *t*-statistics may not correctly capture the non-normality, heteroskedasticity and serial correlation of the residuals as seen in Table 1.

To deal with these estimation issues, we employ KNT's bootstrap approach. First, we store the residuals $\hat{\epsilon}_t^i$ from each fund-level benchmark regression. These residuals have zero alpha by construction, and thus they represent abnormal returns under the null hypothesis of zero alpha. For each bootstrap replication $b = 1, \dots, B$ (with $B = 1{,}000$) of the T_i returns of fund i, we generate random month indices $s_1^b, \dots, s_{T_i}^b$, where each s_t^b takes a value from 1 to T_i, with replacement. These month indices are then used to construct resampled fund returns

$$r_{i,t}^b = \sum_{k=1}^{K} \hat{\beta}_k^i F_{k,t} + \hat{\epsilon}_t^i, \quad t = s_1^b, \dots, s_{T_i}^b,$$

which again have zero alpha by construction. We then re-run the benchmark regressions on these resampled returns to obtain bootstrapped alphas and their *t*-statistics.

For each bootstrap replication $b = 1, \dots, B$ we generate a cross-section of fund-level alphas $\hat{\alpha}_i^b$ and *t*-statistics $\hat{t}_{\hat{\alpha}_i}^b$ across the $i = 1, \dots, N$ funds. Given a desired performance quantile (e.g., median), we calculate the quantile in each of the *B* cross-sections, which gives the distribution of the performance quantile under null hypothesis. We can then compare this null distribution against the performance quantile calculated from the observed values $\hat{\alpha}^i$ or $\hat{t}_{\hat{\alpha}^i}$. The bootstrapped *p*-value is the probability that the null-distributed quantile is above the observed quantile.

Figure 1 compares the density of observed alpha *t*-statistics and the respective bootstrapped null distribution. Panel A reproduces Figure 2 from KNT. Note the density under the null is symmetric around zero, whereas the observed distribution is skewed with significant mass to the right of the null distribution, suggesting that top managers are skilled. Our replication in Panel B is quite similar, whereas in Panel C the right tail is greatly diminished during the 2003–2020 extension, consistent with declining levels of abnormal performance.

1.2.2. Estimating Bayesian posterior alphas

KNT show that hedge fund performance prediction can be improved with the Bayesian estimation methodology of Pástor and Stambaugh (2002). In short, KNT combine FH7 benchmark regressions with information on non-benchmark assets to produce improved estimates of FH7 alpha. For each individual fund, they use two non-benchmark assets: the HFR style index corresponding to the fund's self-reported broad style and the HFR fund-of-fund index. When studying a fund-of-fund, only the fund-of-fund style index benchmark is used. Because the non-benchmark assets are correlated with fund returns, and have longer histories, their correlation structure with benchmark assets can be used to improve alpha estimates.

Like KNT, we estimate Bayesian posterior alphas using the most recent 24 returns. A detailed description of the method is in the appendix.

1.3. Results

1.3.1. Bootstrap

Table 2 shows the bootstrap tests at various performance percentiles. The KNT results are reproduced in Panel A and our 1994–2002 replication is in Panel B. In each panel, the first and second rows report OLS FH7 alpha in percent per month and its bootstrapped *p*-value. The third and fourth rows report the *t*-statistic of alpha based on heteroscedasticity and autocorrelation consistent standard errors and its bootstrapped *p*-value. In all cases the match between our results and those in KNT is extremely tight. At the 90th percentile, for example, KNT report an OLS alpha of 1.24% and a *t*-statistic of 3.78 while we find an OLS alpha of 1.36% and an identical *t*-statistic of 3.78. When measured by alpha *t*-statistic, in both Panels A and B all bottom percentiles have insignificant performance, and all top percentiles have significant performance. The cross-sectional distributions are slightly more diffuse in our replication, which likely reflects our use of six databases instead of the three in KNT, allowing us to capture more cross-sectional variation. As reported in Panel C for the 2003–2020 extension, the level of performance has shifted substantially to the left. The OLS alpha at the 90th percentile, for example, is reduced from 1.36% to 0.67% but remains significant, suggesting that performance prediction might still be worthwhile despite the overall decline in hedge fund performance.

KNT assess the impact on the distribution of fund-level performance of four procedures used to mitigate potential biases: dropping the first 12 monthly returns of each fund to control for backfill; de-smoothing following the procedure in Getmansky et al. (2004); allowing for structural breaks in factor loadings identified by the Bai and Perron (1998) test; and a bootstrap that allows for cross-sectional dependence in residuals. Table 3 shows the 10th and 90th percentiles of alpha and the *t*-statistic of alpha after applying these data treatments. Panel A reproduces results listed in KNT, whereas Panels B and C show our replications using the same time period as KNT and the more recent extension, respectively. These sensitivity analyses do not affect the results: in all cases the top decile alpha *t*-statistics remain significant.

In unreported analyses, we also replicate quite well the distribution of fund-level performance for the top and bottom five funds as well as style subsets as studied in KNT. These results provide evidence that our sample replicates the salient properties of the sample constructed by KNT. We turn next to the central analysis of performance persistence.

1.3.2. Persistence

KNT test for performance persistence following the procedure in Carhart (1997). Funds are sorted annually at the beginning of each year based on performance measured over the prior 24 months. Portfolios of the sorted funds are equally weighted monthly and are reconstituted annually. The performance of each portfolio is then estimated once over the concatenated annual holding periods.[4]

Table 4 shows results for annual sorts based on the OLS FH7 alpha in Panel A and its corresponding *t*-statistic in Panel B. Panel A1 reproduces results reported in KNT. Both the top and bottom deciles generate annualized holding period alpha of about 5% and the difference is not significant. This result indicates that the OLS alpha is too noisy to distinguish skilled from unskilled managers. Indeed, portfolios formed from the top and bottom 1% of funds feature dramatic reversal, with the top portfolio generated an FH alpha of −0.95% per year versus the bottom portfolio's 8.24% per year.[5] Our replication in Panel A2 features a similar reversal for the top and bottom 1%. In both Panel A1 and A2, note that the sort on alpha results in portfolios with

[4] For the portfolio formed on January 1, 2020, the holding period is only six months, as our data end in June 2020.

[5] This result is consistent with simulation evidence in Carpenter and Lynch (1999) regarding spurious reversals when survivorship is related to multi-period performance.

volatility that rises the further one moves from the median portfolio of funds because of the noise in OLS alpha as a sorting variable.

When sorting on the *t*-statistic of OLS alpha in Panels B1 and B2, volatility is smallest for the top portfolio and increases monotonically for the others, because *t*-statistics are larger for those funds with less residual risk. More importantly, there is greater evidence of persistence for the top funds. As shown in Panel B1, for example, KNT find the top decile generates a holding period FH alpha of 6.05% versus 4.19% for the bottom decile, though the difference is not statistically significant. Our replication in Panel B2 features a slightly lower holding period alpha, but we find the same qualitative result that differences across the top and bottom decile portfolios are insignificant.

Table 5 shows results for annual sorts based on either the Bayesian posterior FH7 alpha or its corresponding *t*-statistic. Panel A1 reproduces the results reported in KNT for the sort on alpha. In stark contrast to the OLS sorts, here there is robust evidence of performance persistence. Holding period alpha is 8.21% for the top decile in Panel A1, and it declines monotonically to 2.40% for the bottom decile. The difference between the top and bottom deciles is significant at the 1% level. Our replication in Panel A2 does not perform as well. In particular, the holding period alpha of our top decile is lower than that of decile 2 and 3, and the spread between our top and bottom deciles is insignificant. One possible explanation for the difference between our results and those in KNT is that our sample contains higher cross-sectional variation, as noted in Table 2, and this results in higher volatility for the top and bottom quantiles than in KNT as shown in Panel A2.

When the sort is based on the *t*-statistic of the Bayesian posterior alpha, however, we are able to closely replicate KNT's results as shown in Panels B1 and B2. KNT report a highly significant holding period alpha of 6.56% for the top decile and a significant spread of 3.15%

across the top and bottom deciles. Similarly, we find a highly significant alpha of 6.34% for the top decile and a significant spread of 4.38% across the top and bottom deciles.

We investigate the robustness of these results to more recent observations of hedge fund performance, as well as more advanced data treatments, later in the paper. Next, we turn attention to replication of the persistence tests in JMN.

2. Replication of Jagannathan, Malakhov, and Novikov (2010)

2.1. Data

Using HFR data, JMN present persistence tests over the May 1996–April 2005 period. Given that their sample ends in April 2005 and their working paper was first released in February 2006, they most likely used a mid-to-late 2005 HFR snapshot. To replicate their results, we use the November 2007 HFR snapshot, which is the version we have that is closest to that used by JMN.

Following JMN, we retain only funds with USD-denominated returns reported net of all fees and adjust for backfill bias by retaining fund returns starting from the date the fund was added to the database. The earliest addition date, May 1st, 1996, therefore, defines the start of the sample and explains the unique convention in JMN of measuring annual returns from May through April.

JMN use 32 HFR style indices. The mapping from a fund's self-reported style to a corresponding style index is nearly unambiguous. However, roughly 800 funds in the November 2007 snapshot that survive the data filters described above use the CTA-like fund styles "Managed Futures" and "Foreign Exchange," neither of which have close counterparts in the list of HFR style indices. We test the hypothesis that JMN simply dropped all funds that use these two fund styles by doing the same and comparing resulting fund counts and summary statistics of returns. Table 6 reproduces the number of funds year-by-year, and cross-sectional average and median monthly return, as well as standard deviation, reported in JMN alongside our replication. The number of funds is almost identical across the two samples, and the summary statistics are typically within a few basis points.

2.2. Methodology

JMN measure hedge fund performance by "relative alpha" against a three factor benchmark $X_t = \left[R_t^{mkt}, \eta_t^{J,self}, \eta_t^{K,aux}\right]'$ consisting of the CRSP VW equity market excess return (R_t^{mkt}), the unsmoothed excess return on fund's self-reported style index ($\eta_t^{J,self}$), and the unsmoothed excess return on an auxiliary style index ($\eta_t^{K,aux}$) that is chosen via model selection. Each hedge fund style index I_t^J is assumed to follow an MA(2) process

$$I_t^J = \gamma_0^J \eta_t^J + \gamma_1^J \eta_{t-1}^J + \gamma_2^J \eta_{t-2}^J$$

with identification constraint $\gamma_0^J + \gamma_1^J + \gamma_2^J = 1$, from which the unsmoothed excess returns η_t^J are extracted as in Getmansky et al. (2004).[6] Similarly, the reported fund-level excess returns $r_{i,t}$ are assumed to follow an MA(s) process

$$r_{i,t} = \sum_{k=0}^{s} \theta_k^i r_{i,t-k}^{un}$$

where $r_{i,t}^{un}$ are the unobserved, true fund excess returns, and θ_k^i are the fund-specific smoothing parameters that sum to one. Combining this MA(s) process with the benchmark regression

$$r_{i,t}^{un} = \alpha_i + X_t \beta_i + \epsilon_{i,t}$$

[6] JMN somewhat confusingly claim using the benchmark $X_t = \left[R_t^{mkt}, I_t^{J,self}, I_t^{K,aux}\right]'$, but later reveal that they employ the unsmoothed indices η_t^J as the actual factors.

yields

$$r_{i,t} = \alpha_i + \sum_{k=0}^{s} X_{t-k}\delta_{k,i} + u_{i,t},$$

where $\delta_{k,i} = \theta_k^i \beta_i$ are the betas under smoothing and $u_{i,t} = \sum_{k=0}^{s} \theta_k^i \epsilon_{i,t-k}$ are the MA(s) errors.[7]

Following JMN, we estimate this benchmark regression on 72-month windows $t = T, \dots, T+71$ consisting of 36 monthly evaluation-period returns followed by 36 monthly prediction-period returns. For each fund and 72-month window, the auxiliary style index $\eta_t^{K,aux}$ and number of lag terms s are chosen to minimize the Schwarz (1978) Bayesian information criterion. We allow for two lag terms and eliminate the auxiliary style index if the model fit is improved without it. On these 72-month windows, the benchmark regression is also modified to allow for different intercepts on the two 36-month periods

$$r_{i,t} = \alpha_{zi} + \sum_{k=0}^{s} X_{t-k}\delta_{k,i} + u_{i,t},$$

where $z = 0$ for $t < T + 36$ and $z = 1$ otherwise. This yields the evaluation-period alpha α_{0i} and prediction-period alpha α_{1i}.[8] Tests of persistence are based on the cross-sectional relationship between α_{0i} and α_{1i}. In the original May 1996–April 2005 sample of JMN, this constitutes four overlapping 72-month subsamples: May 1996–April 2002, May 1997–April 2003, May 1998–April 2004, and May 1999–April 2005. Table 7 shows that the adjusted-R^2 values are similar between JMN's reported values and our baseline replication.

[7] In theory the $\delta_{k,i}$ should be linearly dependent, but we do not model this dependency.

[8] In this 72-month regression, based on Equation (7) in JMN, betas are assumed constant across the evaluation and prediction periods. However, latter parts of JMN, such as comparisons of adjusted- R^2 across the two periods, suggest that they may have used two separate 36-month regressions instead. In unreported analysis, we found that the single 72-month regression replicates JMN's results slightly better, so we use it throughout.

In studies of performance, researchers typically must deal with missing return observations. This is not simply an issue of econometric estimation, because there might be information revealed by a manager's endogenous decision to neglect to report a return, or to stop reporting all together. In the estimation stage, we drop funds from the analysis if any returns are missing in the evaluation period. If returns are only missing from the prediction period, we run the benchmark regression on the 36 evaluation-period returns only and set the prediction period alpha to missing. In selecting the optimal auxiliary style index, we ignore indices with missing returns over the estimation window.

2.3. Results

2.3.1. Regression tests for persistence

JMN's first analysis of performance persistence involves OLS cross-sectional regressions of prediction period alpha on evaluation period alpha in each of their four 72-month subsamples. Table 8 shows our replication of these regressions. Panels A and B indicate that our inference is qualitatively the same as JMN in only two of the four subsamples, the first and the last, for which we both find positive and significant slope coefficients, i.e., prediction period alpha is positively related to evaluation period alpha. In the first subsample, though, the slope in JMN is more than double ours, whereas, in the last subsample, our slope is roughly 50% larger than reported in JMN. As JMN note, the prediction period alpha α_1 is prone to outliers, so they rerun the analysis after winsorizing α_1 at 1% and 99% levels. Panels C and D show replication is improved with the winsorized prediction period alphas. Our slope coefficients are almost identical to those in JMN for three of the four subsamples. The main difference between our results and those of JMN is that we find negative persistence in the 1997–2003 cross-section (t = −2.57), whereas JMN finds no persistence (t = −0.18) in this cross-section.

Table 9 shows the corresponding analysis using weighted least squares (WLS), where both α_0 and α_1 are divided by the standard error of α_0 to produce *t*-statistic $t_{\alpha_0} = \frac{\alpha_0}{\sigma_{\alpha_0}}$ and a stylized *t*-statistic $t^*_{\alpha_1} = \frac{\alpha_1}{\sigma_{\alpha_0}}$. Here the problem with the 1997–2003 cross-section vanishes, and we are able to replicate JMN very well, especially when winsorizing $t^*_{\alpha_1}$ in Panels C and D. In both JMN and our replication, there is a positive and significant relationship between estimation period and prediction period alphas in all four subsamples, and our slope coefficients are reasonably close to theirs in magnitude.

In subsequent analysis, JMN control for missing prediction-period alphas due to funds having stopped reporting during the prediction period by modeling the probability of the two main reasons for stopping reporting: liquidation and voluntary non-reporting due to fund being closed to new investors. JMN estimate parameters of the model via GMM and find that evidence of persistence is somewhat weaker, wherein the slope coefficient on prediction period alphas is positive in all four subsamples but statistically significant in only two. In our unreported attempt at replicating the GMM model, we found no evidence of persistence. The non-replicability is likely due to the sensitivity of parameter estimates in the rather complicated GMM model.

2.3.2. Portfolio tests of persistence

JMN conduct another test of performance persistence by sorting funds into equally-weighted portfolios by the relative alpha *t*-statistic (t_{α_0}). They then measure the subsequent performance of the sorted portfolios and present the evaluation and prediction period portfolio performance as FH7 alphas. If a fund stops reporting during the out-of-sample prediction period, the portfolio is rebalanced assuming returned capital is distributed equally across the remaining funds. JMN do not state whether they rebalance fund weights monthly or use buy-and-hold weights. For simplicity, we use monthly rebalancing as in the KNT analysis.

Following JMN, Table 10 shows tertile portfolio sorts, where 33% of funds enter each of Inferior, Neutral, and Superior portfolios. Panel A reproduces the results reported in JMN. The Superior portfolios deliver significant positive alpha in each of the four evaluation periods, and significant positive alpha in three of the four out-of-sample holding periods.[9] Panel B lists the results of our replication using the 2007 download of the HFR database. We find the same qualitative result, with the Superior portfolio delivering positive significant alpha in the same three out-of-sample periods as JMN, with monthly alpha within a few basis points.

In summary, using the 2007 HFR database, we are generally able to replicate the central results in JMN, and find robust evidence of persistence for three of the four subsamples using both cross-sectional regressions of alpha on lagged alpha as well as the performance of sorted portfolios out-of-sample.

3. Horse Race

3.1 Preliminaries

KNT and JMN follow different approaches in their out-of-sample analyses, so it is not straightforward to execute a "horse race" between these two measures. We opt to frame the comparison between KNT and JMN using the standard Carhart (1997) methodology with non-overlapping one-year holding periods. KNT use this procedure whereas JMN's portfolio sorts utilize overlapping three-year holding periods. In addition, JMN has no minimum AUM filter for

[9] As a caveat, note that a substantial number of funds stop reporting during the holding period, even for the Superior portfolios. If one applies an ad-hoc return penalty to reflect possible censoring of poor performance, as in Titman and Tiu (2011), the out-of-sample performance of all portfolios will drop.

their hedge fund sample and use non-standard May rebalancing dates. As a preparation to a side-by-side comparison of the predictive power of the two measures, we study the sensitivity of JMN's results to varying these three methodological differences. We continue to use the original JMN approach to estimate the relative alpha *t*-statistic as the sorting variable for portfolio construction but use the KNT set of quantiles (e.g., deciles as opposed to the tertiles used in JMN).

Table 11 reports the results. The first specification repeats the JMN analysis with no size threshold and May formation dates, but a one-year as opposed to a three-year holding period. The top decile generates a significant FH7 annual alpha of 3.84% versus the 6.34% reported in Table 5, Panel B2, which uses the KNT Bayesian alpha *t*-statistic as the sorting variable. The spread between the top and bottom deciles is an insignificant 3.12% versus a significant 4.38% in Table 5 using the KNT approach.

To test for sensitivity to fund size, the next specification adds an AUM threshold of 20 million USD as in KNT, applied at formation time. The top decile performance is improved slightly to a significant 4.52%. Note that the bottom decile also features a significant alpha of roughly the same magnitude, though this result disappears with all other specifications when using a January sort, indicating a lack of robustness in the JMN procedure. All other specifications maintain the AUM filter. The third specification forms the portfolios at the beginning of each January, as in KNT, instead of May, for an out-of-sample period of January 2000 through December 2005. The top decile performance is unaffected but the spread between the top and bottom deciles is now a significant 3.76%, similar to the KNT results in Table 5.

The fourth and the fifth specifications in Table 11 continue to use the AUM threshold and January-formation as above but utilize different datasets: the 2020 HFR snapshot and the aggregate database used in our KNT replication, respectively.[10] In both cases the performance of the top decile is again significant and similar in magnitude to the other specifications. The 2020 HFR snapshot yields a weakly significant 10% alpha spread of 1.90%, while the aggregate database improves to a statistically significant 3.31%. This result is intuitive: the more comprehensive a dataset the more likely it is to include a substantial fraction of skilled managers.

We conclude that the predictive power of the JMN relative alpha *t*-statistic is maintained when focusing on the top decile and using a standard one-year holding period, an AUM filter, and a January formation date.[11] That said, the magnitude of the top decile alpha is noticeably smaller than that generated by the KNT Bayesian alpha, foreshadowing the results of the horse race we turn to next.

3.2. Horse race

Table 12 shows the persistence results over the post-publication periods for both fund-weighted and asset-weighted buy-and-hold portfolios. We use the original KNT and JMN approaches to estimate predictive alpha *t*-statistic measures. We use one-year holding periods, rebalance portfolios each January, and apply a 20 million USD AUM threshold for fund inclusion. The performance of each portfolio is then estimated once over the concatenated annual holding periods.

Panel A lists the results for the KNT sort on the Bayesian posterior FH7 alpha *t*-statistic. With fund-weighting, the top decile generates a significant FH7 alpha of 2.85%, whereas with

[10] For the aggregate database, we can only utilize funds that report their HFR-specific narrow style to the 2020 HFR snapshot. However, we have checked that the results are qualitatively unchanged if we add the non-HFR-reporting funds and use five coarse styles.

[11] The spread between the top and bottom deciles is highly sensitive to the various specifications we use, but our focus is generally the performance of the top decile as this is a more tangible measure of the benefit to investors.

asset-weighting, the top decile's alpha is insignificant, suggesting that smaller funds are generating much of the portfolio alpha. The top percentile generates an even larger and more significant alpha. Spreads between the FH7 alpha of the top and bottom deciles are highly significant and are 5.19% and 3.93% for the two weighting schemes, respectively. Thus, over the post-publication sample of 2003–2020, the KNT sort continues to display predictive power, though the holding period alpha of the top decile is less than half the performance in the earlier period as reported in Table 5.

Panel B lists the results for the JMN sort on the relative alpha *t*-statistic. Here, neither weighting scheme produces significant top decile portfolio alpha. The top percentile portfolios, however, do feature a significant alpha almost as large as those using the KNT predictor. The top-bottom decile alpha spreads are significant at just over 2% for the two weighting schemes yet fall short of the respective spreads produced by KNT Bayesian alpha *t*-statistic. Of the two models, the KNT Bayesian alpha *t*-statistic appears to be the superior performance predictor in the post-publication samples. The JMN approach does generate holding period alpha for the top percentile, but this quantile would likely be less feasible for an investor to access.

3.3. Why does KNT outperform JMN?

We next explore and compare the drivers of the predictive power of the KNT Bayesian alpha and JMN relative alpha. Table 13 shows how the results from Table 12 are affected when we modify elements of KNT and JMN to pinpoint the reasons for the superior ability of KNT to predict hedge fund performance.

Panel A presents the holding period performance of portfolios formed by variations of the KNT methodology. First, under "Table 12," we reproduce the main results from Table 12. Second, under "Short window," we use a rolling 36-month window for estimating the relation between benchmark and non-benchmark assets, as opposed to the original expanding window. The results are essentially unchanged, suggesting that the Pàstor and Stambaugh (2002) motivation for using the longer time series available for non-benchmark assets is not the source of the incremental predictive power. Third, under “FH7 benchmark,” we remove the non-benchmark assets completely, thereby eliminating the necessity of a Bayesian analysis, and simply use an OLS regression with the FH7 benchmark to measure alpha. The results are essentially unchanged. The top decile alpha is 2.80%, for example, compared to the 2.85% generated by the full KNT model in Table 12. The value added by the Bayesian methodology is not as clear as originally reported by KNT, suggesting that the relative importance of the non-benchmark assets has weakened over time.

Panel B presents the holding period performance of hedge fund portfolios sorted by variations of the JMN relative alpha. First, under “Table 12,” we reproduce the main results from Table 12. Second, under “No lags,” we calculate the JMN relative alpha without the lag structure in the factor model that JMN use to incorporate serial correlation due to asset illiquidity. Surprisingly, this substantially improves the performance of the spread portfolios: the difference in alpha between the top and bottom deciles increases from 2.13% in Table 12 to 3.22%. This result suggests that JMN’s lag structure improves in-sample fit at the expense of out-of-sample predictability. Third, under “FH7 benchmark,” we maintain the lack of lag structure and replace the equity factor in JMN’s factor model with the full FH7 benchmark. This improves the performance of the spread portfolios even further by better identifying the worst funds.

Taken together, this analysis indicates that the JMN relative alpha suffers from overfitting relative to the KNT methodology. None of the variants of the JMN measure can generate a top decile with a significant holding period alpha, whereas all variants of the KNT measure feature a significant top decile alpha of about 2.80%. The KNT use of non-benchmark assets in a Bayesian setting offers some incremental improvement by better identifying the worst performers. However, most of the predictability would be achieved via a simple OLS regression on the FH7 benchmark.[12] This result suggests that the improvement in predictability reported by KNT may be a feature only of the hedge fund landscape in their early sample, as it does not hold using our much longer time series.

4. Real-Time Extensions

4.1. Real-time data

Our replications reported thus far assume that investors can observe a reported return at the end of the corresponding month. For example, we assume the return for a given fund in December 2001 was observed by the investor in time to estimate performance, sort funds into deciles, and invest at the beginning of January 2002. In contrast, as described below, our extensions in this section explicitly model realistic frictions faced by investors, such as reporting delays, hence our use of the term "real-time" data. We remove funds that are closed to new investment, so our tests provide implementable guidance for investors.[13]

We treat the data two ways to generate a sample that more accurately reflects the information available to an investor at a portfolio formation date as well as the investor's opportunity set. First, we address the backfill issue. Consider a fund manager who started reporting to a database in January 2001 and uploaded a three-year track record beginning in January 1998.

[12] We also explored sensitivity to many other minor implementation details. For KNT Bayesian alpha, for example, we tried different levels mispricing uncertainty. For JMN relative alpha, we tested disabling the style index level unsmoothing, as well as using the most recent 24 returns in the estimation (instead of the full 36 months of returns). Results were generally not sensitive to these details, so we omit them for brevity.

[13] We require a fund to have at least one open share class at portfolio formation, determined using the exact closing and reopening dates where available, and otherwise the most recent open-for-investment dummy. This rule is conservative as it eliminates share classes with no information on closing status, and for closed funds with missing closing date assumes the fund to have been closed since inception.

Clearly, investors who rely on the databases were not aware of the fund prior to January 2001. Thus, we only allow the investor to hold funds following the initiation of reporting, in the spirit of Jorion and Schwarz (2019). Once a fund starts reporting, we do allow the investor to use older observations for the purpose of performance assessment for the KNT measure, but not for the JMN measure, in line with their respective methodologies.[14] Second, to allow for reporting delays by funds as well as the time it takes for databases to be updated and analyzed by investors, we assume the investor observes returns with a three-month lag at each portfolio formation date.

In portfolio sorts involving real-time data, when a fund stops reporting during the holding period, we assume that the proceeds are reinvested at a risk-free rate like Edelman et al. (2013) and Joenväärä et al. (2021).[15] Titman and Tiu (2011) suggest applying a delisting adjustment if a fund stops reporting during a holding period to reverse the censoring of unreported poor returns of failing funds. However, our focus is the performance of top-decile funds which typically do not stop reporting because of pending failure; top-decile funds are far more likely to cease reporting to data vendors because they are closed for new investment.

4.2. Impact of real-time data corrections

Table 14 demonstrates how the predictive power of the KNT and JMN measures is impacted by the introduction of the information frictions discussed in Section 4.1. Columns marked “Real-time” incorporate the two data treatments discussed in Section 4.1: an exact backfill

[14] In effect this means that the formation-time backfill-adjustment only affects of KNT results, not the JMN results, as the latter are implicitly backfill-adjusted during the JMN estimation stage.

[15] Joenväärä et al. (2021) show that best-performing top and worst-performing bottom decile funds are likelier to drop from databases. However, to ensure bottom-decile funds are more likely to cease reporting, as highlighted by Liang and Park (2010), do not drive our results; our unreported analysis finds that our results are robust to the Titman and Tiu (2011) adjustment.

correction and a three-month lag between an evaluation period and holding period. Columns marked "As-reported" do not incorporate these treatments and are from Table 12.

Panel A shows the results for KNT Bayesian alpha t-statistic. The real-time data treatments have little effect on the top quantiles (first percentile through first decile), whose alphas all remain statistically significant. However, the untreated data allow for better detection of the worse funds in the bottom quantiles. As a result, the top-bottom decile alpha spread decreases from 5.19% to 4.24%, while remaining highly significant.

Panel B shows the results when sorting by JMN relative alpha t-statistic. As in Panel A, the real-time data treatments weaken the ability to detect worse funds, which reduces the top-bottom decile alpha spread from a significant 2.13% to a weakly significant 1.54% (t = 1.81). The respective percentile spread reduces more dramatically from 4.76% to 2.20% (t = 1.73). The treatments improve the top decile portfolio alpha to a significant 1.68%, which however remains below the 2.96% alpha from the KNT sort in Panel A. As in Section 3, of the two alpha t-statistics, the KNT measure remains the superior predictor.

To illustrate how persistence has changed throughout 1994−2020, Figure 2 plots rolling 5-year FH7 alphas (Panel A) and their t-statistics (Panel B) for top and bottom deciles and their spread, when using KNT's Bayesian alpha t-statistic to form portfolios, with real-time data treatments. The top and bottom deciles' alphas generally decrease over time, though top decile FH7 alphas remain consistently positive. The decrease is steeper for the bottom decile, especially from 2011 onwards, resulting in a dramatic increase in spread between the deciles. These results indicate the decline in aggregate hedge fund performance documented in Bollen et al. (2021), and elsewhere. The figure further confirms that the Bayesian alpha t-statistic continues to offer an informative signal throughout the sample.

A more exacting approach to deal with data biases, including the possibility of data revisions and the elimination of poorly performing funds from the databases over time, is to work with an archive of downloads. In this approach, at each portfolio formation date we only use data from the respective December downloads to estimate past performance. This captures the true reporting delays present at each year-end snapshot, as well as eliminates the effect of later backfilling and data revisions. As shown in the appendix, our simpler real-time approach generates results which are comparable to those obtained when using the archived snapshots. We conclude, therefore, that an investor can approximate actual performance reasonably well with the most recent downloads of the databases, so long as an exact backfill correction and an evaluation lag are applied to the data.

4.3. Long-term persistence

Most hedge fund studies examine persistence over relatively short horizons, such as the annual holding periods used in the bulk of our analysis. A typical investor's holding period is longer than a year, due to lock-up restrictions and the time it takes to assess performance and decide whether to reallocate to other managers (Joenväärä et al. 2019). To address this concern, in addition to the one-year holding period, in this section we also form real-time portfolios by sorting funds every second or third year in January to test whether performance persists over longer horizons. The performance of each portfolio is then estimated once over the concatenated annual, two-year, and three-year holding periods.

We conduct the portfolio sorts using KNT's Bayesian posterior FH7 alpha *t*-statistic, as it generated the most robust evidence of persistence in the previous sections. Panel A of Table 15 shows the persistence results over the full 1994–2020 period for fund-weighted buy-and-hold portfolios when we use one-year, two-year and three-year holding periods. Only the extreme percentiles and top decile generate positive and significant holding period alpha for all three holding periods. Top decile FH7 alpha declines from 3.21% for the one-year holding period to 2.36% for the three-year holding period. Spreads between the top and bottom decile FH7 alpha are significant for all holding periods but decline from 3.90% per year for the one-year holding period to 2.93% for the three-year holding period. Thus, though there is a notable decline in the magnitude of performance over the longer holding periods, the top decile and spread portfolios continue to generate positive and significant alpha.

Panel B lists the results for asset-weighted buy-and-hold portfolios when we use one-year, two-year and three-year holding periods. For the top decile, and the spread between the top and bottom deciles, the three-year holding period results are about 0.5% per year lower than when using one-year holding periods. However, they are still economically important indicating that the out-of-sample ability of top managers persists at time-scales relevant for investors. Thus, over our extended sample, KNT's sort identifies a top decile that generates positive and significant alpha and at least marginally significant spread between the top and bottom deciles even when using our real-time data treated to reflect reporting frictions and when using longer two-year and three-year holding periods.

Our results indicate that a real-time investor can exploit persistence even when using relatively low-frequency rebalancing, a concern raised by Joenväärä et al. (2019). This finding is further illustrated in Figure 3 plotting FH7 alphas (Panel A) and their *t*-statistics (Panel B) over time for one to 36-month formation lags. We find post-sort alphas are positive for top decile portfolios and negative for bottom decile portfolios for 20 to 25 months after portfolio formation. The top decile's annualized FH7 alpha stays at about 2.5% for each month of the three-year holding period. This degree of long-term persistence implies that short-term autocorrelation in fund returns does not contaminate our persistence results.

4.4. Fund-of-Funds

Our tests are all based on the hedge fund return data observed in databases as a proxy for the funds available to investors; however, some hedge funds do not report their returns to commercial databases at all, as studied by Agarwal et al. (2013), Edelman et al. (2013), Aiken et al. (2013), and Barth et al. (2023). We address this form of selection bias by studying the performance persistence of the fund-of-funds in our union of six commercial databases. The pioneering work of Fung and Hsieh (2000, 2002) recommends using fund-of-funds data rather than individual hedge funds. They argue that the returns of fund-of-funds are a more accurate representation of the returns earned by hedge fund investors because some fund-of-funds invest in individual funds that do not report to commercial databases. Such funds tend to be closed to new investors because of superior performance and capacity constraints. In addition, the fund-of-funds returns reflect individual funds' returns in liquidation or when redemptions are suspended. Such funds' returns tend to be poor, as shown by Aiken et al. (2015), and are often not reported to commercial databases.

Table 16 reports the real-time performance persistence results for fund-of-funds in the entire 1994–2020 period, with three different holding periods (one-year, two-year, and three-year), and either fund-weighted (Panel A) or value-weighted (Panel B). As with the analysis in Section 4.3, we conduct the portfolio sorts using KNT's Bayesian posterior FH7 alpha *t*-statistic. Table 16 shows that fund-of-funds deliver significant top decile alphas and top-bottom decile alpha spreads in all cases except for the top decile alpha using asset-weighted portfolios with a three-year holding period. This result indicates that fund-of-funds returns are persistent. In the appendix (see Figure A1 and Table A4), we conduct these tests using gross-of-first layer-fee returns. We find that fund-of-funds generate consistently significant top decile alphas and top-bottom decile alpha spreads for both fund-weighted and asset-weighted portfolios even at the three-year holding periods. Although these results do not give direct evidence that fund-of-funds can access individual funds not in commercial databases, they suggest that fund-of-fund performance persists (Fung et al., 2008) even after our "real-time" correction.

4.5. Decreasing returns to scale

As noted previously, prior research shows hedge fund performance declined substantially following the 2008–2009 global financial crisis. Bollen et al. (2021) explore several explanations but cannot distinguish between the impact of growth in aggregate AUM and outsized returns of passive markets fueled by sustained monetary stimulus. To gain additional insight, we conduct tests of decreasing returns to scale.

Prior studies find mixed evidence of a scale effect. Agarwal et al. (2009) and Yin (2016) report evidence of decreasing returns to scale at the fund level. In contrast, Rzakhanov et al. (2019) find no relation between size and performance after controlling for a measure of strategy capital flow, and Forsberg et al. (2022) find no fund-level scale effect after controlling for industry size. As argued by Cao et al. (2021), standard tests do not adequately address the interrelation between managerial skill, prior fund performance, and fund size. If only a subset of managers possesses the skill to reliably deliver alpha, for example, then searching for a relation between lagged size and subsequent performance will be plagued by a great deal of noise. Consequently, we control for managerial skill by incorporating lagged values of a fund's Bayesian alpha *t*-statistic as a proxy, as well as interactions between the *t*-statistic and lagged size.

Table 17 lists the results of regressions of hedge fund performance as measured by FH7 alpha in year t computed as the sum of the intercept and monthly residuals where factor loadings are estimated over years $t - 2$ through $t - 1$. Explanatory variables include lagged size measured as ln(AUM) at the end of year $t - 1$ and lagged performance as measured by the Bayesian t-statistic of alpha measured over the prior two years. "Late" is an indicator variable equal to one for years $t \geq 2008$ and zero otherwise, motivated by the time variation in aggregate hedge fund performance.

Model 1 indicates a positive relation between size and performance which helps offset the intercept of −2.323% per year. Model 2 reflects the extremely large shift in alpha over time, wherein the intercept is 1.943% in the pre-2008 period and nearly 6% lower thereafter. The relation between size and performance likewise weakens post-2008. Model 3 drops the intercept in favor of year fixed effects and includes lagged skill as a predictor. The coefficient on lagged skill is positive and significant, consistent with persistence, and the relation between size and performance disappears, suggesting the importance of controlling for skill when evaluating decreasing returns to scale. Model 4 includes an interaction between lagged skill and size. The coefficient is negative and significant, consistent with the idea that the more skilled the manager, the greater is the deleterious effect of scale, i.e., these funds have more to lose. Model 5 includes a triple interaction between lagged size, skill, and the Late indicator. Decreasing returns to scale was more pronounced earlier in the sample and weakens considerably more recently. For every standard deviation increase in lagged skill (3.92), doubling the fund size would decrease alpha by 46 basis points in the earlier period, but only by 17 basis points in the later period. We interpret this as evidence that the increase in industry size has eroded most of the aggregate available alpha, so that the impact of scale at the fund level is less pronounced.

A final remark on decreasing returns to scale is salient. Models 3 through 5 all show positive and significant coefficients on lagged skill of a magnitude indicating high persistence. This may appear contradictory to our main findings regarding weaker performance of the top decile in the years following publication of JMN and KNT. Note, however, that the regression analysis does not differentiate between persistent good and persistent bad performance. Figure 2 shows that the performance of the bottom decile as ranked by KNT's Bayesian *t*-statistic is much more persistent than that of the top decile.

5. Summary

Replication studies by Choi and Zhao (2021) and Riley (2021) of foundational work in mutual fund performance appraisal successfully reproduce evidence of persistence reported in Carhart (1997) and Kosowski et al. (2006) but find that it is absent in the years following the publication of the earlier papers. Similarly, we can reproduce evidence of persistence in hedge fund performance following the methodologies developed by KNT and JMN but find that performance persistence in hedge funds has weakened in a more recent sample. That said, KNT's Bayesian alpha *t*-statistic continues to provide a useful information signal throughout our extended sample and allows for the formation of a top-decile portfolio that delivers positive abnormal performance.

Our analyses yield some practical guidance for academics and institutional investors. We find that top portfolios of hedge funds sorted on JMN's relative alpha do not feature out-of-sample performance as robust as those sorted by KNT's Bayesian alpha, likely the result of overfitting in-sample. In fact, over the full sample, a sort on simple OLS FH7 alpha performances almost as well the KNT procedure. We show that researchers can rely on the most recent downloads of commercial databases for realistic performance appraisal, so long as they implement the exact backfill of Jorion and Schwarz (2019) and apply a reasonable lag between the end of an evaluation period and the beginning of a holding period to account for reporting delays. We also confirm prior research that finds in the years following the publication of KNT and JMN, aggregate hedge fund performance has declined substantially. This implies that investors can only justify a continued allocation to hedge funds if they monitor and rebalance their holdings regularly.

References

Ackermann C., R. McEnally, and D. Ravenscraft, (1999). The performance of hedge funds: Risk, return and incentives. *Journal of Finance* 54(3), 833–874.

Agarwal V., N.D. Daniel, and N.Y. Naik (2009). Role of managerial incentives and discretion in hedge fund performance. *Journal of Finance* 64(5), 2221–2256.

Agarwal, V., V. Fos, and W. Jiang, (2013). Inferring reporting-related biases in hedge fund databases from hedge fund equity holdings. *Management Science* 59(6), 1271–1289.

Aggarwal R.K., and P. Jorion (2010a). Hidden survivorship bias in hedge fund returns. *Financial Analysts Journal* 66(2), 69–74.

Aggarwal R.K., and P. Jorion (2010b). The performance of emerging hedge funds and managers. *Journal of Financial Economics* 96(2), 238–256.

Akey, P., A. Robertson, and M. Simutin, (2021). Noisy factors. University of Toronto Working Paper.

Aiken A.L., C.P. Clifford, and J.E. Ellis (2013) Out of the dark: Hedge fund reporting biases and commercial databases. *Review of Financial Studies* 26(1), 208–243.

Aiken A.L., C.P. Clifford, and J.E. Ellis (2015) Hedge funds and discretionary liquidity restrictions. *Journal of Financial Economics* 116(1), 197–218.

Ardia, D., L. Barras, P. Gagliardini, and O. Scaillet, (2024). Is it alpha or beta? Decomposing hedge fund returns. *Journal of Financial Economics*, 154, 103805.

Aragon, G., and V. Nanda, (2017). Strategic delays and clustering in hedge fund reported returns. *Journal of Financial and Quantitative Analysis* 52, 1–35.

Bai, J., P. Perron, (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.

Bali, T., S. Brown, and K. Demirtas, (2013). Do hedge funds outperform stocks and bonds? *Management Science* 59, 1887–1903.

Barth, D., J. Joenväärä, M. Kauppila, R. Wermers, (2023). A bias-free assessment of the hedge fund industry: A new evaluation of total assets, alphas, and the flow-performance relation. Office of Financial Research Working Paper 20-01.

Berk J.B. and J. van Binsbergen, (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics* 118(1), 1–20.

Bhardwaj G., G.B. Gorton, and K.G. Rouwenhorst, (2014). Fooling some of the people all of the time: The inefficient performance and persistence of commodity trading advisors. *Review of Financial Studies* 27(11), 3099–3132.

Bollen, N., J. Joenväärä, and M. Kauppila, (2021). Hedge fund performance: End of an era? *Financial Analysts Journal* 77, 109–132.

Brown, S., T. Fraser, and B. Liang (2008). Hedge fund due diligence: A source of alpha in a hedge Fund portfolio strategy. *Journal of Investment Management* 6, 23–33.

Brown, S., W. Goetzmann, and R. Ibbotson, (1999). Offshore hedge funds: Survival and performance 1989–95. *Journal of Business* 72(1), 91–117.

Brown S., Y. Lu, S. Ray, and M. Teo (2018). Sensation seeking and hedge funds. *Journal of Finance* 73, 2871-2914.

Carpenter, J., and A. Lynch (1999). Survivorship bias and attrition effects in measures of performance persistence. *Journal of Financial Economics* 54, 337-374.

Cao, C., T. Simin, and R. Velthuis (2021). Are hedge fund capacity constraints binding? Evidence on scale and competition. Penn State University Working Paper.

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

Choi, J., and K. Zhao, (2021). Carhart (1997) mutual fund performance persistence disappears out of sample. *Critical Finance Review* 10, 263–270.

Dichev, I., and G. Yu, (2011). Higher risk, lower returns: What hedge fund investors really earn. *Journal of Financial Economics* 100, 248–263.

Edelman D., W. Fung, and D. Hsieh, (2013). Exploring uncharted territories of the hedge fund industry: Empirical characteristics of mega hedge fund firms. *Journal of Financial Economics* 109(3), 734–758.

Fama E.F. and K.R. French, (2008). Mutual fund performance, working paper.

Forsberg, D., D. Gallagher, and G. Warren, (2022). Capacity constraints in hedge funds: The relation between fund performance and cohort size. *Financial Analysts Journal* 78, 57−77.

Fung, W. and D. Hsieh, (2000). Performance characteristics of hedge funds and commodity funds: Natural versus spurious biases. *Journal of Financial and Quantitative Analysis* 35(3), 291–307.

Fung, W. and D. Hsieh, (2002). Hedge fund benchmarks: Information content and biases. *Financial Analysts Journal* 58(1), 22–34.

Fung, W. and D. Hsieh, (2004). Hedge fund benchmarks: A risk-based approach, *Financial Analysts Journal,* 65–80.

Fung, W. and D. Hsieh, (2009). Measurement biases in hedge fund performance data: An update. *Financial Analysts Journal* 65(3), 36–40.

Fung, W., D. Hsieh, N. Naik, and T. Ramadorai, (2008). Hedge funds: Performance, risk, and capital formation, *Journal of Finance* 63(4),1777–1803.

Getmansky, M., A.W. Lo, I. Makarov, (2004). An econometric model of serial correlation and illiquidity in hedge fund returns, *Journal of Financial Economics* 74, 529–609.

Glode, V., and R. Green. (2011). Information spillovers and performance persistence for hedge funds. *Journal of Financial Economics* 101, 1-17.

Grinblatt, M., and S. Titman. (1992). The persistence of mutual fund performance. *Journal of Finance* 47, 1977-1984.

Harris R.S., T. Jenkinson, S.N. Kaplan, and R. Stucke, (2022). Has persistence persisted private equity? Evidence from buyout and venture capital funds, Working paper.

Jagannathan, R., A. Malakhov, and D. Novikov, (2010). Do hot hands exist among hedge fund managers? An empirical evaluation. *Journal of Finance* 65, 217–255.

Joenväärä J., M. Kauppila, R. Kosowski, and P. Tolonen (2021). Hedge fund performance: are stylized facts sensitive to which database one uses? *Critical Finance Review*, 10, 271–327.

Joenväärä J., R. Kosowski, and P. Tolonen (2019) The effect of investment constraints on hedge fund investor returns. *Journal of Financial and Quantitative Analysis* 54(4), 1539–1571.

Jorion, P., and C. Schwarz, (2014). The delisting bias in hedge fund databases. *Journal of Alternative Investments* 16(3), 37–47.

Jorion, P., and C. Schwarz, (2019). The fix is in: Properly backing out backfill bias. *Review of Financial Studies* 32, 5048–5099.

Kosowski, R., A. Timmermann, R. Wermers, H. White, (2006). Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance* 61, 2551–2595.

Kosowski, R., N. Naik, M. Teo. 2007. Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *Journal of Financial Economics* 84, 229–264.

Li, H., X. Zhang, and R. Zhao, (2011). Investing in talents: Manager characteristics and hedge fund performances. *Journal of Financial and Quantitative Analysis* 46, 59-82.

Liang, B., (2000). Hedge funds: The living and the dead. *Journal of Financial and Quantitative Analysis* 35(3), 309–325.

McCahery, J., and F. de Roode, (2001). The lost decade for hedge funds: Three threats. In D. Cumming, S. Johan, and G. Wood (Eds.). The Oxford handbook of hedge funds (p. 35-63). Oxford, United Kingdom: Oxford University Press.

McLean, R., and J. Pontiff, (2016). Does Academic Publication Destroy Stock Return Predictability? *Journal of Finance* 71, 5–32.

Pástor, L. and R.F. Stambaugh, (2002). Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* 63(3), 315–349.

Patton, A., T. Ramadorai, M. Streatfield, (2015). Change you can believe in? Hedge fund data revisions. *Journal of Finance* 70, 963–999.

National Association of College and University Business Officers. 2021. 2020 NACUBO-TIAA study of endowments.

Riley, T. (2021). Can mutual fund stars still pick stocks? A replication and extension of Kosowski, Timmerman, Wermers, and White (2006). *Critical Finance Review* 10, 251–261.

Rzakhanov, Z. and G. Jetley, (2019). Competition, scale and hedge fund performance: Evidence from merger arbitrage. *Journal of Economics and Business*, 105: 105841.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.

Sullivan, R. (2021). Hedge fund alpha: Cycle or sunset? *Journal of Alternative Investments* 23, 55–79.

Titman, S., and C. Tiu, (2011). Do the best hedge funds hedge? *Review of Financial Studies* 24, 123–168.

Yin, C. (2016). The optimal size of hedge funds: Conflict between investors and fund managers. *Journal of Finance* 71, 1857−1893.

Figure 1. Bootstrapped vs Actual Alpha *t*-Statistics

Description: Panel A reproduces the density plots reported in Figure 2 of Kosowski et al. (2007); Panel B shows our replication using data from the same time period 1994–2002; Panel C shows results in the later time period January 2003–June 2020. Panels B and C also show the density of actual statistics with Jorion and Schwarz (2019) backfill correction based on listing dates. **Interpretation:** Kosowski et al. (2007) find a right-skewed performance distribution in Panel A as do we in Panel B, but the distribution has shifted to the left in the more recent sample in Panel C.

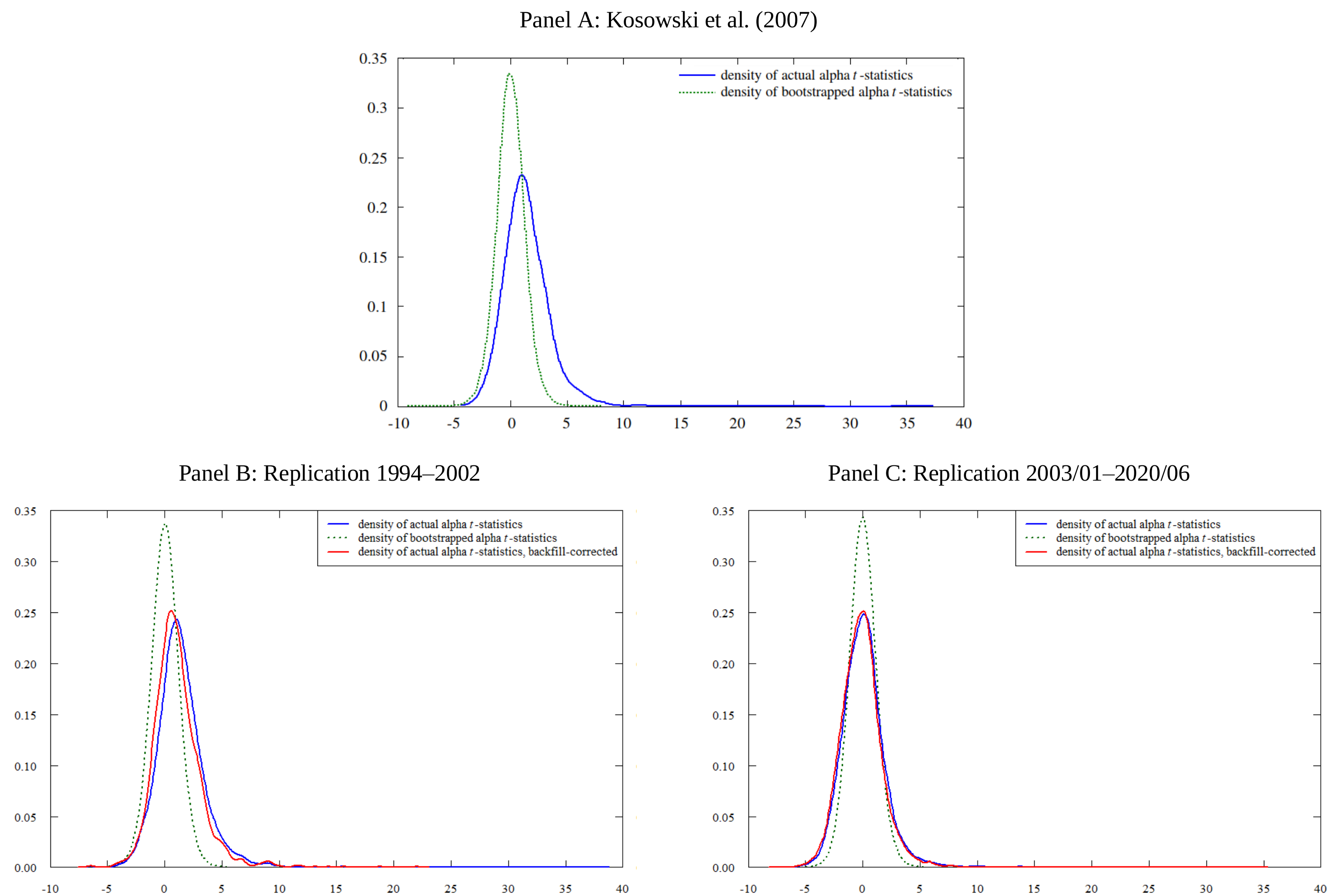
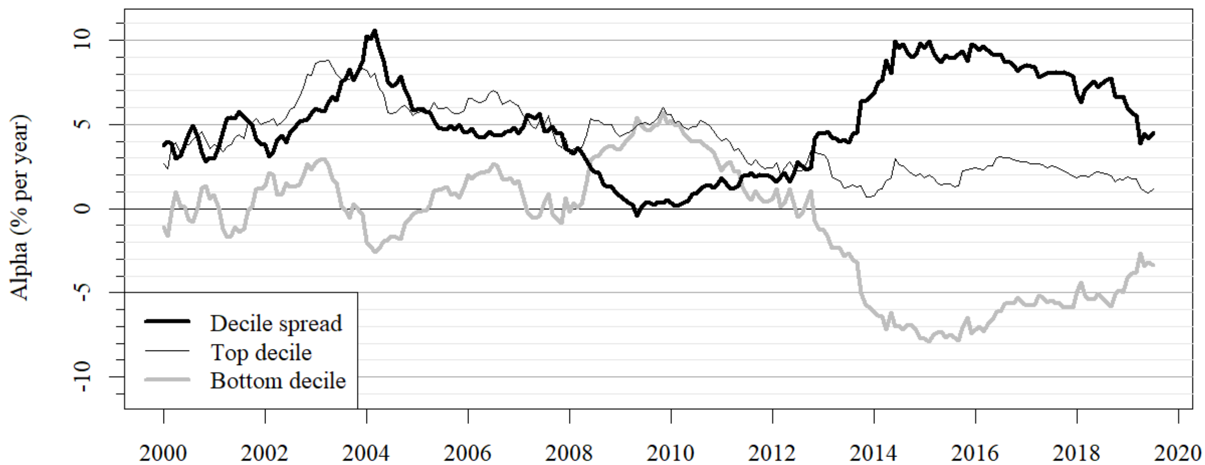


Figure 2. Five-Year Rolling Alphas

Description: five-year rolling seven-factor Fung–Hsieh alphas (Panel A) and their *t*-statistics (Panel B) of portfolios ranked by KNT Bayesian alpha *t*-statistic. The thin black line shows the top decile portfolio; thick grey line the bottom decile portfolio; and thick black line their spread portfolio. Sample period is January 1994–June 2020. **Interpretation:** the spread between top and bottom funds is largely driven by the poor performance of bottom funds in the more recent sample.

Panel A: Rolling alpha estimates



Panel B: Rolling alpha *t*-statistics

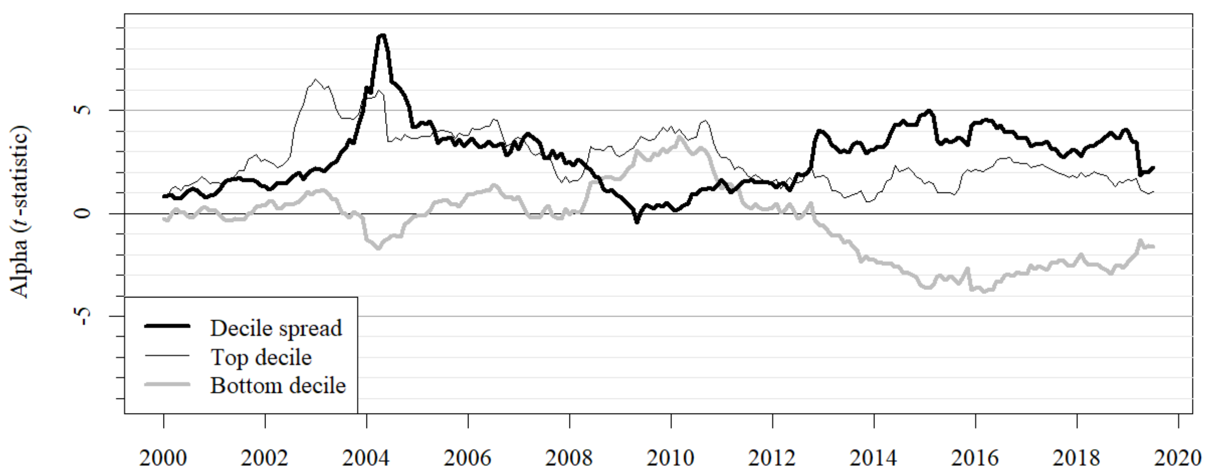
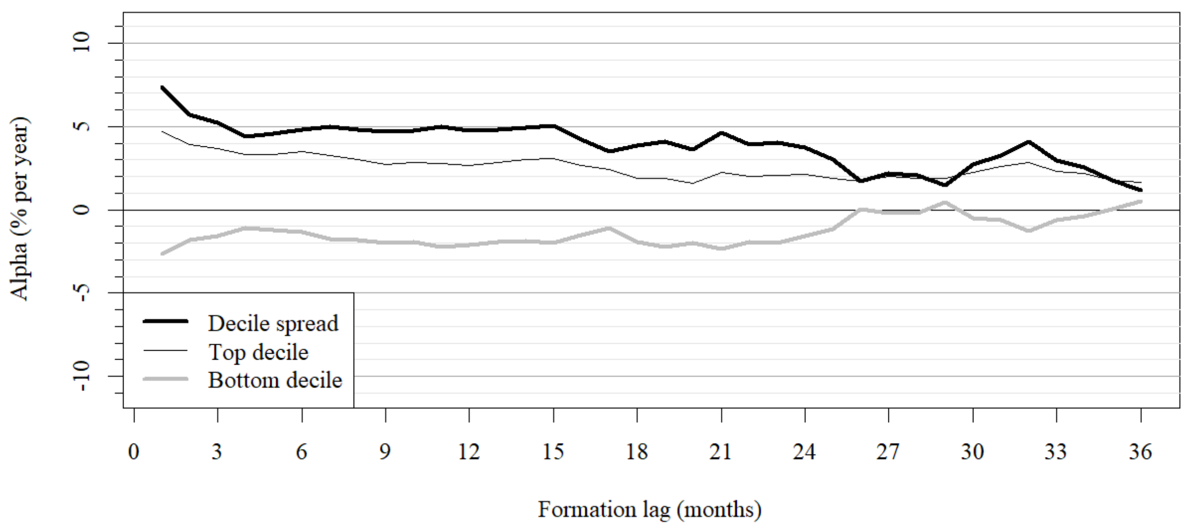


Figure 3. Long-term persistence as a function of formation lag

Description: post-formation Fung-Hsieh alphas (Panel A) and their *t*-statistics (Panel B) of portfolios sorted by *k*-month lagged Bayesian alpha *t*-statistic. The portfolios are formed monthly and held for one month. The fund must be open and database-reporting at month *t* minus *k*. Thin black line shows the top decile portfolio; thick grey line the bottom decile portfolio; and thick black line their spread portfolio. Sample period is January 1994–June 2020. **Interpretation:** the superior performance of top funds persists for up to 36 months post-formation, indicating that the Bayesian alpha *t*-statistic has predictive power for an economically meaningful horizon.

Panel A: Alpha estimates



Panel B: Alpha *t*-statistics

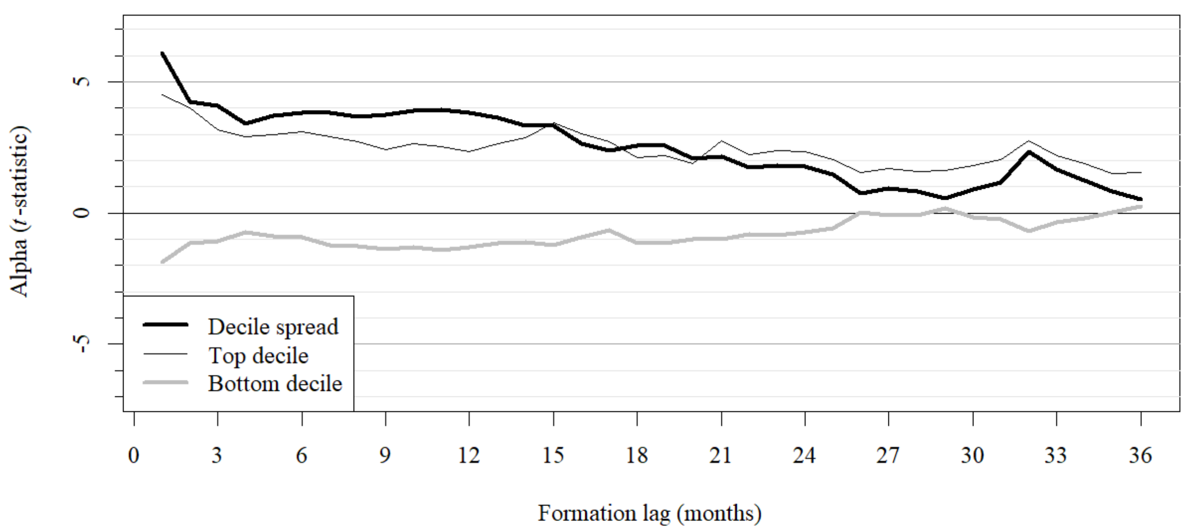


Table 1. Summary Statistics

Description: Panel A reproduces summary statistics of monthly abnormal performance reported in Table 1 of Kosowski et al. (2007); Panel B shows our replication using data from the same time period 1994–2002; Panel C shows results in the later time period January 2003–June 2020. The second through fifth columns show cross-sectional averages of fund-level statistics. The last three columns show percentage of funds with significant non-normality, heteroscedasticity, and serial correlation as measured by the Jarque–Bera, Breusch–Pagan, and Ljung–Box statistics, respectively. **Interpretation:** our replication of Kosowski et al.'s (2007) hedge fund sample has similar performance and descriptive statistics. The more recent sample features worse average performance.

	No. of funds	Alpha (%)	*t*-Alpha	Kurtosis	Skewness	% with J-B $p < 0.1$	% with B-P $p < 0.1$	% with L-B $p < 0.1$
Panel A. Kosowski et al. (2007)								
Long/short equity funds	711	0.51	1.16	3.69	−0.99	69.01	69.01	40.85
Directional trader funds	539	0.41	1.05	3.96	−0.20	51.04	37.50	61.46
Multiprocess funds	278	0.37	1.72	4.25	−0.10	33.33	25.00	16.67
Relative value funds	546	0.42	2.06	5.27	−0.02	22.52	18.02	26.13
Security selection funds	951	0.50	1.14	3.75	0.00	33.96	30.19	33.96
Fund of funds	420	0.27	1.54	4.70	−0.05	39.21	36.56	32.16
All funds	2,734	0.42	1.43	4.29	0.04	38.63	30.03	30.47
Panel B. Replication								
Long/short equity funds	875	0.66	1.26	4.11	0.24	38.06	33.03	27.20
Directional trader funds	424	0.36	0.71	3.92	0.18	29.01	29.72	13.21
Multiprocess funds	271	0.46	2.11	4.60	−0.09	45.39	45.76	33.21
Relative value funds	238	0.39	2.24	6.01	−0.42	47.48	44.54	36.13
Security selection funds	875	0.66	1.26	4.11	0.24	38.06	33.03	27.20
Fund of funds	637	0.29	1.50	4.53	0.00	44.43	47.72	35.16
All funds	2,445	0.46	1.42	4.43	0.07	39.88	38.81	28.38
Panel C. Extension								
Long/short equity funds	5,053	0.01	−0.01	4.03	0.06	35.42	38.27	18.17
Directional trader funds	2,162	−0.05	−0.18	4.08	0.00	32.93	40.29	15.73
Multiprocess funds	1,671	0.11	0.67	5.75	−0.11	48.89	48.11	26.33
Relative value funds	1,833	0.02	0.94	5.62	−0.13	51.50	53.96	30.82
Security selection funds	5,053	0.01	−0.01	4.03	0.06	35.42	38.27	18.17
Fund of funds	3,037	−0.08	−0.12	4.48	−0.34	44.42	58.38	19.89
All funds	13,756	−0.01	0.15	4.56	−0.08	40.80	46.31	20.84

Table 2. Bootstrapped Significance Levels at Different Percentiles

Description: Panel A reproduces performance metrics of the funds at various percentiles reported in Table 2 of Kosowski et al. (2007); Panel B shows our replication using data from the same time period 1994–2002; Panel C shows results in the later time period January 2003–June 2020. In each panel, the first and second rows report the ranked OLS Fung and Hsieh (2004) alpha in percent per month and its bootstrapped *p*-value. The third and fourth rows report the ranked *t*-statistic of alpha based on heteroscedasticity and autocorrelation consistent standard errors and its bootstrapped *p*-value. **Interpretation:** our replication of Kosowski et al.'s (2007) hedge fund sample has a similar cross-sectional distribution of performance measures. The more recent sample features a distribution of performance measures that has shifted to the left.

	1%	3%	5%	10%	90%	95%	97%	99%
Panel A: Kosowski et al. (2007)								
Monthly alpha (%)	−1.61	−0.88	−0.64	−0.33	1.24	1.63	2.08	2.82
p-value (bootstrapped)	0.42	1.00	1.00	1.00	0.00	0.00	0.00	0.00
t-alpha	−2.50	−1.82	−1.41	−0.81	3.78	4.90	5.86	7.87
p-value (bootstrapped)	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00
Panel B: Replication								
Monthly alpha (%)	−1.98	−1.05	−0.73	−0.34	1.36	1.87	2.37	3.68
p-value (bootstrapped)	0.81	0.98	0.97	0.81	0.00	0.01	0.01	0.00
t-alpha	−2.85	−1.99	−1.59	−0.87	3.78	4.94	5.92	8.15
p-value (bootstrapped)	0.98	0.96	0.86	0.83	0.00	0.01	0.00	0.00
Panel C: Extension								
Monthly alpha (%)	−2.03	−1.29	−1.05	−0.74	0.67	1.01	1.35	2.31
p-value (bootstrapped)	0.77	0.99	1.00	0.99	0.00	0.00	0.06	0.10
t-alpha	−3.78	−3.04	−2.66	−2.09	2.25	3.21	4.08	6.84
p-value (bootstrapped)	1.00	0.98	0.98	0.94	0.03	0.01	0.01	0.00

Table 3. Bootstrapped Sensitivity Analysis

Description: Panel A reproduces performance metrics of the 10th and 90th percentiles correcting for four different biases reported in Table 3 of Kosowski et al. (2007); Panel B shows our replication using data from the same time period 1994–2002; Panel C shows results in the later time period January 2003–June 2020. In each panel, the first and second rows report the ranked OLS Fung and Hsieh (2004) alpha in percent per month and its bootstrapped *p*-value. The third and fourth rows report the ranked *t*-statistic of alpha based on heteroscedasticity and autocorrelation consistent standard errors and its bootstrapped *p*-value. The bias corrections are: dropping the first 12 months to control for backfill; de-smoothing following the procedure in Getmansky et al. (2004); allowing for structural breaks in factor loadings identified by the Bai and Perron (1998) test (December 2000 in Panels A and B and December 2007 in Panel C); and a bootstrap that allows for cross-sectional dependence in residuals. **Interpretation:** the performance of top decile funds is robust to corrections for database biases.

	Backfill		Smoothing		Structural Break		Cross-Sectional Dep.	
	10%	90%	10%	90%	10%	90%	10%	90%
Panel A: Kosowski et al. (2007)								
Monthly alpha (%)	−0.41	1.12	−0.39	1.22	−0.02	1.18	−0.23	1.18
p-value (bootstrapped)	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
t-alpha	−0.97	3.76	−0.89	3.55	−0.08	4.69	−0.55	3.95
p-value (bootstrapped)	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Panel B: Replication								
Monthly alpha (%)	−0.45	1.20	−0.41	1.36	−0.37	1.40	−0.34	1.36
p-value (bootstrapped)	0.97	0.16	0.73	0.00	0.77	0.19	0.81	0.00
t-alpha	−1.01	3.73	−0.85	3.32	−0.95	3.99	−0.87	3.78
p-value (bootstrapped)	0.62	0.00	0.85	0.00	0.94	0.00	0.82	0.00
Panel C: Extension								
Monthly alpha (%)	−0.80	0.60	−0.82	0.66	−0.75	0.69	−0.74	0.67
p-value (bootstrapped)	0.98	0.02	1.00	0.00	0.99	0.06	0.98	0.00
t-alpha	−2.16	2.12	−2.00	1.92	−2.11	2.28	−2.09	2.25
p-value (bootstrapped)	0.95	0.01	0.95	0.03	0.95	0.02	0.96	0.02

Table 4. Performance Persistence

Description: Panel A reproduces performance metrics of hedge fund portfolios formed annually by sorting on prior Fung and Hsieh (2004) OLS alpha (Panel A1) or *t*-statistic of alpha (Panel B1) reported in Table 5 of Kosowski et al. (2007); Panels A2 and B2 show our replication. **Interpretation:** our replication of Kosowski et al.'s (2007) performance persistence based on OLS statistics is reasonably close for the top decile, but neither they nor we find a significant spread between the top and bottom deciles.

Panel A. Sort by OLS alpha estimate

	Panel A1. Kosowski et al. (2007)					Panel A2. Replication 1994–2002				
Portfolio	Mean	Vol	Alpha	*t*-stat	*p*-val	Mean	Vol	Alpha	*t*-stat	*p*-val
1%ile	0.89	15.13	−0.95	−0.19	0.43	2.10	21.44	−0.44	−0.06	0.53
5%ile	6.37	11.77	4.32	1.29	0.10	6.58	15.65	3.54	0.79	0.22
Decile 1	7.21	10.27	5.32	1.97	0.03	6.39	13.19	3.81	1.07	0.14
Decile 2	7.25	7.74	5.76	3.04	0.00	6.46	8.02	4.70	2.64	0.00
Decile 3	5.71	5.87	4.78	4.00	0.00	7.09	5.53	6.04	4.37	0.00
Decile 4	6.40	4.87	5.57	5.11	0.00	4.78	5.00	3.98	3.26	0.00
Decile 5	6.51	4.53	5.77	5.82	0.00	4.93	4.78	4.18	3.70	0.00
Decile 6	5.22	4.66	4.42	3.95	0.00	6.24	4.59	5.33	4.54	0.00
Decile 7	5.15	5.67	4.07	2.96	0.00	5.17	6.57	4.05	2.79	0.00
Decile 8	5.87	6.78	4.44	2.71	0.00	4.02	5.74	2.97	2.39	0.01
Decile 9	5.54	6.46	4.24	2.56	0.01	3.99	6.63	2.75	1.59	0.06
Decile 10	6.22	10.84	4.99	1.46	0.08	6.32	10.45	5.23	1.26	0.10
95%ile	7.14	13.72	5.67	1.35	0.09	4.46	13.48	3.42	0.63	0.26
99%ile	9.53	19.74	8.24	1.33	0.10	11.33	22.58	9.96	1.43	0.08
Δ 10%	0.99	10.90	0.33	0.08	0.47	0.07	12.55	−1.42	−0.31	0.62
Δ 1%	−8.65	23.20	−9.19	−1.15	0.13	−9.23	26.74	−10.40	−1.15	0.88

Panel B. Sort by OLS alpha *t*-statistic

	Panel B1. Kosowski et al. (2007)					Panel B2. Replication 1994–2002				
Portfolio	Mean	Vol	Alpha	*t*-stat	*p*-val	Mean	Vol	Alpha	*t*-stat	*p*-val
1%ile	7.34	2.26	7.03	9.76	0.00	6.92	4.01	6.57	4.54	0.00
5%ile	6.64	3.17	6.36	6.26	0.00	6.11	4.77	5.36	3.01	0.00
Decile 1	6.52	4.15	6.05	5.26	0.00	6.50	5.38	5.58	3.14	0.00
Decile 2	6.05	6.19	5.14	3.37	0.00	5.92	7.22	4.70	2.39	0.01
Decile 3	6.53	6.14	5.39	3.78	0.00	5.99	7.10	4.84	3.03	0.00
Decile 4	7.84	6.82	6.36	3.71	0.00	6.24	7.02	4.76	3.40	0.00
Decile 5	4.63	6.87	3.29	2.51	0.01	5.87	6.37	4.59	3.23	0.00
Decile 6	5.69	6.69	4.52	2.95	0.00	4.61	7.10	3.31	2.36	0.01
Decile 7	6.08	7.14	4.78	2.84	0.00	5.66	7.79	4.28	2.22	0.01
Decile 8	6.66	7.22	5.08	2.80	0.00	5.27	6.84	3.80	2.27	0.01
Decile 9	5.18	7.61	3.92	1.74	0.04	4.82	7.23	3.62	1.60	0.05
Decile 10	5.33	8.72	4.19	1.67	0.05	4.11	8.22	3.18	1.08	0.14
95%ile	4.86	9.44	3.85	1.38	0.09	4.69	9.13	3.93	1.16	0.12
99%ile	3.64	12.99	2.84	0.69	0.25	1.81	12.37	2.19	0.54	0.30
Δ 10%	1.19	7.71	1.86	0.73	0.24	2.39	7.59	2.40	0.84	0.20
Δ 1%	3.70	12.50	4.19	0.98	0.16	5.12	12.75	4.38	1.03	0.15

Table 5. Bayesian Performance Persistence

Description: Panel A reproduces performance metrics of hedge fund portfolios formed annually by sorting on prior Bayesian posterior alpha (Panel A1) or Bayesian posterior *t*-statistic of alpha (Panel B1) reported in Table 6 of Kosowski et al. (2007); Panels A2 and B2 show our replication. **Interpretation:** our replication of Kosowski et al.'s (2007) performance persistence based on Bayesian statistics is reasonably close for the top decile. We both find a significant spread between the top and bottom deciles when forming portfolios using the Bayesian alpha *t*-statistic.

Panel A. Sort by Bayesian alpha estimate

	Panel A1. Kosowski et al. (2007)					Panel A2. Replication 1994–2002				
Portfolio	Mean	Vol	Alpha	*t*-stat	*p*-val	Mean	Vol	Alpha	*t*-stat	*p*-val
1%ile	12.06	11.09	11.86	3.26	0.00	9.56	15.57	6.82	1.29	0.10
5%ile	8.24	7.78	7.37	3.47	0.00	6.40	11.92	3.78	0.99	0.16
Decile 1	9.31	7.76	8.21	4.35	0.00	7.23	10.43	4.89	1.66	0.05
Decile 2	7.93	7.15	6.93	4.36	0.00	7.39	7.26	5.67	4.11	0.00
Decile 3	6.04	5.50	5.22	4.37	0.00	6.75	6.07	5.63	3.91	0.00
Decile 4	6.15	4.87	5.54	5.78	0.00	4.83	5.21	3.89	3.06	0.00
Decile 5	5.65	4.95	4.73	4.26	0.00	6.04	4.45	5.24	3.94	0.00
Decile 6	4.67	4.60	3.79	3.64	0.00	5.41	5.26	4.51	4.03	0.00
Decile 7	5.23	5.78	4.18	3.51	0.00	4.11	4.88	3.17	2.56	0.01
Decile 8	5.14	6.20	4.32	3.20	0.00	3.79	5.92	2.63	2.27	0.01
Decile 9	5.13	8.92	3.57	2.09	0.02	4.78	7.22	3.64	1.69	0.05
Decile 10	4.84	13.18	2.40	0.88	0.19	4.33	12.18	2.96	0.70	0.24
95%ile	5.82	14.93	3.30	1.02	0.16	2.18	14.18	0.54	0.11	0.46
99%ile	6.24	16.46	4.61	0.88	0.19	1.42	25.44	−1.92	−0.22	0.59
Δ 10%	4.47	7.84	5.81	2.65	0.01	2.90	9.09	1.94	0.43	0.33
Δ 1%	5.83	16.12	7.25	1.21	0.12	8.13	24.38	8.74	0.83	0.20

Panel B. Sort by Bayesian alpha *t*-statistic

	Panel B1. Kosowski et al. (2007)					Panel B2. Replication 1994–2002				
Portfolio	Mean	Vol	Alpha	*t*-stat	*p*-val	Mean	Vol	Alpha	*t*-stat	*p*-val
1%ile	8.77	4.26	8.58	5.94	0.00	11.46	5.98	11.06	3.92	0.00
5%ile	7.34	5.15	6.98	5.03	0.00	8.06	5.40	7.17	3.19	0.00
Decile 1	7.25	5.61	6.56	4.96	0.00	7.44	6.07	6.34	3.00	0.00
Decile 2	6.71	5.81	5.97	4.92	0.00	5.51	5.98	4.31	3.04	0.00
Decile 3	8.51	5.35	7.74	6.65	0.00	6.87	6.35	5.39	4.46	0.00
Decile 4	5.55	5.75	4.57	3.14	0.00	6.57	6.14	5.21	4.08	0.00
Decile 5	7.05	6.27	6.10	4.34	0.00	4.77	6.64	3.75	2.30	0.01
Decile 6	4.93	6.17	3.77	2.83	0.00	6.41	6.80	5.08	3.37	0.00
Decile 7	6.27	6.71	5.06	3.56	0.00	4.10	6.99	2.65	1.66	0.05
Decile 8	4.33	7.65	3.28	2.20	0.02	5.23	7.18	3.94	2.50	0.01
Decile 9	4.33	9.08	2.58	1.45	0.08	4.68	7.71	3.62	1.58	0.06
Decile 10	5.18	9.85	3.41	1.74	0.04	3.05	8.94	1.96	0.72	0.24
95%ile	4.52	10.12	2.99	1.27	0.10	2.03	10.19	1.07	0.29	0.39
99%ile	2.88	9.17	2.04	0.70	0.24	−2.80	15.15	−3.47	−0.62	0.73
Δ 10%	2.07	6.36	3.15	1.76	0.04	4.38	6.50	4.38	1.58	0.06
Δ 1%	5.89	8.13	6.53	2.30	0.01	14.25	14.55	14.53	2.96	0.00

Table 6. Yearly Distribution of Hedge Funds

Description: Panel A reproduces the total number of funds that reported during a year and the mean, median, and standard deviation of monthly excess returns as reported in Jagannathan et al. (2010). A year represents the period from May of that year through April of the next year. Panel B shows corresponding statistics in our replicating sample. **Interpretation:** we can closely match the sample in Jagannathan et al. (2010).

Panel A. Jagannathan et al. (2010) Table V

Year	No. of Funds	Mean Return	Median Return	*SD*
1996	1,123	0.57%	0.61%	5.10%
1997	1,326	1.14%	0.86%	5.31%
1998	1,436	−0.19%	0.23%	7.98%
1999	1,479	1.50%	0.67%	7.97%
2000	1,546	−0.40%	0.12%	7.12%
2001	1,851	0.12%	0.24%	4.64%
2002	2,183	−0.09%	0.13%	4.34%
2003	2,744	1.11%	0.76%	3.31%
2004	3,274	0.23%	0.20%	2.86%

Panel B. Replication

Year	No. of Funds	Mean Return	Median Return	*SD*
1996	1,120	0.58%	0.61%	5.26%
1997	1,312	1.20%	0.88%	5.46%
1998	1,410	−0.18%	0.23%	8.15%
1999	1,430	1.58%	0.67%	8.15%
2000	1,488	−0.47%	0.10%	7.22%
2001	1,808	0.11%	0.23%	4.75%
2002	2,120	−0.10%	0.12%	4.40%
2003	2,710	1.10%	0.75%	3.34%
2004	3,405	0.22%	0.20%	2.86%

Table 7. Summary Statistics of Relative Performance Model Adjusted-R^2

Description: Panel A reproduces four sets of the cross-sectional mean, median, and standard deviation of the adjusted-R^2 from the relative performance model estimated over a three-year evaluation period and a three-year prediction period as reported in Jagannathan et al. (2010). The performance model is run on the combined six-year period and produces 72 monthly residuals, from which we calculate two separate adjusted-R^2. Panel B shows corresponding results using our replicating sample. **Interpretation:** we can closely match the sample in Jagannathan et al. (2010).

Panel A. Jagannathan et al. (2010) Table VII

	Evaluation Period Adjusted-R^2			Prediction Period Adjusted-R^2		
Cross-section	Mean	Median	SD	Mean	Median	SD
1996–1999 to 1999–2002	0.49	0.52	0.26	0.45	0.46	0.28
1997–2000 to 2000–2003	0.50	0.55	0.26	0.44	0.45	0.27
1998–2001 to 2001–2004	0.50	0.52	0.25	0.47	0.51	0.27
1999–2002 to 2002–2005	0.44	0.45	0.28	0.53	0.58	0.26

Panel B. Replication

	Evaluation Period Adjusted-R^2			Prediction Period Adjusted-R^2		
Cross-section	Mean	Median	SD	Mean	Median	SD
1996–1999 to 1999–2002	0.55	0.59	0.25	0.52	0.54	0.26
1997–2000 to 2000–2003	0.57	0.63	0.24	0.42	0.43	0.29
1998–2001 to 2001–2004	0.59	0.63	0.23	0.42	0.46	0.29
1999–2002 to 2002–2005	0.53	0.55	0.25	0.49	0.53	0.27

Table 8. OLS Regression Results

Description: Panels A and C reproduce parameter estimates of cross-sectional OLS regressions of prediction period alpha on evaluation period alpha with and without winsorizing, respectively, as reported in Jagannathan et al. (2010). Persistence is measured by the slope coefficient. Panels B and D show results from our replication. **Interpretation:** we can only partially replicate the original results, indicating noticeable noise in sorts based on alpha.

Panel A. Jagannathan et al. (2010) Table IX

	Intercept			Slope		
Cross-section	Estimate	*t*-statistic	*p*-value	Estimate	*t*-statistic	*p*-value
1996–1999 to 1999–2002	−0.15	−1.17	0.24	0.38	3.15	0.00
1997–2000 to 2000–2003	−0.05	−0.41	0.68	0.29	2.40	0.02
1998–2001 to 2001–2004	−0.15	−1.64	0.10	0.06	0.69	0.49
1999–2002 to 2002–2005	−0.27	−5.02	0.00	0.13	2.97	0.00

Panel B. Replication

	Intercept			Slope		
Cross-section	Estimate	*t*-statistic	*p*-value	Estimate	*t*-statistic	*p*-value
1996–1999 to 1999–2002	−0.06	−1.10	0.27	0.14	1.99	0.05
1997–2000 to 2000–2003	−0.09	−2.18	0.03	−0.12	−2.63	0.01
1998–2001 to 2001–2004	−0.01	−0.29	0.77	0.06	1.70	0.09
1999–2002 to 2002–2005	0.04	1.47	0.14	0.21	5.70	0.00

Panel C. Jagannathan et al. (2010) with Winsorized α_1, Table X

	Intercept			Slope		
Cross-section	Estimate	*t*-statistic	*p*-value	Estimate	*t*-statistic	*p*-value
1996–1999 to 1999–2002	−0.07	−1.01	0.31	0.11	1.73	0.09
1997–2000 to 2000–2003	−0.10	−1.64	0.10	−0.01	−0.18	0.85
1998–2001 to 2001–2004	−0.20	−4.55	0.00	0.03	0.86	0.39
1999–2002 to 2002–2005	−0.23	−7.65	0.00	0.15	5.99	0.00

Panel D. Replication with Winsorized α_1

	Intercept			Slope		
Cross-section	Estimate	*t*-statistic	*p*-value	Estimate	*t*-statistic	*p*-value
1996–1999 to 1999–2002	−0.05	−0.96	0.34	0.12	1.92	0.06
1997–2000 to 2000–2003	−0.10	−2.60	0.01	−0.11	−2.57	0.01
1998–2001 to 2001–2004	−0.01	−0.24	0.81	0.05	1.57	0.12
1999–2002 to 2002–2005	0.06	2.71	0.01	0.15	5.08	0.00

Table 9. Weighted Least Squares Regression Results

Description: Panels A and C reproduce parameter estimates of cross-sectional weighted least squares regressions of prediction period stylized *t*-statistic of alpha on evaluation period *t*-statistic of alpha with and without winsorizing, respectively, as reported in Jagannathan et al. (2010). Persistence is measured by the slope coefficient. Panels B and D show results from our replication. **Interpretation:** weighted least squares tightens the match between the original results and our replication by reducing the impact of noise on the regression parameter estimates. The combination of weighted least squares and winsorizing results in a very close replication.

Panel A. Jagannathan et al. (2010) Table XI

	Intercept			Slope		
Cross-section	Estimate	*t*-statistic	*p*-value	Estimate	*t*-statistic	*p*-value
1996–1999 to 1999–2002	0.13	0.63	0.53	0.53	4.48	0.00
1997–2000 to 2000–2003	−0.29	−1.42	0.16	0.35	2.90	0.00
1998–2001 to 2001–2004	−0.26	−0.94	0.35	0.04	0.26	0.79
1999–2002 to 2002–2005	−0.73	−5.91	0.00	0.21	3.28	0.00

Panel B. Replication

	Intercept			Slope		
Cross-section	Estimate	*t*-statistic	*p*-value	Estimate	*t*-statistic	*p*-value
1996–1999 to 1999–2002	0.23	1.46	0.15	0.20	3.49	0.00
1997–2000 to 2000–2003	0.03	0.27	0.79	0.19	4.47	0.00
1998–2001 to 2001–2004	0.07	0.66	0.51	0.17	4.28	0.00
1999–2002 to 2002–2005	0.24	2.17	0.03	0.42	10.05	0.00

Panel C. Jagannathan et al. (2010) with Winsorized $t^*_{\alpha_1}$, Table XII

	Intercept			Slope		
Cross-section	Estimate	*t*-statistic	*p*-value	Estimate	*t*-statistic	*p*-value
1996–1999 to 1999–2002	0.17	1.04	0.30	0.39	4.25	0.00
1997–2000 to 2000–2003	−0.06	−0.53	0.60	0.19	2.65	0.01
1998–2001 to 2001–2004	−0.41	−4.74	0.00	0.10	1.91	0.06
1999–2002 to 2002–2005	−0.59	−6.47	0.00	0.20	4.16	0.00

Panel D. Replication with Winsorized $t^*_{\alpha_1}$

	Intercept			Slope		
Cross-section	Estimate	*t*-statistic	*p*-value	Estimate	*t*-statistic	*p*-value
1996–1999 to 1999–2002	0.24	1.65	0.10	0.24	4.48	0.00
1997–2000 to 2000–2003	0.03	0.31	0.76	0.14	3.61	0.00
1998–2001 to 2001–2004	0.08	1.00	0.32	0.16	5.05	0.00
1999–2002 to 2002–2005	0.29	3.65	0.00	0.29	9.67	0.00

Table 10. Three Relative Performance Ranked Portfolios with 33% Cutoff

Description: portfolios are formed and ranked according to the *t*-statistic of relative alpha in the evaluation period with a 33% cutoff. The Fung and Hsieh (2004) portfolio past and out-of-sample alpha, as well as the out-of-sample appraisal ratio, are reported. Panel A shows results from Jagannathan et al. (2010). Panel B shows results from our replication using 2007 HFR snapshot. '*', '**', and '***' denote significance at the 10%, 5%, and 1% significance level, respectively. **Interpretation:** performance persists in three of the four subsamples. We can closely match the results in Jagannathan et al. (2010).

Panel A. Jagannathan et al. (2010) Table XXII

Cross-section	Portfolio	Funds at Formation	Survived Funds	Past Alpha	Out-of-Sample Alpha	Appraisal ratio
1996–1999 to 1999–2002	Inferior	163	99	−0.1075	0.0960	0.0659
	Neutral	167	109	0.1995	0.0991	0.0736
	Superior	163	110	0.7497***	0.3292***	0.5083
1997–2000 to 2000–2003	Inferior	223	140	−0.0641	0.0202	0.0200
	Neutral	227	148	0.5016**	0.0074	0.0107
	Superior	223	170	0.9010***	0.0943	0.1666
1998–2001 to 2001–2004	Inferior	239	148	−0.1032	0.3116**	0.4341
	Neutral	245	172	0.4540**	0.2026**	0.3727
	Superior	239	187	0.8738***	0.2693***	0.6795
1999–2002 to 2002–2005	Inferior	252	156	−0.0921	−0.0138	−0.0219
	Neutral	259	172	0.5705***	0.1214	0.2158
	Superior	252	191	0.8685***	0.1768*	0.6494

Panel B. Replication with 2007 HFR database

Cross-section	Portfolio	Funds at Formation	Survived Funds	Past Alpha	Out-of-Sample Alpha	Appraisal ratio
1996–1999 to 1999–2002	Inferior	159	86	−0.3063	0.3146	0.2174
	Neutral	158	105	0.2184	0.2003	0.1351
	Superior	158	105	0.6519***	0.2852*	0.3310
1997–2000 to 2000–2003	Inferior	213	127	−0.1701	0.0432	0.0475
	Neutral	212	144	0.5444**	−0.1338	−0.2148
	Superior	212	149	0.9322***	−0.0310	−0.0591
1998–2001 to 2001–2004	Inferior	227	135	0.0060	0.0884	0.1348
	Neutral	226	164	0.5566***	0.1994*	0.3415
	Superior	227	165	0.9844***	0.2145***	0.5021
1999–2002 to 2002–2005	Inferior	239	141	0.0678	0.0943	0.1331
	Neutral	238	149	0.5499***	0.0985	0.1640
	Superior	238	182	0.9676***	0.1794**	0.4482

Table 11. JMN Persistence with a One-Year Holding Period

Description: this table shows the predictive power of JMN relative alpha *t*-statistic when using portfolio sorts with a one-year holding period as in Carhart (1997) and KNT. The first specification uses 2007 the HFR snapshot without an AUM threshold, with portfolios formed at the beginning of each May, following JMN. The second specification adds a reasonable formation-time AUM threshold of 20 M USD as in KNT. The out-of-sample period in both of these specifications is May 1999 through April 2005. The third specification addresses calendar sensitivity by forming the portfolios at the beginning of each January, for an out-of-sample period of January 2000 through December 2005. The fourth and the fifth specifications utilize the 2020 HFR snapshot and our aggregate database, respectively, to isolate the sensitivity to database choice. **Interpretation:** the predictive power of the JMN relative alpha *t*-statistic is maintained when focusing on the top decile and using a standard one-year holding period, an AUM filter, and a January formation date. The top decile's performance is not as high using the JMN relative alpha than when using the KNT Bayesian alpha to form portfolios.

	Specification 1 2007 HFR snapshot No AUM threshold May formation		Specification 2 2007 HFR snapshot AUM ≥ 20 M USD May formation		Specification 3 2007 HFR snapshot AUM ≥ 20 M USD January formation		Specification 4 2020 HFR snapshot AUM ≥ 20 M USD January formation		Specification 5 Aggregate database AUM ≥ 20 M USD January formation	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic	Alpha	*t*-statistic	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	5.31	3.40	5.23	4.04	6.78	5.16	6.38	2.73	5.71	2.79
5%ile	4.58	3.17	5.04	2.71	5.27	3.46	4.89	5.12	4.67	4.24
Decile 1	3.84	2.67	4.52	3.71	4.41	4.39	3.49	4.12	4.43	5.32
Decile 2	2.12	0.98	1.19	0.62	2.07	1.90	1.71	1.69	1.50	0.91
Decile 3	2.50	2.48	2.44	1.95	0.60	0.60	0.28	0.17	2.26	1.84
Decile 4	2.12	1.19	3.03	2.51	1.45	1.73	1.04	0.78	1.71	1.07
Decile 5	0.07	0.03	1.74	0.69	-0.88	-0.62	-1.06	-0.89	1.78	1.18
Decile 6	1.41	0.50	1.52	0.71	0.22	0.16	-1.28	-1.06	1.14	0.93
Decile 7	2.30	1.09	2.26	1.16	0.44	0.32	-1.50	-0.97	2.63	2.48
Decile 8	-2.14	-1.09	0.21	0.13	-2.06	-1.81	-1.70	-1.36	0.98	0.92
Decile 9	4.97	1.73	1.71	0.64	-2.30	-1.07	1.26	0.92	2.50	1.94
Decile 10	0.71	0.29	4.90	2.44	0.64	0.51	1.59	1.19	1.12	0.83
95%ile	1.27	0.74	4.76	3.68	1.80	1.18	1.99	1.34	2.71	2.15
99%ile	2.21	0.69	1.57	0.78	6.17	1.84	7.05	2.44	5.90	2.22
Spread 10%	3.12	1.21	-0.37	-0.22	3.76	3.34	1.90	1.67	3.31	2.58
Spread 1%	3.10	1.14	3.66	1.71	0.61	0.23	-0.67	-0.24	-0.20	-0.09

Table 12. Post-Publication Performance Persistence

Description: listed are tests of the predictive power of KNT Bayesian alpha *t*-statistic (Panel A) and JMN relative alpha *t*-statistic (Panel B) outside the sample periods used in the original papers, using our union of six databases. The estimation of the alpha *t*-statistics follows the original papers. Portfolios are formed at the beginning of January and held for one year. **Interpretation:** sorting on the KNT Bayesian alpha *t*-statistic results in higher out-of-sample performance for the top decile, and a larger spread between the top and bottom deciles, than sorting on the JMN relative alpha *t*-statistic.

Panel A. Sort by Bayesian alpha *t*-statistic (2003–2020)

	Fund-weighted		Asset-weighted	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	5.78	3.48	4.11	3.03
5%ile	4.21	3.43	2.05	1.69
Decile 1	2.85	2.45	1.61	1.20
Decile 2	0.50	0.46	-0.45	-0.35
Decile 3	0.12	0.11	-0.20	-0.17
Decile 4	-0.20	-0.18	-0.94	-0.89
Decile 5	-0.25	-0.25	-1.37	-1.17
Decile 6	-0.99	-0.94	-0.51	-0.50
Decile 7	-0.93	-0.83	-1.47	-1.13
Decile 8	-1.47	-1.15	-1.27	-0.97
Decile 9	-2.24	-1.64	-2.43	-1.53
Decile 10	-2.34	-1.53	-2.32	-1.44
95%ile	-2.35	-1.43	-2.48	-1.36
99%ile	-1.71	-0.69	-3.63	-2.06
Spread 10%	5.19	3.80	3.93	2.71
Spread 1%	7.49	3.13	7.74	3.79

Panel B. Sort by relative alpha *t*-statistic (2006–2020)

	Fund-weighted		Asset-weighted	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	4.00	3.88	3.24	2.75
5%ile	1.27	1.48	1.10	1.02
Decile 1	1.35	1.44	1.49	1.42
Decile 2	0.95	0.71	0.67	0.53
Decile 3	0.15	0.15	-0.43	-0.29
Decile 4	0.03	0.02	-0.05	-0.04
Decile 5	-0.56	-0.47	-1.76	-1.22
Decile 6	-0.60	-0.54	-1.60	-1.13
Decile 7	-0.82	-0.66	-0.02	-0.02
Decile 8	-1.19	-1.13	-1.10	-1.13
Decile 9	-1.06	-1.04	-1.44	-1.24
Decile 10	-0.78	-0.74	-0.78	-0.68
95%ile	-0.72	-0.73	-0.23	-0.19
99%ile	-0.76	-0.65	-1.85	-1.32
Spread 10%	2.13	2.99	2.27	2.79
Spread 1%	4.76	3.88	5.09	3.72

Table 13. Variations of KNT and JMN

Description: this table tests the sensitivity of performance persistence to the choice of benchmark and non-benchmark assets. "Table 12" shows the "Fund-weighted" results from Table 12. In Panel A, "Short window" uses a 36-month rolling window for non-benchmark assets and "FH7 benchmark" eliminates non-benchmark assets. In Panel B, "No lags" removes the evaluation-stage lag terms from JMN relative alpha and "FH7 benchmark" uses the full FH7 benchmark instead of the equity market factor. **Interpretation:** the JMN relative alpha suffers from overfitting relative to the KNT methodology. Most of the KNT predictability would be achieved via a simple OLS regression on the FH7 benchmark when using the more recent data.

Panel A: Variants of KNT Bayesian alpha *t*-statistic (2003–2020)

	Table 12		Short window		FH7 benchmark	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	5.78	3.48	6.15	3.99	5.43	9.63
5%ile	4.21	3.43	4.23	3.41	3.72	3.98
Decile 1	2.85	2.45	2.76	2.36	2.80	2.79
Decile 2	0.50	0.46	-0.02	-0.02	0.56	0.47
Decile 3	0.12	0.11	0.49	0.45	-0.29	-0.23
Decile 4	-0.20	-0.18	-0.32	-0.28	0.09	0.08
Decile 5	-0.25	-0.25	-0.22	-0.22	-0.60	-0.51
Decile 6	-0.99	-0.94	-1.05	-1.02	-0.74	-0.63
Decile 7	-0.93	-0.83	-0.90	-0.79	-1.67	-1.37
Decile 8	-1.47	-1.15	-1.39	-1.11	-1.29	-1.02
Decile 9	-2.24	-1.64	-1.91	-1.44	-1.72	-1.26
Decile 10	-2.34	-1.53	-2.31	-1.56	-1.91	-1.34
95%ile	-2.35	-1.43	-2.36	-1.50	-1.73	-1.11
99%ile	-1.71	-0.69	-2.11	-0.99	-1.24	-0.80
Spread 10%	5.19	3.80	5.07	3.89	4.70	2.73
Spread 1%	7.49	3.13	8.26	3.91	6.66	3.73

Panel B: Variants of JMN relative alpha *t*-statistic (2006–2020)

	Table 12		No lags		FH7 benchmark	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	4.00	3.88	4.98	7.16	5.18	6.47
5%ile	1.27	1.48	2.41	2.10	2.19	1.85
Decile 1	1.35	1.44	1.69	1.33	1.74	1.60
Decile 2	0.95	0.71	0.63	0.49	0.91	0.72
Decile 3	0.15	0.15	0.05	0.05	-0.37	-0.29
Decile 4	0.03	0.02	-0.09	-0.09	-0.06	-0.06
Decile 5	-0.56	-0.47	-0.52	-0.43	-0.66	-0.55
Decile 6	-0.60	-0.54	-0.21	-0.20	-0.45	-0.37
Decile 7	-0.82	-0.66	-0.05	-0.05	0.03	0.03
Decile 8	-1.19	-1.13	-1.44	-1.17	-1.05	-0.89
Decile 9	-1.06	-1.04	-1.04	-1.01	-0.61	-0.56
Decile 10	-0.78	-0.74	-1.53	-1.51	-2.01	-1.68
95%ile	-0.72	-0.73	-1.97	-1.64	-2.91	-2.28
99%ile	-0.76	-0.65	-1.45	-1.02	-4.15	-1.40
Spread 10%	2.13	2.99	3.22	3.10	3.75	4.08
Spread 1%	4.76	3.88	6.43	4.26	9.33	3.77

Table 14. Impact of real-time data corrections

Description: this table compares the predictive powers of KNT Bayesian and JMN relative alpha *t*-statistics when using data as-reported to commercial databases versus data treated using the backfill correction of Jorion and Schwarz (2019) and a three-month lag between an investor's evaluation period and holding period. Panel A shows results KNT Bayesian alpha *t*-statistic whereas whereas Panel B shows results for the JMN relative alpha *t*-statistic. Portfolios are sorted at in January and held for one year. The "As-reported" results correspond to the "Fund-weighted" results from Table 12. **Interpretation:** the KNT Bayesian alpha *t*-statistic is more robust to real-time data corrections than the JMN relative alpha *t*-statistic, and results in substantially higher top-decile performance and a far larger spread between the top and bottom deciles.

Panel A: Sort by Bayesian alpha *t*-statistic (2003–2020)

	Real-time		As-reported	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	6.26	6.16	5.78	3.48
5%ile	4.29	5.21	4.21	3.43
Decile 1	2.96	3.45	2.85	2.45
Decile 2	1.33	1.50	0.50	0.46
Decile 3	0.62	0.68	0.12	0.11
Decile 4	-0.51	-0.55	-0.20	-0.18
Decile 5	-0.54	-0.52	-0.25	-0.25
Decile 6	-0.56	-0.60	-0.99	-0.94
Decile 7	-0.67	-0.66	-0.93	-0.83
Decile 8	-1.23	-1.05	-1.47	-1.15
Decile 9	-0.75	-0.61	-2.24	-1.64
Decile 10	-1.29	-1.00	-2.34	-1.53
95%ile	-1.18	-0.81	-2.35	-1.43
99%ile	0.01	0.00	-1.71	-0.69
Spread 10%	4.24	3.96	5.19	3.80
Spread 1%	6.25	3.50	7.49	3.13

Panel B: Sort by relative alpha *t*-statistic (2006–2020)

	Real-time		As-reported	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	2.36	2.45	4.00	3.88
5%ile	1.79	2.52	1.27	1.48
Decile 1	1.68	2.19	1.35	1.44
Decile 2	1.46	1.27	0.95	0.71
Decile 3	0.63	0.64	0.15	0.15
Decile 4	-0.03	-0.03	0.03	0.02
Decile 5	0.14	0.15	-0.56	-0.47
Decile 6	0.30	0.33	-0.60	-0.54
Decile 7	0.02	0.02	-0.82	-0.66
Decile 8	-0.69	-0.76	-1.19	-1.13
Decile 9	0.59	0.60	-1.06	-1.04
Decile 10	0.14	0.16	-0.78	-0.74
95%ile	0.04	0.04	-0.72	-0.73
99%ile	0.15	0.12	-0.76	-0.65
Spread 10%	1.54	1.81	2.13	2.99
Spread 1%	2.20	1.73	4.76	3.88

Table 15. Real-Time Performance Persistence 1994–2020

Description: this table tests the predictive power of KNT Bayesian alpha *t*-statistic for fund-weighted returns (Panel A) and asset-weighted returns (Panel B) in our full January 1994–June 2020 period and union of six databases. Portfolios are sorted at in January, including only funds reporting as of portfolio formation. Results in this and all subsequent tables are based on data treated using the backfill correction of Jorion and Schwarz (2019) and a three-month lag between an investor's evaluation period and holding period. **Interpretation:** the top decile and spread portfolios continue to generate positive and significant alpha over the full sample, though there is a notable decline in the magnitude of performance compared to the original KNT results.

Panel A. Sort by Bayesian alpha *t*-statistic: Fund-weighted returns

	One-year Holding period		Two-year Holding period		Three-year Holding period	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	6.99	6.95	6.47	6.11	5.84	6.12
5%ile	3.95	4.09	3.34	3.94	3.05	3.31
Decile 1	3.21	3.47	2.36	2.58	2.36	2.55
Decile 2	2.30	2.75	1.30	1.54	2.27	2.04
Decile 3	1.23	1.22	0.51	0.48	1.20	1.00
Decile 4	1.13	1.01	0.26	0.23	0.72	0.51
Decile 5	0.29	0.33	0.19	0.20	-0.18	-0.20
Decile 6	0.20	0.21	0.05	0.06	-0.89	-1.05
Decile 7	0.38	0.36	0.19	0.18	0.11	0.10
Decile 8	-0.35	-0.31	-0.69	-0.70	-0.86	-0.81
Decile 9	1.02	0.82	0.59	0.51	-0.05	-0.04
Decile 10	-0.69	-0.50	-0.54	-0.40	-0.58	-0.45
95%ile	-1.20	-0.77	-0.15	-0.09	-0.21	-0.16
99%ile	-1.12	-0.55	-0.33	-0.17	-0.34	-0.19
Spread 10%	3.90	3.17	2.90	2.33	2.93	2.29
Spread 1%	8.12	3.78	6.80	3.74	6.18	3.32

Panel B. Sort by Bayesian alpha *t*-statistic: Asset-weighted returns

	One-year Holding period		Two-year Holding period		Three-year Holding period	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	4.64	4.21	4.88	3.96	4.19	4.45
5%ile	3.59	3.74	2.71	2.81	2.92	3.17
Decile 1	3.13	2.79	2.03	1.98	2.53	2.62
Decile 2	0.47	0.44	-0.26	-0.25	1.48	1.02
Decile 3	1.53	1.17	0.53	0.37	0.65	0.43
Decile 4	0.34	0.22	0.42	0.26	0.17	0.08
Decile 5	0.93	0.77	0.92	0.67	0.18	0.17
Decile 6	-0.12	-0.09	-0.71	-0.46	-0.20	-0.17
Decile 7	-0.59	-0.52	-0.01	0.00	-0.35	-0.32
Decile 8	-0.79	-0.63	-0.77	-0.62	-0.80	-0.68
Decile 9	0.52	0.33	0.79	0.54	0.02	0.02
Decile 10	-0.76	-0.44	-0.89	-0.55	-0.87	-0.56
95%ile	-1.23	-0.60	-0.90	-0.48	-1.10	-0.71
99%ile	-3.29	-1.38	-1.86	-0.77	-2.21	-1.15
Spread 10%	3.89	2.50	2.92	1.81	3.39	2.09
Spread 1%	7.92	3.11	6.74	2.86	6.41	3.43

Table 16. Real-Time Fund-of-fund Performance Persistence 1994–2020

Description: listed are results of performance persistence tests for fund-of-funds based on the KNT Bayesian alpha *t*-statistic. Panel A shows results using fund-weighted returns and Panel B using asset-weighted returns in our full January 1994–June 2020 period and union of six databases. Portfolios are sorted in January, including only funds reporting as of portfolio formation. Results are based on data treated using the backfill correction of Jorion and Schwarz (2019) and a three-month lag between an investor's evaluation period and holding period. **Interpretation:** the predictive power of the KNT Bayesian alpha *t*-statistic is robust to using fund-of-fund returns as a means of dealing with a censorship bias in hedge fund databases.

Panel A. Sort by Bayesian alpha *t*-statistic: Fund-weighted returns

	One-year Holding period		Two-year Holding period		Three-year Holding period	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	5.40	4.59	5.66	4.57	4.50	2.96
5%ile	3.73	3.64	3.10	2.96	2.59	2.13
Decile 1	3.40	3.09	2.79	2.45	2.14	1.85
Decile 2	1.21	1.08	0.52	0.45	0.56	0.47
Decile 3	0.92	0.59	0.36	0.20	0.26	0.10
Decile 4	0.21	0.16	0.00	0.00	1.30	0.84
Decile 5	-0.58	-0.46	-0.31	-0.24	-0.47	-0.40
Decile 6	-0.27	-0.24	-0.02	-0.02	-1.58	-1.13
Decile 7	-1.32	-1.10	-0.99	-0.74	-0.98	-0.91
Decile 8	-0.85	-0.76	-0.92	-0.79	0.19	0.17
Decile 9	-0.15	-0.13	-0.88	-0.75	-1.72	-1.05
Decile 10	-2.68	-1.39	-2.43	-1.41	-2.46	-1.29
95%ile	-2.97	-1.26	-2.25	-1.06	-1.53	-0.74
99%ile	-1.09	-0.51	-3.24	-2.57	-2.71	-1.38
Spread 10%	6.07	2.91	5.22	2.84	4.59	2.21
Spread 1%	6.49	2.57	8.91	5.12	7.21	3.03

Panel B. Sort by Bayesian alpha *t*-statistic: Asset-weighted returns

	One-year Holding period		Two-year Holding period		Three-year Holding period	
Portfolio	Alpha	*t*-statistic	Alpha	*t*-statistic	Alpha	*t*-statistic
1%ile	3.86	1.83	4.71	2.20	4.56	1.61
5%ile	3.72	2.15	2.55	1.84	2.45	1.44
Decile 1	3.51	2.17	2.61	1.91	2.12	1.48
Decile 2	1.34	0.94	0.27	0.21	0.87	0.63
Decile 3	0.50	0.32	-0.18	-0.10	-0.05	-0.02
Decile 4	0.29	0.22	0.34	0.22	2.12	1.12
Decile 5	0.35	0.16	0.84	0.32	-0.02	-0.02
Decile 6	-0.17	-0.12	0.22	0.15	-1.51	-1.01
Decile 7	-2.62	-1.48	-2.19	-1.23	-2.74	-1.45
Decile 8	-0.94	-0.70	-0.51	-0.38	-0.09	-0.08
Decile 9	-1.02	-0.58	-0.34	-0.19	-0.18	-0.12
Decile 10	-2.08	-0.94	-2.34	-1.24	-3.16	-1.22
95%ile	-3.02	-1.35	-1.03	-0.44	-1.06	-0.53
99%ile	-1.68	-0.79	-3.01	-2.15	-3.98	-1.78
Spread 10%	5.59	2.56	4.95	2.36	5.27	1.84
Spread 1%	5.54	1.95	7.73	3.52	8.54	2.74

Table 17. Decreasing Returns to Scale

Description: listed are results of regressions of hedge fund performance as measured by FH7 alpha (%) in year t as a function of lagged size and lagged performance as measured by the Bayesian t-statistic of alpha measured over the prior two years. "Late" is an indicator variable equal to one for years $t \geq 2008$ and zero otherwise. "Size" is ln(AUM) measured at the end of year $t - 1$. **Interpretation:** the interaction between lagged fund size and lagged fund performance reveals evidence of decreasing returns to scale for high-performing funds.

	Model				
	1	2	3	4	5
Size	0.125	0.535	-0.055	-0.049	0.147
	(2.21)	(4.22)	(-0.39)	(-0.36)	(0.89)
Size x *t*-statistic				-0.072	-0.168
				(-3.35)	(-6.23)
Size x Late		-0.446			-0.223
		(-3.16)			(-1.06)
Size x *t*-statistic x Late					0.105
					(2.75)
Late		-5.845			
		(-7.83)			
t-statistic x Late					-0.411
					(-1.42)
t-statistic			0.494	0.869	1.250
			(3.54)	(4.80)	(6.41)
Intercept	-2.323	1.943			
	(-7.72)	(2.90)			
Adj. R^2	0.000	0.034	0.102	0.103	0.103
Observations	59,602	59,602	59,602	59,602	59,602
Style FEs?	No	No	Yes	Yes	Yes
Year FEs?	No	No	Yes	Yes	Yes
SEs clustered by fund?	No	No	Yes	Yes	Yes
SEs clustered by year?	No	No	Yes	Yes	Yes