Thousands of Alpha

Giglio. S., Liu. Y., Xiao. D.

The Review of Financial Studies, 2020.

Motivation

- Data snooping is a major concern in empirical asset pricing. With multiple testing comes the concern that as more tests are performed, an increasing number of them will be positive purely due to chance.
- Anomalies potentially have exposures to unknown risk factors and fund managers may trade common factors that are not observable.
- Many time series of returns have short histories or missing records, and these returns likely follow a factor model.

Existing Literatures

- The existing literature in asset pricing is aware of the data-snooping with multiple testing, taken in response two alternative approaches.
- One has been to abandon the multiple-testing problem altogether: to ask whether any fund beats the benchmark, or whether funds on average beat the benchmark (White 2000; Kosowski et al. 2006; Fama and French 2010)
- The second approach imports statistical methods that directly control the FWER or the FDR, facing omitting factors and missing data (Barras et al. 2010; Bajgrowicz and Scaillet 2012; Harvey et al. 2015)

Existing Literatures

- Targeting the question of whether any error was made, earlier work mainly focuses on controlling the FWER (Simes 1986; Holm 1979). These procedures guard against any single false discovery and hence are overly conservative.

- The FDR control takes into account the number of erroneous rejections and controls the expected proportion of errors among the rejected hypotheses (Benjamini and Hochberg 1995; Benjamini and Yekutieli 2001; Storey et al. 2004).

Existing Literatures

- Alternative methods guard against the drawbacks of the standard FDR approach in the presence of dependence (Leek and Storey 2008; Romano and Wolf 2005; Fan and Han 2016), however, do not exploit the factor structure of asset returns, nor are they directly applicable to an unbalanced panel.
- There is also a burgeoning body of research that applies machine learning methods to push the frontiers of empirical asset pricing (Kozak et al. 2017; Freyberger et al. 2017; Giglio and Xiu 2017; Feng et al. 2020; Kelly et al. 2017; Gu et al. 2018), however, do not directly control the number of false discoveries.

Methodology

- Our framework is based on a combination of three key ingredients: factor analysis, Fama-MacBeth regressions, and false discovery control.
- In a first step, we use time-series regressions to estimate fund exposures to (observable) benchmark factors. We further apply PCA or matrix completion to residuals to recover the missing commonalities.
- Next, we implement cross-sectional regressions like Fama-MacBeth to estimate the risk premiums of the factors and the alphas relative to the augmented benchmark model.
- Finally, we build t-statistics for these alphas and apply the B-H procedure for the FDR control.

Background

- We begin with a description of the model. We assume the $N \times 1$vector of excess returns r_t follows

$$r_t = \alpha + \beta\lambda + \beta(f_t - E(f_t)) + u_t$$

- Where f_t is a $K \times 1$vector of factors and u_t is the idiosyncratic component. The parameter λ is a $K \times 1$ vector of factor risk premiums

- The objective is to find individual funds with truly positive alphas.

$$H_0^i: \alpha_i \leq 0, i = 1, \dots, N. \longrightarrow H_0^i: \alpha_i = 0, i = 1, \dots, N.$$

$$GRS\ test: \quad H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_N = 0$$

The former is a multiple-testing problem that addresses which funds have significantly positive alphas. In contrast, the latter addresses whether there exists (at least one) fund whose alpha is significantly different from zero.

Background

- The possibility that many of the tests will look significant by pure chance, even if their true alpha is zero.
- For example, suppose there are 1,000 funds available, with only 10% of them having positive alphas. Conducting 1,000 tests independently would yield $1000 \times (1 - 10\%) \times 5\% = 45$ false positive alphas, in addition to $1000 \times 10\% = 100$ true positive alphas
- Consequently, among the $100 + 45 = 145$ "skilled" fund managers we find, almost one-third of them are purely due to luck.

Controlling FDR

- Let $H_0 \subset \{1, \ldots, N\}$ denote the set of indices for which the corresponding null hypotheses are true, R be the total number of rejections in a sample, and let F be the number of false rejections in that sample:

$$F = \sum_{i=1}^{N} 1\{i \leq N : t_i > c_i \; and \; i \; \in H_0\}, \qquad R = \sum_{i=1}^{N} 1\{i \leq N : t_i > c_i\},$$

- In a specific sample, we can obviously observe R, but we cannot observe F. we write the FDP and its expectation, FDR, as

$$FDP = \frac{F}{\max\{R, 1\}}, \qquad FDR = E(FDP)$$

- For comparison, we can also write the per-test error rate, $E(F)/N$, and the $FWER, P(F \geq 1) \leq \tau$

B-H procedure

- The FDR control procedure strikes a balance between these two approaches. It accepts a certain number of false discoveries as the price to pay to gain power in detecting true rejections.
- Algorithm 1(B-H procedure) Let N_0 be the num. of true null hypotheses.
- We still do not know N_0, so we replace it with some upper bound M

$$F(p) \underset{\approx}{^{(a)}} N_0 P(p_i < p | \alpha_i \leq 0) \underset{\leq}{^{(b)}} N_0 P(p_i < p | \alpha_i = 0) \underset{=}{^{(c)}} N_0 p \approx Mp$$

$$\boxed{E\left(\frac{F(P)}{R(P)}\right) \leq \tau} \qquad p \leq \frac{\tau R(p)}{M} = \frac{\tau \sum_{i=1}^{N} 1\{i \leq N: p_i > p\}}{\mathrm{M}}$$

$$p^* = max\left\{p \in (0,1): p \leq \frac{\tau \sum_{i=1}^{N} 1\{i \leq N: p_i > p\}}{\mathrm{M}}\right\}$$

Alpha Screening

- Based on the above discussion, not surprisingly, the count of negative alphas adversely affects the power of the B-H procedure.

$$F(p) \underset{\approx}{\overset{(a)}{}} \boxed{N_0 P(p_i < p | \alpha_i \leq 0) \underset{\leq}{\overset{(b)}{}} N_0 P(p_i < p | \alpha_i = 0)} \underset{=}{\overset{(c)}{}} N_0 p \approx Mp$$

- Holding the number of non-negative alphas constant, as the count of negative alphas increases, the critical value shrinks

- We tackle this problem by using a simple dimension reduction technique, the screening method in the context of testing for inequalities.

Alpha Screening

- The idea is that when some of the alphas are “overwhelmingly negative" their corresponding hypotheses could be simply eliminated from the set of candidate hypotheses.
- This would reduce the total count of hypotheses and thereby improve the power of FDR control. Based on this idea, we propose to reduce the set of funds to

$$\hat{I} = \left\{ i \leq N : t_i > -\log(\log T)\sqrt{logN} \right\}$$

- Where the threshold depends on the sample size and the cross-sectional dimension. We thereby can safely consider a smaller set of funds $\hat{I}$ for FDR control.

Compare with BSW(2010)

- Alternatively, Barras et al. (2010) apply a simple adjustment proposed by Storey (2002) to improve the power of the B-H procedure.

- Specifically, they suggest replacing $\hat{k}\tau = N$ in the cutoff value by $\hat{k}\tau = N_0$, where N_0 is the number of true null hypotheses that can be estimated using $\widehat{N}_0 = (1-\lambda)^{-1}\sum_{i=1}^{N}\{p_i > \lambda\}$, where $\lambda \in (0,1)$ is a tuning parameter.

- Therefore one would expect $N_0(1-\lambda)$ of the p-values to lie within the interval $(\lambda, 1)$ for any sufficiently large λ. Replacing $N\ with\ N_0 < N$ thereby increases the power of the procedure.

- However, this adjustment is not applicable in the context of testing null hypotheses that are inequalities, under which the p-values are no longer uniformly distributed.

Compare with others

- By eliminating the true negatives, the remaining null alphas are close to zero, so that we can safely increase the critical value, and consequently enhance the power of the procedure.
- In a setting similar to ours, Harvey and Liu (2018) propose to increase statistical power by dropping funds that appear only for a small number of periods.
- This approach shares the same spirit as our alpha screening step and is typically used in the literature (e.g., Fung and Hsieh (1997) require at least 36 months of data).

Estimating alpha

- When the benchmark includes non-tradable factors, estimating Equation (1) requires two-pass Fama-MacBeth regressions.

$$r_t = \alpha + \beta\lambda + \beta(f_t - E(f_t)) + u_t$$

- The first stage estimates β using time-series regressions of individual fund returns onto the benchmark factors
- And the second stage involves a cross-sectional regression of average returns onto the estimated β, where the residuals of this regression yield estimates of alpha, denoted as α

Estimating obstacles

- When the benchmark includes non-tradable factors, estimating Equation (1) requires two-pass Fama-MacBeth regressions.

$$r_t = \alpha + \begin{bmatrix} \beta_o & \beta_l \end{bmatrix} \begin{bmatrix} f_{o,t} \\ f_{l,t} \end{bmatrix} + u_t = \underbrace{\alpha + \beta_l \lambda_l}_{\text{“alpha”}} + \beta_o f_{o,t} + \underbrace{\beta_l (f_{l,t} - \mathbb{E} f_{l,t}) + u_t}_{\text{“idiosyncratic” error}}$$

- The latent factors in f_l contribute to the total risk premiums, a bias $\beta_l \lambda_l$ would arise in the estimated “alpha“. Even if f_l is not priced, omitting f_l would still lead to an omitted variable bias for when f_l is correlated with f_o.
- Secondly, f_l plays the role of “idiosyncratic" error.
- Thirdly, leaving f_l in the residuals produces strong correlation among the alpha test statistics, which invalidates the independence assumption.

Observable factors only

- When all factors are observable (not necessarily tradable), we can directly estimate using the classical two-pass regression
- S1a. Run time-series regressions and obtain the OLS estimator β

$$\hat{\beta} = \left(RM_{1_T}F'\right)\left(FM_{1_T}F'\right)^{-1}$$

- S2. Run a cross-sectional regression of $\bar{r}$ on the estimated $\hat{\beta}$ to obtain $\hat{\lambda}$

$$\hat{\lambda} = \left(\hat{\beta}'M_{1_N}\hat{\beta}\right)^{-1}\left(\hat{\beta}'M_{1_N}\bar{r}\right)$$

- S3. Estimate the estimated risk premiums from average returns:

$$\hat{\alpha} = \bar{r} - \hat{\beta}\hat{\lambda}$$

Latent factors only

- In the case that some factors are missing from an observable factor model, the first-step time-series regressions are no longer consistent.
- In this case, we follow Giglio and Xiu (2017) and proceed by rewriting Equation(1) into a statistical factor model:

$$\bar{R} = \beta\bar{V} + \bar{U}$$

- Simply replacing S1a in Algorithm 3 with S1b below leads to a new algorithm for estimating in this scenario:
- S1b. Let $S_R = \frac{1}{T}\bar{R}\bar{R}'$ be the $N \times N$ sample covariance matrix of R. Conduct the principal component analysis of S_R:

Latent factors only

- In the case that some factors are missing from an observable factor model, the first-step time-series regressions are no longer consistent.

- In this case, we follow Giglio and Xiu (2017) and proceed by rewriting Equation(1) into a statistical factor model:

$$\bar{R} = \beta\bar{V} + \bar{U}$$

- Simply replacing S1a in Algorithm 3 with S1b below leads to a new algorithm for estimating in this scenario:

- S1b. Let $S_R = \frac{1}{T}\bar{R}\bar{R}'$ be the $N \times N$ sample covariance matrix of R. Conduct the principal component analysis of S_R:

Latent factors only

- S1b. Let $S_R = \frac{1}{T}\bar{R}\bar{R}'$ be the $N \times N$ sample covariance matrix of R. Conduct the principal component analysis of S_R: set

$$\hat{\beta} = \sqrt{N}(b1, \dots, b_K)$$

- where $b_1, \dots, b_K$ are the K eigenvectors of S_R, corresponding to its largest K eigenvalues.
- S2 & S3 are the same as in Algorithm 3.
- This procedure therefore uses the principal components of returns as factors and uses them as a benchmark to estimate the alphas.

General case

- We now present the most general case, where $f_{o,t}$ is a $K_o \times 1$ vector of observable factors, and $f_{l,t}$ is a $K_l \times 1$ vector of latent factors,

$$r_t = \alpha + \begin{bmatrix} \beta_o & \beta_l \end{bmatrix} \begin{bmatrix} \lambda_o \\ \lambda_l \end{bmatrix} + \begin{bmatrix} \beta_o & \beta_l \end{bmatrix} \begin{bmatrix} f_{o,t} - \mathbb{E} f_{o,t} \\ f_{l,t} - \mathbb{E} f_{l,t} \end{bmatrix} + u_t$$

- To estimate in this case, we combine S1a and S1b, and then proceed with S2 and S3 as in Algorithm 3.
- Specifically, we first obtain $\hat{\beta}_o$ from time-series regressions using observable factors alone, and then obtain $\hat{\beta}_l$ by applying PCA to the covariance matrix of residuals from time-series regressions. The estimated $\hat{\beta}_o$ and $\hat{\beta}_l$ are stacked together as $\hat{\beta}$.

General case

- S1.a. Run time-series regressions and obtain the OLS estimator $\hat{\beta}_o$ and residual matrix Z, where $F_o = (f_{o,1}, f_{o,2}, \dots, f_{o,T})$:

$$\hat{\beta}_o = \left(RM_{1_T}F_o'\right)\left(F_oM_{1_T}F_o'\right)^{-1}, \qquad Z = \bar{R} - \hat{\beta}_o\bar{F}_o$$

- S1. b. Let $S_Z = \frac{1}{T}ZZ'$ be the $N \times N$ sample covariance matrix of Z. Let

$$\hat{\beta}_l = \sqrt{N}(b1, \dots, b_{K_l})$$

- where $b_1, \dots, b_{K_l}$ are the K_l eigenvectors of S_Z, corresponding to its largest K_l eigenvalues. The resulting $\hat{\beta}$ is given by

$$\hat{\beta} = (\hat{\beta}_o, \hat{\beta}_l)$$

General case

- The presence of such bias does not affect the inference for alphas, thanks to the invariance of alpha to the rotation of the factors
- Formally, we can show that $\hat{\beta}_o \xrightarrow{P} \beta_o + \beta_l H_1$ for some matrix H_1, where $\beta_l H_1$ denotes the omitted variable bias. Hence the probability limit of β_o is still spanned by $\beta = (\beta_o, \beta_l)$. As a result, we have established that

$$\hat{\beta} = (\hat{\beta}_o, \hat{\beta}_l) \xrightarrow{P} \beta H$$

- The resulting alpha estimate remains consistent because it is invariant to rotations (the rotation matrix H is canceled with its inverse in $\hat{\lambda}$) and is thus not affected by the omitted variable bias.

Dealing with missing data

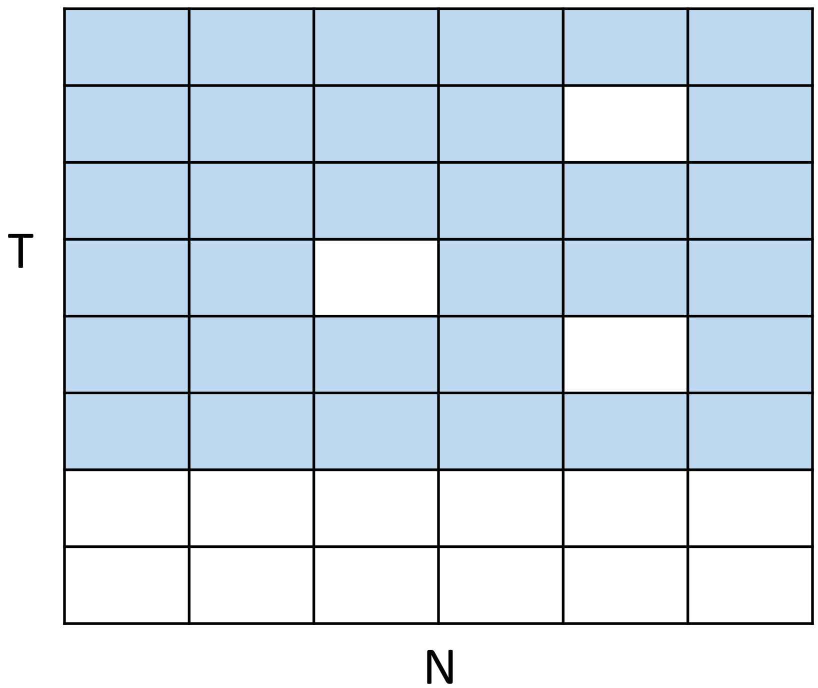
- It is not uncommon in finance applications to deal with unbalanced panels. Many hedge funds last for short periods of time, then liquidate, and many new funds pop up.
- It is therefore important that the estimators we propose work in the presence of missing data. We adopt the matrix completion method from the recent machine learning literature.
- A notable application of this method is the so-called Netflix problem in predicting customers' ratings of movies based on existing ratings. The key assumption behind this algorithm is that the full matrix is approximately low-rank.

Dealing with missing data

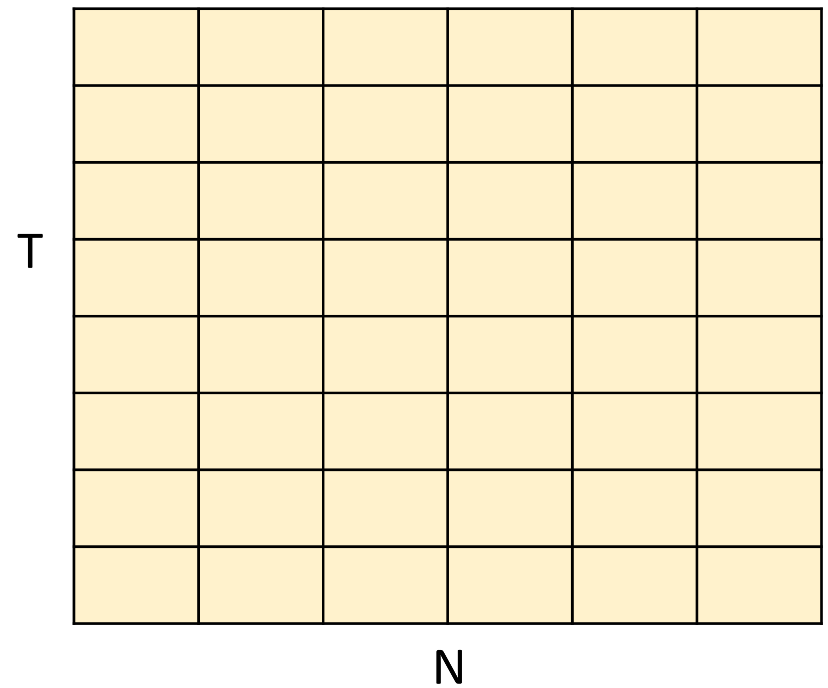
- When data are missing, the common approach in the literature is to adopt an EM algorithm (e.g., Stock and Watson 2002; Su et al. 2019).
- In contrast, our matrix completion approach is much faster; hence it is particularly appealing for large dimensional return matrices.
- The goal is to recover an $N \times T$ low-rank matrix X. Suppose that Z is an $N \times T$ matrix (the “noisy version" of X), which can be written as $Z = X + E$, and E is the noise.

$$\boxed{Z = X + E}$$

Matrix X (low-rank)



Matrix Z (noisy version of X)



Matrix completion

- Suppose Z is not fully observed and Ω is an $N \times T$ matrix whose (i, t)-th element $\omega_{it} = 1\{z_{it}$ is observed$\}$.
- Using this notation, econometricians can only observe $Z \circ \Omega$ and Ω, where $\circ$ represents the element-wise matrix product.
- We employ the following nuclear-norm penalized regression approach to recover X:

$$\hat{X} = arg \min_X \left|\left|(Z - X) \circ \Omega\right|\right|^2 + \lambda_{NT} \left|\left|X\right|\right|_n \ (16)$$

- $\left|\left|X\right|\right|_n$ denotes the matrix nuclear norm and $\lambda_{NT} > 0$ is a tuning parameter. By penalizing the singular values of X, the algorithm achieves a low-rank matrix as the output. The latent factors and betas can then be estimated via the associated singular vectors of $\hat{X}$

Estimate with Matrix completion

- Let N_t denote the set of funds that are observed at time t and T_i denote the collection of time points on which the $i-th$ fund return is observed:
- $N_t = \{i \in \{1, \dots, N\}: r_{it}\ is\ observed\}, \ \ T_i = \{t \in \{1, \dots, T\}: r_{it}\ is\ observed\}$
- We first estimate an observable factor models to calculate the residual matrix Z
- S1. a. Obtain $\hat{\beta}_o$ and the residual matrix $Z = (z_{it})_{N\times T}$:

$$\hat{\beta}_{o,i} = \left(F_{o,i}M_{1_{T_i}}F'_{o,i}\right)^{-1}\left(F_{o,i}M_{1_T}R_i\right)$$

$z_{it} = r_{it} - \bar{r}_i - \hat{\beta}'_{o,i}\left(f_{o,t} - \bar{f}_{o,t}\right)$ $\qquad r_{it}$ is observable; otherwise z_{it} is missing.

- S1. b. Conduct matrix completion, with Z in Equation (16) constructed above, and obtain a low-rank matrix $\hat{X}$

Matrix completion

$$\bar{R} = \beta\bar{V} + \bar{U}$$

- Estimate the latent factors and their loadings using $\hat{X}$:

$$\hat{v}_{l,t} = \left(\sum_{i\in N_t} b_i b_i'\right)^{-1} \sum_{i\in N_t} b_i z_{it}, \qquad t = 1, \dots, T$$

$$\hat{\beta}_{l,t} = \left(\sum_{i\in T_i} \hat{v}_{l,t}\hat{v}_{l,t}'\right)^{-1} \sum_{i\in T_i} \hat{v}_{l,t} z_{it}, \qquad i = 1, \dots, N$$

- The resulting $\hat{\beta}$ is given by $\hat{\beta} = (\hat{\beta}_o, \hat{\beta}_l)$ and the factors $\hat{v}_t = \left(f_{o,t} - \bar{f}_o, \hat{v}_{l,t}'\right)'$, where $\bar{f}_o = \frac{1}{T}\sum_{t=1}^{T} f_{o,t}$, we assume no data are missing for observable factors
- S2 is the same as in Algorithm 3 with inputs $\hat{\beta}$ from above and $\bar{r}_i$, which yields the estimate $\hat{\lambda}$

Matrix completion

- S3. Estimate and de-bias the estimates of α:

$$\hat{\alpha}_i = \bar{r}_i - \hat{\beta}_i' \hat{\lambda} + \hat{A}_i, \qquad i = 1, \ldots, N.$$

- S4. Calculate the standard error as follows:

$$se(\hat{\alpha}_i) = \frac{1}{\sqrt{T_i}} \hat{\sigma}_i, \qquad \hat{\sigma}_i^2 = \frac{1}{T_i} \sum_{t \in T_i} \hat{u}_{it}^2 \left(1 - \hat{v}_i' \hat{\Sigma}_f^{-1} \hat{\lambda}\right)^2$$

Where $\hat{u}_{it} = \mathrm{r_{it}} - \bar{r}_i - \hat{\beta}_i' \hat{v}_t$ is the residual, and $\hat{\Sigma}_f = \frac{1}{T} \sum_{t=1}^{T} \hat{v}_t \hat{v}_t'$

- S5. Calculate the t-statistics and p-values:

$$t_i = \frac{\hat{\alpha}_i}{se(\hat{\alpha}_i)}, p_i = 1 - \Phi(t_i), \ i = 1, \ldots, N$$

Construct p-values with bootstrap

- S0. Generate a bootstrap sample of r_{it}^{*} by resampling weighted residuals:
- S1. Obtain $\hat{\beta}^{*} = (\hat{\beta}_1^{*}, \ldots, \hat{\beta}_N^{*})$:
- S2. use $\bar{r}_i^{*}$ and $\hat{\beta}^{*}$, which yields the estimate $\hat{\lambda}^{*}$.
- S3. Estimate and de-bias the estimates of α:
- S4. Repeat S0-S3 for B times and denote the estimates from S3 as $\{\hat{\alpha}_{i,b}^{*}:$ $i = 1, \ldots, N, b = 1, \ldots, B\}$. Compute the bootstrap p-values as

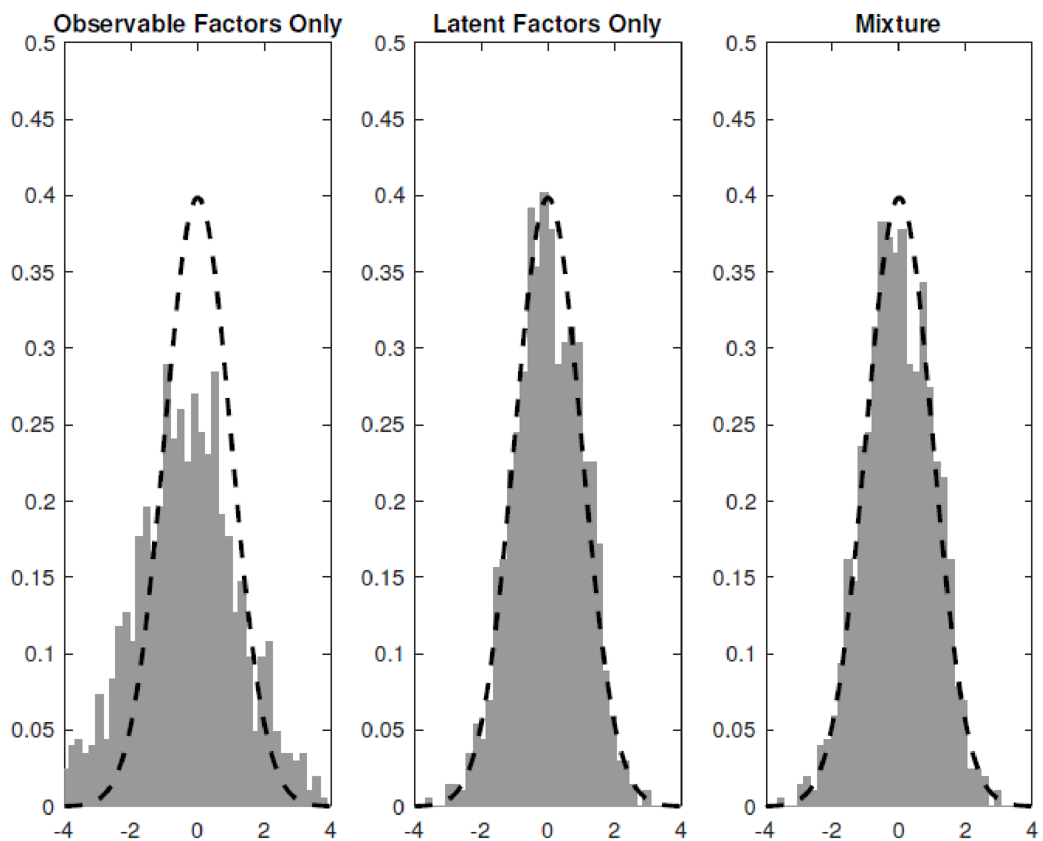
$$p_i = \frac{1}{B} \sum_{b=1}^{B} 1\{\hat{\alpha}_{i,b}^{*} > \hat{\alpha}_i\}, \qquad i = 1, \ldots, N$$

Simulation

- With respect to the data-generating process, we consider a five-factor model for hedge fund returns.
- To examine the impact of missing data on our procedures, we also resample the exact missing pattern of these funds. Throughout the simulations, we fix T = 240 and N = 1000, and on average over 70% of all records are missing.
- We vary the simulated cross-sectional distribution of alphas to check the performance of the FDR, we simulate the alphas from a mixture of two Gaussian distributions, $N(-2\sigma;\ \sigma^2)\ and\ N(2\sigma;\ \sigma^2)$, with mixture probabilities p1 and p2, plus a point mass at zero.

Simulation

- The figure plots the histograms of the standardized alpha estimates for one fixed fund using Algorithms 3 ,4 (latent factors only), and 5 (mixture),



Simulation

Monte Carlo simulation results

			(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
# of observable factors			5	4	0	4	4	4	4	4	4	4	5
# of latent factors			0	0	5	1	1	1	1	0	0	0	0
Missing data								✓	✓		✓	✓	
p_1	p_2												
0.1	0.1	FDR	5.20	8.14	4.81	4.85	5.55	8.36	5.58	47.78	27.83	45.28	35.27
		FDP std.	3.50	15.27	3.23	3.27	3.52	5.11	4.51	6.70	8.69	6.17	7.21
		Avg. power	64.81	53.62	64.62	64.58	65.45	49.85	42.70	60.98	41.17	51.45	79.88
		FNR	3.79	4.95	3.80	3.81	3.72	5.32	6.00	4.44	6.26	5.38	2.31
0.1	0.2	FDR	4.38	6.18	4.12	4.13	4.76	6.55	5.45	33.41	20.99	26.81	19.56
		FDP std.	2.48	11.24	2.41	2.43	2.61	3.44	3.18	4.94	5.32	4.73	5.31
		Avg. power	69.02	57.68	68.78	68.72	69.63	54.70	51.72	64.19	47.44	51.78	80.43
		FNR	7.15	9.53	7.20	7.21	7.03	10.14	10.72	8.77	11.81	11.11	4.83
0.1	0.3	FDR	3.67	4.88	3.44	3.44	3.92	5.19	4.59	24.62	16.07	17.91	12.31
		FDP std.	2.02	8.56	1.90	1.93	2.02	2.53	2.42	3.87	3.80	4.09	3.84
		Avg. power	71.29	60.14	71.10	71.03	71.96	57.47	55.48	66.18	50.39	52.25	80.43
		FNR	10.86	14.44	10.92	10.94	10.64	15.32	15.89	13.52	17.84	17.40	7.94
0.2	0.1	FDR	4.50	7.24	4.17	4.21	5.31	7.87	5.39	43.60	24.84	42.10	32.12
		FDP std.	3.33	14.22	3.01	3.07	3.52	4.92	4.62	6.81	8.35	6.22	7.44
		Avg. power	63.40	52.36	63.18	63.14	64.55	49.28	43.17	59.53	40.04	50.63	78.54
		FNR	4.03	5.19	4.05	4.05	3.91	5.50	6.10	4.66	6.50	5.57	2.50
0.2	0.2	FDR	3.79	5.41	3.53	3.53	4.42	5.97	5.13	29.77	18.26	24.06	17.29
		FDP std.	2.32	10.20	2.20	2.22	2.51	3.18	3.00	4.76	4.98	4.57	5.16
		Avg. power	68.15	56.97	67.92	67.84	69.33	54.56	51.90	63.35	46.67	51.35	79.66
		FNR	7.43	9.79	7.48	7.49	7.18	10.30	10.82	8.96	12.06	11.28	5.05
0.3	0.1	FDR	3.90	6.46	3.63	3.64	5.04	7.11	5.08	39.19	21.89	38.58	28.93
		FDP std.	3.02	13.11	2.80	2.83	3.40	4.70	4.43	6.90	8.05	6.28	7.34
		Avg. power	62.07	51.28	61.85	61.84	63.69	48.75	43.47	58.12	38.90	49.82	77.25
		FNR	4.27	5.42	4.29	4.29	4.09	5.68	6.22	4.88	6.75	5.76	2.70

Simulation

- First of all, we find it critical to use alpha tests that take into account omitted factors, by comparing columns c and d with column b.
- Second, comparing columns d and e, we find alpha screening less conservative and more powerful, when the unskilled funds p1, is large.
- Third, the standard bootstrap method in various scenarios does not cope well with missing data and missing factors.
- Fourth, columns b, h, i, and j clearly show that missing a factor tends to increase the standard errors of the FDP.
- Finally, without any B-H type control, the false discovery rate can exceed 35% among the experiments we consider even in the most ideal setting

Hedge fund returns data

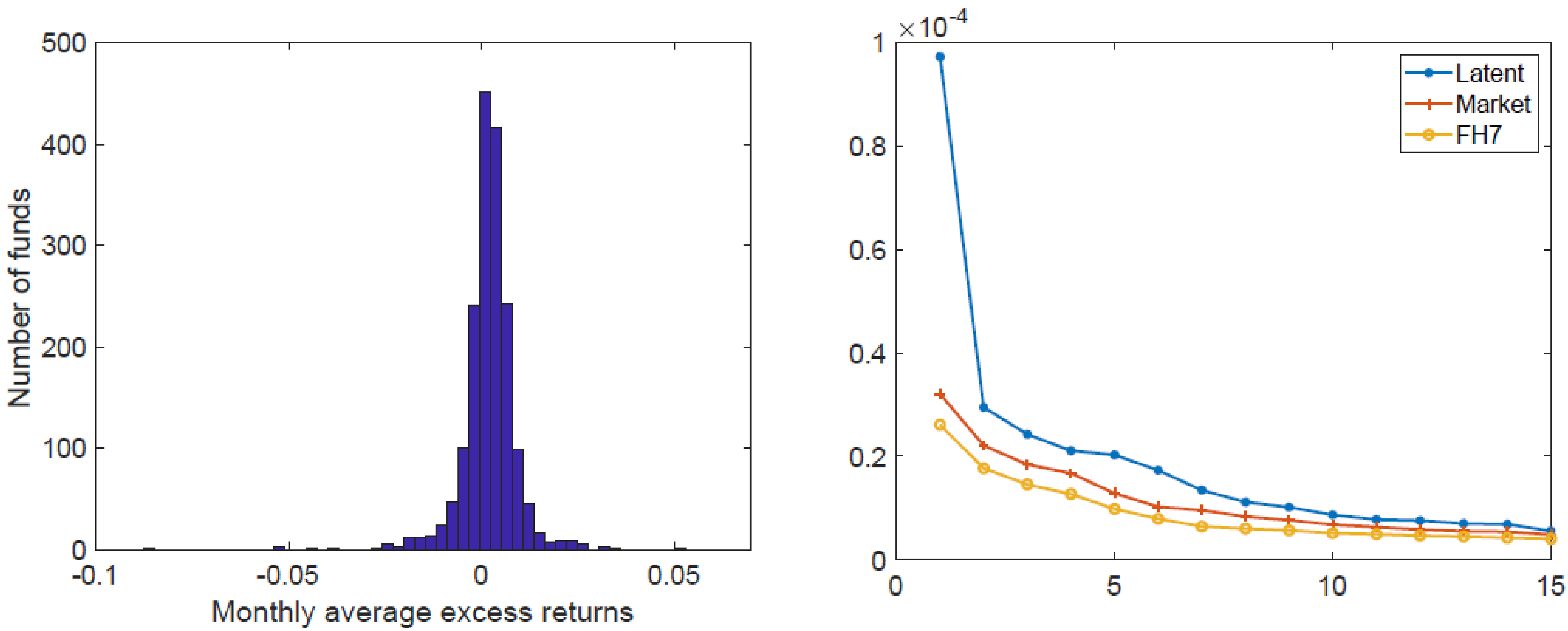
- We apply our methodology to the Lipper TASS hedge funds data set, covering the time period 1994-2018. We follow closely the bias correction and data-cleaning procedures of Sinclair (2018)
- As is standard in this literature, we only focus on funds that have a sufficiently long time series. Based on our simulations, we choose a minimum period of 36 months.
- After applying these filters, we are left with 1,761 funds in our data set.

Benchmark model

- We consider two standard benchmark models. Our baseline model will be the Fung and Hsieh (2004)seven-factor model.
- The model includes market, size, a bond factor, a credit risk factor, and three trend-following factors(related to bonds, currencies & commodities).
- As an alternative, we also consider the Agarwal and Naik (2004) model
- Which includes the Fama-French-Carhart four factors (market, size, value, and momentum factors) plus two option-based factors (an out-of-the-money call and an out-of-the-money put factor).

Hedge fund returns data

- To get a sense of the factor structure of hedge fund returns, the blue line in panel B of Figure 2 shows the first 15 eigenvalues of the excess returns in our panel.



A. Average excess returns: Histogram

B. Scree plot of eigenvalues

In-sample analysis

- We begin with an in-sample analysis in which we compare the funds selected by our FDR control methodology to those selected using different methodologies.

In-sample results

	Mixed FDR		Only observable			No screen	No screen	Only latent	Alternative models	
	Bootstrap	Avar	FDR	$p<.05$	$p>.05$	FDR	observ. FDR	FDR	HL	BSW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Average alpha	70.4	68.7	66.2	62.9	4.8	70.6	66.4	61.4	65.8	65.4
Average t-stat	4.6	4.2	5.8	4.3	.3	4.7	5.9	4.3	18.8	5.4
Fraction selected	.19	.24	.10	.20	.80	.18	.10	.19	.01	.12
p-val. alpha = 0	$<.01$	$<.01$	$<.01$	$<.01$	.04	$<.01$	$<.01$	$<.01$	$<.01$	$<.01$
p-val., alpha = (1)	-	.04	$<.01$	$<.01$	$<.01$	.66	$<.01$	$<.01$	.05	$<.01$

Out-of-sample analysis

- The baseline (first row) uses yearly rebalancing, three latent factors, and funds that are alive for at least 36 months. The other robustness tests, add monthly rebalancing,, We next consider a battery of additional robustness tests for our baseline specification:

  a) Add a 1 or 3-month lag between portfolio formation and evaluation
  b) Restricting the estimation to funds with data for at least 5 years;
  c) Using five instead of three latent factors;
  d) Using the Agarwal and Naik (2004) benchmark model (Fama-French-Carhart four factors plus two option factors);
  e) Using the Evestment data, an entirely different hedge fund data set;
  f) Using equal-weighted instead of value-weighted alphas.

Out-of-sample analysis

A. Alphas

	Mixed FDR Bootstrap	Mixed FDR Avar	Only observable FDR	Only observable $p<.05$	Only observable $p>.05$	No screen FDR	No screen Observ. FDR	Only latent FDR	Alternative models HL	Alternative models BSW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Baseline	22.3	21.0	15.1	10.7	12.5	21.8	16.8	13.8	10.4	11.6
Yearly rebal., 1-month lag	22.6	20.8	18.2	12.4	12.8	22.3	17.9	14.2	8.9	14.3
Yearly rebal., 3-month lag	20.9	20.7	17.9	12.2	13.4	20.5	18.7	13.7	9.0	12.0
Monthly rebal.	27.3	26.7	21.9	18.7	14.7	26.9	18.2	23.1	43.5	21.7
Monthly rebal., 1-month lag	25.9	22.6	20.5	16.0	12.8	26.1	21.7	21.1	33.2	20.1
Monthly rebal., 3-month lag	24.9	21.6	19.3	15.5	12.5	24.4	20.8	20.3	–16.1	19.2
Min 60 months	25.1	24.6	18.3	15.7	13.4	26.6	19.6	16.9	14.3	15.7
5 latent factors	20.5	18.3	15.5	10.5	12.3	20.4	18.0	13.1	11.4	10.7
Agarwal and Naik (2004)	22.5	17.0	18.4	2.6	17.0	22.0	18.1	11.3	–27.6	9.2
Evestment, annual rebal.	24.7	20.7	7.4	13.9	16.3	24.6	8.2	17.6	–10.0	14.0
Evestment, monthly rebal.	35.5	39.4	18.6	32.6	3.5	35.8	16.0	25.5	16.0	25.6
Only U.S.	25.4	21.6	15.4	10.3	16.0	24.8	15.4	11.9	10.7	14.4
Equal-weighted returns	25.7	25.0	19.8	16.6	6.8	25.4	19.9	20.8	18.9	19.6

B. p-values

	Mixed FDR Bootstrap	Mixed FDR Avar	Only observable FDR	Only observable $p<.05$	Only observable $p>.05$	No screen FDR	No screen Observ. FDR	Only latent FDR	Alternative models HL	Alternative models BSW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Baseline	.02	.04	.12	.29	.30	.02	.08	.18	.40	.24
Yearly rebal., 1-month lag	.01	.04	.05	.20	.30	.02	.06	.16	.47	.13
Yearly rebal., 3-month lag	.02	.04	.06	.20	.28	.02	.05	.18	.47	.21
Monthly rebal.	$<.01$	.01	.03	.06	.25	$<.01$	.07	.02	$<.01$	.02
Monthly rebal., 1-month lag	$<.01$	.04	.05	.11	.33	.01	.03	.05	$<.01$	.04
Monthly rebal., 3-month lag	.01	.05	.06	.12	.35	.02	.04	.06	.70	.05
Min 60 months	.01	.02	.07	.11	.30	$<.01$	.06	.11	.10	.11
5 latent factors	.03	.07	.11	.30	.31	.03	.06	.20	.36	.28
Agarwal and Naik (2004)	.07	.19	.10	.83	.27	.07	.11	.38	.65	.45
Evestment, annual rebal.	.16	.23	.66	.49	.39	.17	.63	.29	.52	.48
Evestment, monthly rebal.	.05	.03	.31	.10	.88	.05	.38	.13	.32	.22
Only U.S.	$<.01$	.03	.10	.30	.16	$<.01$	.10	.23	.29	.14
Equal-weighted returns	.01	.02	.06	.14	.57	.01	.06	.07	.04	.07

Out-of-sample analysis

Out-of-sample results: Selection

A. Fraction selected

	Mixed FDR Bootstrap	Mixed FDR Avar	Only observable FDR	Only observable $p < .05$	Only observable $p > .05$	No α-scr. FDR	No α-scr. Obs. FDR	Only lat. FDR	Alt. models HL	Alt. models BSW	# funds used	# funds alive
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Baseline	.25	.27	.17	.29	.71	.25	.17	.33	.06	.20	982.8	408.0
Yearly rebal., 1-month lag	.25	.27	.17	.29	.71	.25	.17	.33	.06	.20	982.8	408.0
Yearly rebal., 3-month lag	.25	.27	.17	.29	.71	.25	.17	.33	.06	.20	982.8	408.0
Monthly rebal.	.29	.31	.17	.29	.71	.28	.23	.35	.07	.20	475.8	391.7
Monthly rebal., 1-month lag	.28	.31	.17	.29	.71	.28	.17	.35	.07	.20	958.7	392.7
Monthly rebal., 3-month lag	.28	.30	.17	.29	.71	.27	.17	.35	.06	.20	959.4	394.3
Min 60 months	.37	.38	.23	.35	.65	.37	.17	.44	.15	.30	960.5	246.1
5 latent factors	.27	.29	.17	.29	.71	.26	.17	.35	.06	.20	982.8	408.0
Agarwal and Naik (2004)	.23	.27	.08	.22	.78	.22	.08	.28	.01	.12	1066.8	442.0
Evestment, annual rebal.	.45	.45	.36	.47	.53	.45	.35	.58	.22	.53	297.2	244.8
Evestment, monthly rebal.	.46	.45	.38	.48	.52	.46	.38	.55	.23	.57	274.5	230.3
Only U.S.	.31	.34	.22	.34	.66	.30	.21	.43	.07	.23	784.4	326.1
Equal-weighted returns	.26	.27	.17	.29	.71	.25	.17	.33	.06	.20	982.8	408.0

B. Average AUM

	Mixed FDR Bootstrap	Mixed FDR Avar	Only observable FDR	Only observable $p < .05$	Only observable $p > .05$	No screen FDR	No screen Obs. FDR	Only latent FDR	Alternative models HL	Alternative models BSW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Baseline	53.4	57.8	43.1	63.1	55.9	52.8	41.0	65.6	11.1	47.1
Yearly rebal., 1-month lag	52.6	57.9	42.0	63.1	55.9	51.9	41.6	65.5	11.1	47.2
Yearly rebal., 3-month lag	52.4	57.9	42.8	62.7	56.3	51.8	40.5	65.6	11.0	47.4
Monthly rebal.	55.7	61.0	42.4	62.7	55.6	55.0	32.8	66.1	11.0	47.1
Monthly rebal., 1-month lag	55.5	60.9	42.3	63.0	55.5	54.8	41.4	65.9	11.2	47.8
Monthly rebal., 3-month lag	55.3	60.6	42.7	63.1	55.7	54.6	41.3	65.9	11.1	48.1
Min 60 months	46.2	47.9	33.0	46.7	32.2	46.0	41.4	51.2	17.1	40.8
5 latent factors	56.5	60.7	42.7	63.8	55.2	55.5	40.5	67.8	11.1	47.7
Agarwal and Naik (2004)	52.5	61.1	25.1	58.2	66.0	52.2	24.5	61.1	1.1	37.3
Evestment, annual rebal.	177.4	174.9	129.0	172.8	135.5	176.9	123.7	206.9	73.3	169.0
Evestment, monthly rebal.	162.6	161.1	118.7	156.1	137.6	162.1	116.1	192.3	67.9	164.8
Only U.S.	51.5	56.7	43.8	62.2	46.6	50.7	42.9	67.6	12.5	45.7

Conclusion

- This paper presents a rigorous framework to address the data-snooping concerns that arise when applying multiple testing in the asset pricing context.
- Our paper builds an FDR control test that is valid when the benchmark includes nontradable, and is robust to the presence of omitted factors and an unbalanced data panel, which makes it particularly suitable for many finance applications.
- We also illustrate this procedure by applying it to the evaluation of hedge fund performance.

Consideration

- This paper presents a rigorous framework to address the data-snooping concerns that arise when applying multiple testing in the asset pricing context.
- Our paper builds an FDR control test that is valid when the benchmark includes nontradable, and is robust to the presence of omitted factors and an unbalanced data panel, which makes it particularly suitable for many finance applications.
- We also illustrate this procedure by applying it to the evaluation of hedge fund performance.