Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power? A Reply

Laurent Barras, Olivier Scaillet, and Russ Wermers*

This version: October 24, 2019

JEL Classification: C11, G12, G23
Keywords: False Discovery Rate, Multiple Testing, Mutual Fund Performance

ABSTRACT

Andrikogiannopoulou and Papakonstantinou (AP; 2019) conduct an inquiry into the bias of the False Discovery Rate (FDR) estimators of Barras, Scaillet, and Wermers (BSW; 2010). In this Reply, we replicate their results, then further explore the bias issue by (i) using different parameter values, and (ii) updating the sample period. Over the original period (1975-2006), we show how reasonable adjustments to the parameter choices made by BSW and AP results in a sizeable reduction in the bias relative to AP. Over the updated period (1975-2018), we further show that the performance of the FDR improves dramatically across a large range of parameter values. Specifically, we find that the probability of misclassifying a fund with a true alpha of 2% per year is 32% (versus 65% in AP). Our results, in combination with those of AP, indicate that the use of the FDR in finance should be accompanied by a careful evaluation of the underlying data generating process, especially when the sample size is small.

*Barras is at McGill University (Desautels Faculty of Management), Scaillet is at the University of Geneva (Geneva Finance Research Institute (GFRI)) and at the Swiss Finance Institute (SFI), and Wermers is at the University of Maryland (Smith School of Business). We thank the Editor, Stefan Nagel, and two anonymous referees for very helpful feedback.

I Introduction

In a paper published in the Journal of Finance, Barras, Scaillet, and Wermers (2010; henceforth, BSW) apply a new econometric approach—the False Discovery Rate (FDR)—to the field of mutual fund performance. BSW use this approach with two main objectives in mind: (i) estimate the proportions, π_0 and π_A, of zero and non-zero alpha funds in the population, and (ii) form portfolios of funds which differ in their ability to generate future (out-of-sample) positive alphas.[1]

It is well known from theory that the FDR estimators, $\hat{\pi}_0$ and $\hat{\pi}_A$, are potentially biased, as $\hat{\pi}_0$ overestimates π_0, and $\hat{\pi}_A$ underestimates π_A (e.g., Genovese and Wasserman (2004), Storey (2002)). This issue arises when non-zero alpha funds are difficult to detect in the data, either because their alphas are small, or the estimation noise is large. Whereas using a conservative estimator, $\hat{\pi}_0$, is beneficial for the selection of a subset of truly outperforming funds (objective (ii) of BSW), it may result in an undesirable level of bias in the evaluation of performance in the overall fund population (objective (i) of BSW). To quantify this bias, BSW perform a simulation analysis using a set of parameter values, and conclude that both $\hat{\pi}_0$ and $\hat{\pi}_A$ are close to their true values.

Andrikogiannopoulou and Papakonstantinou (2019; henceforth, AP) challenge the simulation analysis of BSW based on their choice of parameter values. First, they correctly point out that the number of assumed fund return observations is set too high by BSW, which artificially reduces the estimation noise compared to the actual sample. Second, they consider smaller (in magnitude) values for the true alphas of non-zero alpha funds. Incorporating these changes, AP conclude, from simulations based on revised parameter values, that the proportions $\hat{\pi}_0$ and $\hat{\pi}_A$ are heavily biased. Thus, AP question the usefulness of the FDR for evaluating the performance of mutual funds (objective (i) of BSW).

We recognize that AP conduct an important critical evaluation of the FDR approach for the benefit of empirical researchers. The FDR has been increasingly used as an estimation approach in mutual fund performance and other areas of finance because it is simple and fast—simply put, it amounts to estimating a simple average based on t-statistics.[2] This contrasts with alternative approaches which impose much more structure to improve estimation performance. For instance, Bayesian/parametric approaches

[1] Henceforth, "zero-alpha" ("non-zero alpha") refers to the set of funds that have a *true* model alpha equal to zero (different from zero).

[2] See, among others, Augustin, Brenner, and Subrahmanyan (2019), Bajgrowicz and Scaillet (2012), Barras (2019), Cuthbertson, Nitzsche, and O'Sullivan (2008), Ferson and Chen (2019), Giglio, Liao, and Xiu (2018), Scaillet, Treccani, and Trevisan (2019).

require a numerical, computer-intensive estimation of π_0 and π_A using sophisticated Markov Chain Monte Carlo (MCMC) and Expectation-Maximisation (EM) methods. If the FDR approach is to be discarded, the costs to researchers in terms of time and complexity are potentially large.[3]

In this Reply to AP, we pursue some simple goals. First, we seek to replicate the results of BSW and AP. Second, we wish to provide further extensions of the BSW and AP simulation results by (i) using parameter values that we believe better reflect the data generating process of mutual fund alphas in the real-world, and (ii) extending their original sample period (1975-2006). In conducting this analysis, we develop an analytical approach that allows us to quantify the bias of the FDR approach without recourse to simulations.[4] This approach is simple and allows for a completely transparent comparison of the results in BSW, AP, and our Reply.

Our results reveal that the changes in parameter values materially affect the evaluation of the FDR approach. Over the original period (1975-2006), we find that the bias decreases significantly relative to AP once we (i) consider alternative values for the fund residual volatility, and (ii) account for the relations between the different parameters, which are motivated by theory and supported by empirical evidence. To illustrate, AP find that a fund with a true alpha of 2% per year is misclassified as a zero alpha fund 65% of the time. We instead find that the misclassification probability is equal to 44%. Therefore, the evidence documented here is more nuanced than in BSW and AP. Whereas we confirm that the initial analysis of BSW may be too optimistic, we do not find the high levels of bias documented by AP.

Next, we provide researchers with updated information about the bias by examining the period 1975 to 2018 (relative to the original period 1975-2006 studied by BSW and AP). Over this updated period, the performance of the FDR improves significantly, regardless of the procedure for choosing the parameter values. Using the same procedure as in BSW and AP, we find that the probability of misclassifying a fund with an alpha of 2% per year drops to 37%. With the procedure proposed here, the misclassification probability further drops to 32%. Therefore, we believe that the FDR approach does a

[3] If the FDR is discarded, researchers also lose the benefits of the FDR as a nonparametric approach. Specifically, the FDR makes no assumptions on the shape of the alpha distributions of negative/positive alpha funds. Therefore, it is less susceptible to potential misspecification errors that arise when researchers make incorrect specification assumptions about the shape of these distributions.

[4] To explain, we build on the analysis of BSW and AP, making the same assumptions regarding the Data Generating Process (DGP) for the cross-section of mutual fund returns. Given these assumptions, we show that a simulation analysis is unnecessary—instead, we use the properties of the Student t-distribution to derive exact expressions for the probability that the FDR misclassifies non-zero alpha funds, and the bias of the FDR proportions $\hat{\pi}_0$ and $\hat{\pi}_A$.

good job at capturing the salient features of the cross-sectional alpha distribution.

Our Reply has general implications beyond the context of mutual fund performance evaluation. First, it highlights the importance of carefully analyzing the properties of the data to assess the benefits of the FDR. This analysis is particularly important if the sample period is relatively short. Second, it provides a simple approach for evaluating the bias of the FDR without recourse to simulations—a definite benefit to empiricists who wish to explore the performance of the FDR approach for their empirical application.

The remainder of our reply is as follows. Section II presents the general context behind the critique of AP. Section III present the methodology for computing the bias of the FDR proportions. Section IV replicates the results of BSW and AP. Section V presents our analysis and summarize our main findings. The appendix provides additional information about the methodology.

II The Context

To place our Reply to AP in context, we first provide a brief background. Specifically, we first discuss the applications of the FDR approach to mutual fund performance, as specified in BSW, including the procedure for estimating the FDR proportions of zero alpha and non-zero alpha funds. We then provide a review of the bias of the FDR proportions.

A The FDR Approach

A.1 Applications of the FDR Approach

As explained in their introduction (p. 180), BSW apply the FDR approach to a large panel of mutual funds to meet two objectives: (i) estimate the proportions, π_0 and π_A, of zero alpha and non-zero alpha funds; (ii) select subgroups of funds, based on these proportion estimates, to determine whether mutual fund performance persists out-of-sample.[5]

For the first objective, BSW use the FDR as an estimation approach. This approach proposed by Storey (2002) is straightforward—it uses as only inputs the p-values of the fund alphas obtained from a regression of the individual fund returns on the benchmark returns. Whereas the p-values of non-zero alpha funds are typically close to zero because

[5] In the first paragraph of the introduction (p. 180), BSW summarize these two objectives as follows: "... It is natural to wonder how many fund managers possess true stockpicking skills, and where these funds are located in the cross-sectional estimated alpha distribution. From an investment perspective, precisely locating skilled funds maximizes our chances of achieving persistent outperformance."

their alphas are different from zero, the p-values of zero alpha funds generally take large values. By choosing a small interval, I_p, close to one, we can, therefore, bifurcate the two categories of funds—zero and non-zero alpha funds—and estimate the FDR proportions π_0 and π_A (see Figure 2 of BSW (p. 188) for a graphical explanation).

For the second objective, BSW use the FDR as a multiple testing approach. The basic idea is to conduct a test of the null hypothesis for each fund i, $H_{0,i}$: $\alpha_i = 0$ (for $i = 1, 2, ...$), to select funds with positive alphas. When conducting this multiple testing approach among several thousand funds, we are likely to have many false discoveries—true zero alpha funds which, by chance, exhibit non-zero estimated alphas of a significant magnitude. Using the FDR allows us to explicitly control for these false discoveries. In other words, it is the extension of the notion of Type I error from single to multiple hypothesis testing.

The critique of AP pertains to the first objective of BSW. Therefore, we now exclusively focus on the properties of the FDR proportions, and refer the reader interested in multiple testing (including its application to finance) to the extensive literature devoted to this issue (see, among others, Bajgrowicz and Scaillet (2012), Bajgrowicz, Scaillet, and Treccani (2016), BSW, Benjamini and Hochberg (1995), Efron (2010), Storey (2002)).

A.2 The FDR Proportions

To describe the procedure for estimating the FDR proportions, we denote by $P[I_p(\lambda)]$ the probability that the fund p-value falls in the interval $I_p(\lambda) = [\lambda, 1]$, where λ denotes the interval lower bound. We can write $P[I_p(\lambda)]$ as a weighted average: $P[I_p(\lambda)] = \pi_0 P_0[I_p(\lambda)] + \pi_A P_A[I_p(\lambda)]$, where $P_0[I_p(\lambda)]$ and $P_A[I_p(\lambda)]$ denote the probabilities that the p-values of truly zero alpha and non-zero alpha funds fall in $I_p(\lambda)$. Because the p-values of non-zero alpha funds are typically close to zero, most of them do not fall in $I_p(\lambda)$. Building on this insight, we can set $P_A[I_p(\lambda)] = 0$ to obtain the estimator of π_0 proposed by Storey (2002) and used by BSW (p. 188):

$$\hat{\pi}_0(\lambda) = \frac{\hat{P}[I_p(\lambda)]}{P_0[I_p(\lambda)]} = \frac{\frac{1}{N}\sum_{i=1}^{N} \mathit{1}\{p_i \in I_p(\lambda)\}}{1-\lambda}, \tag{1}$$

where the numerator $\hat{P}[I_p(\lambda)] = \frac{1}{N}\sum_{i=1}^{N} \mathit{1}\{p_i \in I_p(\lambda)\}$ is the empirical counterpart of $P[I_p(\lambda)]$, N is the total number of funds, $\mathit{1}\{p_i \in I_p(\lambda)\}$ is an indicator function equal to 1 if the p-value of fund i falls in $I_p(\lambda)$, and zero otherwise. The denominator $P_0[I_p(\lambda)]$ is equal to $1 - \lambda$ because the p-values of zero alpha funds follow a uniform distribution.

In this Reply, we propose a numerically equivalent, yet simpler approach for computing $\hat{\pi}_0(\lambda)$ which replaces the p-values with the t-statistics of the fund alphas (see Barras (2019), Efron (2010)).[6] We denote by $I(\lambda) = [-a(\lambda), a(\lambda)]$ an interval centered around zero with bounds equal to $\pm a(\lambda)$, and by $P_0[I(\lambda)]$, and $P_A[I(\lambda)]$ the probabilities that the t-statistics of zero alpha and non-zero alpha funds fall in $I(\lambda)$. We set the bounds $\pm a(\lambda)$ such that $P_0[I(\lambda)]$ remains equal to $P_0[I_p(\lambda)] = 1 - \lambda$. In other words, the bounds $\pm a(\lambda)$ correspond to the quantiles at $\lambda/2$ and $1 - \lambda/2$ of the t-statistic distribution for the zero alpha funds.[7] Because the t-statistics of non-zero alpha funds are typically far from zero, most of them do not fall in $I(\lambda)$. Setting $P_A[I(\lambda)] = 0$, we can therefore rewrite $\hat{\pi}_0(\lambda)$ in Equation (1) as

$$\hat{\pi}_0(\lambda) = \frac{\hat{P}[I(\lambda)]}{P_0[I(\lambda)]} = \frac{\frac{1}{N}\sum_{i=1}^{N} \mathit{1}\{t_i \in I(\lambda)\}}{P_0[I(\lambda)]}, \tag{2}$$

where $\hat{P}[I(\lambda)] = \frac{1}{N}\sum_{i=1}^{N} \mathit{1}\{t_i \in I(\lambda)\}$, and $\mathit{1}\{t_i \in I(\lambda)\}$ is an indicator function equal to one if t_i falls in the interval $I(\lambda)$, and zero otherwise. Similarly, we compute the proportion of non-zero alpha funds $\hat{\pi}_A(\lambda) = 1 - \hat{\pi}_0(\lambda)$ as

$$\hat{\pi}_A(\lambda) = \hat{P}[I_n(\lambda)] - \hat{\pi}_0(\lambda)P_0[I_n(\lambda)] = \frac{1}{N}\sum_{i=1}^{N} \mathit{1}\{t_i \in I_n(\lambda)\} - \hat{\pi}_0(\lambda)P_0[I_n(\lambda)], \tag{3}$$

where the interval $I_n(\lambda)$ is equal to $[-\infty, -a(\lambda)] \cup [a(\lambda), +\infty]$. Furthermore, we can split $\hat{\pi}_A(\lambda)$ to estimate the proportions of funds with negative/positive alphas, π_A^- and π_A^+, as:

$$\begin{aligned} \hat{\pi}_A^-(\lambda) &= \hat{P}[I_n^-(\lambda)] - \hat{\pi}_0(\lambda)P_0[I_n^-(\lambda)] = \frac{1}{N}\sum_{i=1}^{N} \mathit{1}\{t_i \in I_n^-(\lambda)\} - \hat{\pi}_0(\lambda)P_0[I_n^-(\lambda)], \quad (4) \\ \hat{\pi}_A^+(\lambda) &= \hat{P}[I_n^+(\lambda)] - \hat{\pi}_0(\lambda)P_0[I_n^+(\lambda)] = \frac{1}{N}\sum_{i=1}^{N} \mathit{1}\{t_i \in I_n^+(\lambda)\} - \hat{\pi}_0(\lambda)P_0[I_n^+(\lambda)], \quad (5) \end{aligned}$$

where the intervals $I_n^-(\lambda)$ and $I_n^+(\lambda))$ are defined as $[-\infty, -a(\lambda)]$ and $[a(\lambda), +\infty]$. These two estimators are the same as those used by BSW (p. 189), except that we use a t-statistic formulation (instead of a p-value formulation).

[6] A t-statistic formulation is also used, among others, by Harvey, Liu, and Zhu (2016) to determine the appropriate significance threshold in multiple hypothesis tests of factor risk premia.

[7] The bounds $\pm a(\lambda)$ are easy to compute after specifying the t-statistic distribution for the zero alpha funds. In our bias analysis presented below, it is given by a Student t-distribution (Equation (11)).

B The Bias of the FDR Proportions

B.1 Theoretical Analysis

At the heart of the AP critique is the potential bias of the estimated FDR proportions $\hat{\pi}_0$ and $\hat{\pi}_A$. As discussed by Barras (2019), we can formalize this point using Equations (2)-(3). Suppose that the t-statistics for some non-zero alpha funds turn out to be close to zero, either because their true alphas are small, the estimation noise is large, or both. In this case, we have $P_A[I(\lambda)] > 0$, instead of $P_A[I(\lambda)] = 0$. Therefore, we can show that, on average, $\hat{\pi}_0(\lambda)$ overestimates π_0, and $\hat{\pi}_A(\lambda)$ underestimates π_A:

$$\begin{aligned} E[\hat{\pi}_0(\lambda)] &= \frac{E[\hat{P}[I(\lambda)]]}{P_0[I(\lambda)]} = \pi_0 + \delta(\lambda)\pi_A > \pi_0, \quad (6) \\ E[\hat{\pi}_A(\lambda)] &= E[\hat{P}[I_n(\lambda)]] - E[\hat{\pi}_0(\lambda)]P_0[I_n(\lambda)] = \pi_A - \delta(\lambda)\pi_A < \pi_A, \quad (7) \end{aligned}$$

where we define the ratio $\delta(\lambda)$ as

$$\delta(\lambda) = \frac{P_A[I(\lambda)]}{P_0[I(\lambda)]} = \frac{P_A[I(\lambda)]}{1-\lambda}. \quad (8)$$

Taking the expectations of Equations (4)-(5), we can further show that $\hat{\pi}_A^-(\lambda)$ and $\hat{\pi}_A^+(\lambda)$, are also biased (see the appendix).

We can write the ratio $\delta(\lambda)$ as the relative bias of $\hat{\pi}_A(\lambda)$, i.e.,

$$\delta(\lambda) = \frac{\pi_A - E[\hat{\pi}_A(\lambda)]}{\pi_A}. \quad (9)$$

Equation (9) implies that we can interpret $\delta(\lambda)$ as the misclassification probability associated with the FDR approach. For instance, a value of 40% for $\delta(\lambda)$ implies a 40% probability that non-zero alpha funds are incorrectly classified as zero alpha funds. Because $\delta(\lambda)$ is a relative measure, we can compute it without having to specify the true proportions π_0, π_A^-, and π_A^+. For this reason, $\delta(\lambda)$ provides a convenient measure of the bias implied by the FDR approach.

The potential bias of the estimated FDR proportions has been known for some time—it is extensively discussed in the statistical papers cited by BSW (Genovese and Wasserman (2004), Storey (2002), Storey, Taylor, and Sigmund (2004)). In addition, several finance papers explicitly discuss the potential bias of the FDR proportions, and propose alternative approaches that impose more parametric assumptions on the fund alpha distribution (e.g., Chen, Cliff, and Zhao (2017), Ferson and Chen (2019), Harvey and Liu (2018)).[8] However, these papers provide limited information about the magnitude of the bias in the context of mutual funds.[9]

B.2 Quantitative Analysis

Two studies provide a quantitative assessment of the bias of the FDR proportions. BSW performs a simulation analysis calibrated on the data over the period 1975-2006. The results reported in the appendix of BSW reveal that both the bias and the variance of $\hat{\pi}_0(\lambda)$ and $\hat{\pi}_A(\lambda)$ are small (p. 4 (appendix)): "The simulation results reveal that the average values of our estimators closely match the true values, and that their 90% confidence intervals are narrow".

More recently, AP perform another simulation analysis using the same period 1975-2006. They consider different scenarios in which funds have different assumed alphas and/or shorter return time-series than those examined in BSW. Contrary to BSW, AP find that $\hat{\pi}_0(\lambda)$ and $\hat{\pi}_A(\lambda)$ are heavily biased. Motivated by these results, AP question the applicability of the FDR approach for evaluating performance of the mutual fund industry (p. 2): "Overall, our results raise concerns about the applicability of the FDR in fund performance evaluation and more widely in finance where the signal-to-noise ratio in the data is similarly low".

These two studies reach different conclusions regarding the bias of the FDR proportions. To understand the reasons for these large differences and explain our views on the properties of the FDR estimated proportions, we propose a novel methodology that allows for a simple and transparent analysis of the bias.

III Methodology for Computing the Bias

To present our methodology, we first discuss the main distributional assumptions we make on the mutual fund returns. Building on these assumptions, we then propose a simple analytical approach to compute the bias without recourse to simulations.

[8] If these specification assumptions are correct, the proposed estimators achieve a better performance than the FDR estimators. However, if they are incorrect, the proposed estimators are plagued by misspecification errors and could potentially be more biased than the FDR estimators.

[9] One exception is the paper by Ferson and Chen (2019) which performs a simulation analysis to examine the properties of the FDR estimators. However, the results are difficult to compare with those in BSW and AP, mainly because the simulations are calibrated on hedge fund data.

A Mutual Fund Returns

To specify the time-series properties of fund returns, we use exactly the same Data Generating Process (DGP) as the one proposed by BSW (p. 3 (appendix)). This DGP is used by both BSW and AP. Therefore, our analysis guarantees a fair comparison between the results documented in (i) BSW, (ii) AP, and (iii) our Reply.

We assume that the excess net return $r_{i,t}$ of each fund i ($i = 1, ..., N$) during each month t ($t = 1, ..., T$) is written as

$$r_{i,t} = \alpha_i + b_i r_{m,t} + s_i r_{smb,t} + h_i r_{hml,t} + m_i r_{mom,t} + \varepsilon_{i,t} = \alpha_i + \beta_i' f_t + e_{i,t}, \tag{10}$$

where α_i is the fund net alpha, $f_t = (r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t})'$ denote the vector of benchmark excess returns for market, size, value, and momentum, $\beta_i = (b_i, s_i, h_i, m_i)'$ is the vector of fund betas, and $e_{i,t}$ is the error term (uncorrelated across funds). We also assume that f_t and $e_{i,t}$ are independent and normally distributed: $f_t \sim N(0, V_f)$, $e_{i,t} \sim N(0, \sigma_e^2)$, where V_f is the factor covariance matrix, and σ_e^2 denotes the residual variance.

With these assumptions, the t-statistic t_i follows a Student t-distribution (t_c) with $T - 5$ degrees of freedom if the fund alpha is null:

$$t_i \sim t_c(T), \quad \text{if } \alpha_i = 0. \tag{11}$$

Otherwise, t_i follows a non-central Student t-distribution (t_{nc}) with $T - 5$ degrees of freedom and a non-centrality parameter $\bar{t}$ equal to $\frac{\alpha}{\sigma_e/\sqrt{T}}$ if the fund alpha is different from zero:

$$t_i \sim t_{nc}(\alpha, \sigma_e, T), \quad \text{if } \alpha_i = \alpha \neq 0. \tag{12}$$

These distributional results are exact (i.e., they hold for fixed T) because the error term is independent, identically normally distributed, and orthogonal to the factors.

B A Simple Analytical Approach

BSW and AP estimate the bias of $\hat{\pi}_0(\lambda)$ and $\hat{\pi}_A(\lambda)$ via a simulation approach for which they need to specify all the parameters in Equation (10). However, this approach is unnecessary given the above assumptions. We can instead use an analytical approach which produces exact values for the bias directly from the Student t-distributions.

Using an analytical approach is simpler and faster. For one, it only requires that we specify the three parameters of the Student t-distributions (i.e., α, σ_e, and T). It is also immune to simulation noise by construction. Finally, it allows for a completely transparent analysis of the bias—conditional on the choice of parameters, there is only one possible output. In our Reply, we therefore use this analytical approach to study the bias of the FDR proportions.

The main computation step is to quantify the misclassification probability in Equation (8). After specifying the values for α, σ_e, and T, we compute $\delta(\lambda)$ as

$$\delta(\lambda) = \frac{F_{nc}(I(\lambda); \alpha, \sigma_e, T)}{1-\lambda}, \tag{13}$$

where $F_{nc}(I(\lambda); \alpha, \sigma_e, T)$ is the probability inferred from the non-central Student distribution t_{nc} over the interval $I(\lambda) = [-a(\lambda), a(\lambda)]$, where the bounds $\pm a(\lambda)$ correspond to the quantiles at $\lambda/2$ and $1-\lambda/2$ of the central Student distribution t_c. We can then substitute $\delta(\lambda)$ into Equations (6)-(7) to obtain the (absolute) bias of $\hat{\pi}_0(\lambda)$ and $\hat{\pi}_A(\lambda)$. As discussed in the appendix, we can also use $\delta(\lambda)$ to compute the (absolute) bias of the proportions of negative/positive alpha funds, $\hat{\pi}_A^-(\lambda)$ and $\hat{\pi}_A^+(\lambda)$.

IV Replication Analysis

We begin our analysis by replicating the results of BSW and AP using the analytical approach presented above. The objectives of this replication analysis are twofold. First, it confirms that the analytical approach yields similar results as those obtained via simulation (as expected from econometric theory). Second, it allows us to discuss the choice of parameter values in BSW and AP, and its impact on the bias of the FDR proportions.

A The Results in BSW

BSW calibrate their simulations using a sample of all active US equity funds (2,076 funds with a minimum number of 60 monthly observations). The sample period begins in January 1975 and ends December 2006 (original sample henceforth). They set σ_e equal to the empirical mean of the residual volatility across funds, and T equal to the total sample size ($\sigma_e = 0.021$, $T = 384$).[10] Using a calibration based on the FDR approach, they further specify that negative alpha funds exhibit an alpha of -3.2% per year, whereas positive alpha funds earn an alpha of 3.8% per year: $\alpha_{ann}^- = -3.2\%$, $\alpha_{ann}^+ = 3.8\%$. Both

[10] To be precise, the empirical mean across all funds is equal to 0.020 (and not 0.021). BSW use a value of 0.021 because they work with a randomly selected subsample of 1,400 funds (out of the 2,076 funds) that allow them to compare the results obtained with and without cross-fund dependence.

α^-_{ann} and α^+_{ann} are divided by 12 to obtain the appropriate monthly values. Finally, they use the estimated FDR proportions as proxies of the true proportions: $\pi_0 = 75\%$, $\pi^-_A = 23\%$, $\pi^+_A = 2\%$. Additional details on the calibration can be found in the appendix of BSW (p. 3-7)

To compute the misclassification probability and the bias terms, we must specify the interval $I(\lambda)$ by choosing the parameter value λ. For simplicity, we initially set λ equal to 0.5—this value is viewed as reasonable by BSW (p. 8 (appendix)), and is used by AP in some of their calculation (see their Figure 1). We discuss alternative values for λ in Section V.A.

We report the results of our replication exercise in Table I. Panel A shows the misclassification probability $\delta(\lambda)$ for the two values of the true alpha, i.e., $|\alpha_{ann}| = \{3.2\%, 3.8\%\}$. In both cases, the probability of misclassifying non-zero alpha funds is close to zero. Consistent with this analysis, Panel B shows that the bias levels for $\hat{\pi}_0(\lambda)$, $\hat{\pi}^-_A(\lambda)$, and $\hat{\pi}^+_A(\lambda)$ are all very small (below 2%). Therefore, our computations closely replicate the results documented by BSW in their Table IA.I (p. 20 (appendix)).

Please insert Table I here

B The Results in AP

We repeat our replication exercise for AP who use a similar sample as BSW (active US equity funds between January 1975 and December 2006). AP improve the initial analysis of BSW along two dimensions. First, they point out that using $T = 384$ overestimates the typical time-series length of mutual fund returns. Using the empirical average of the number of monthly observations across funds, AP set T equal to 150 ($\sigma_e = 0.021$, $T = 150$). Second, they note that the simulation analysis of BSW is incomplete because it only examines two values of alpha. To address this issue, they consider a wider range of values for negative/positive alpha funds, i.e., $\alpha^-_{ann} = -\alpha_{ann}$ and $\alpha^+_{ann} = \alpha_{ann}$, where $\alpha_{ann} = \{1.0\%, 1.5\%, ..., 3.5\%\}$.

AP considers 4 values for the proportion vector $(\pi_0, \pi^-_A, \pi^+_A)'$. Specifically, they choose 4 values for π_0 equal to 94%, 75.0%, 38%, and 6%. For each value of π_0, they determine π^-_A and π^+_A such that the ratio π^-_A/π^+_A remains equal to 11.5 as in BSW (i.e., $23/2 = 11.5$).

Panel A of Table II shows the misclassification probability $\delta(\lambda)$ for each value of the true alpha α_{ann}. Panel B reports the bias of the FDR proportions across the 24 scenarios (4 proportion vectors $\times$ 6 alphas). We find that our analytical approach closely replicates the results obtained by AP. We find that 67% of the funds with an alpha of 2% per year are misclassified as zero alpha funds—a number close to the one reported by AP in their abstract: "65% of funds with economically large alphas of ±2% are misclassified as zero alpha". In addition, the differences in bias for $\hat{\pi}_0(\lambda)$ relative to the original results in AP (their Table V) are, on average, only equal to 0.5% across all scenarios.[11]

Overall, the replication results confirm the analysis of AP. If we change the initial framework of BSW by (i) decreasing the number of observations, and (ii) reducing the alpha that funds possibly exhibit, the FDR proportions become markedly biased.

Please insert Table II here

V Our Analysis

We now turn to the main analysis of our Reply, which contains two main parts. In the first part, we revisit the analysis of BSW and AP by considering alternative values for the parameters and the interval I(λ). In the second part, we update the evidence on the bias by extending the sample period from 2006 to 2018.

A Criticism of BSW and AP

A.1 The Choice of Parameter Values

We agree with AP that the initial analysis of BSW overestimates the number of observations T and focuses on a somewhat limited values of the true alpha of non-zero-alpha funds. In addition to these modifications, we further argue that two changes in parameter values are important to reproduce the salient feature of mutual fund data. These changes, which are not considered by BSW and AP, could materially affect the performance of the FDR estimators.

(i) Residual volatility. Throughout their analysis, BSW and AP set the residual volatility equal to the empirical mean across funds ($\sigma_e = 0.021$). This value is then used as a proxy for the residual volatility of all funds in the population. The main issue with this approach is that the mean is influenced by a few extremely volatile funds. That is, the cross-sectional distribution of residual volatility is heavily skewed (the skewness is equal to 3.9). To obtain a value for σ_e that is more representative of the typical fund, we propose a simple solution: we replace the mean with the median. Applying this change

[11] The tiny observed differences are due to several factors. First, AP only set λ equal to 0.5 to plot their Figure 1. For the simulations, they follow BSW and use the MSE-minimization procedure of Storey (2002) to select λ across values ranging from 0.3 to 0.7. Second, AP do not set $T = 150$ for all funds, but randomly draw values from the empirical distribution in their sample whose median equals 150. Third, AP set the mean of f_t equal to its empirical average (instead of zero). Finally, the simulation analysis does not yield exact values because it is subject to simulation noise.

for both the residual volatility and the number of observations, we obtain $\sigma_e = 0.018$ and $T = 135$.[12]

(ii) Relations between fund parameters. AP extend the analysis of BSW by considering a large set of values for the true alpha α_{ann}. Specifically, they examine different scenarios where α_{ann} varies, but the other parameters σ_e and T stay constant. As theory suggests, these scenarios may not be realistic representations of the real world. If a fund has a high alpha and a low residual volatility, it is highly attractive because of its high information ratio (Treynor and Black (1973)). Stambaugh (2014) shows that as investors allocate more money to this fund, capacity constraints tighten and its alpha moves downward in line with its residual volatility (i.e., lower σ_e means lower α).[13] Furthermore, the model of Berk and Green (2004) implies that funds only deliver positive alphas during the learning phase when they are young (i.e., higher T means lower α). These arguments suggest that an analysis in which α varies should allow for the possibility that σ_e and T vary as well.

We propose a simple calibration approach to capture the relations between parameters. For each value of α_{ann}, we select a total of J_α funds whose estimated alpha $\hat{\alpha}_{ann}$ falls in the bin $\alpha_{ann} \pm 0.5\%$, where the bin width corresponds to the distance between the different values of α_{ann} examined by AP ($\alpha_{ann} = \{1.0\%, 1.5\%, ..., 3.5\%\}$). Extending Equation (13), we then compute the misclassification probability $\delta(\lambda)$ by using the values σ_{e_j} and T_j that are specific to each selected fund j $(j = 1, ..., J_\alpha)$:

$$\delta(\lambda) = \frac{\frac{1}{J_\alpha}\sum_{j=1}^{J_\alpha} F_{nc}(I(\lambda); \alpha, \sigma_{e_j}, T_j)}{1-\lambda}. \quad (14)$$

Table III reports, for each value of alpha α_{ann}, the estimated median residual volatility and median number of return observations among the J_α selected funds. Consistent with the theoretical predictions, we find that funds with lower alphas (i) exhibit lower residual volatility, and (ii) have a higher number of return observations.[14]

Please insert Table III here

In Table IV, we examine the impact of these changes on the misclassification probability.

[12] For consistency, we also use the median for the number of return observations ($T = 135$). Choosing $T = 150$ as in AP would therefore improve the accuracy of the FDR estimators reported in Table IV.

[13] In the model of Stambaugh (2014), there is a one-to-one mapping between α and σ_e because the information ratios of all funds must be equal in equilibrium.

[14] This finding resonates with the empirical evidence in Kosowski et al. (2006; Figure 1 and Table II).

In the row labelled Residual Volatility, we use Equation (13) to compute $\delta(\lambda)$ (with $\sigma_e = 0.018$, $T = 135$, $\lambda = 0.5$). In the row labelled Parameter Relations, we then compute $\delta(\lambda)$ using Equation (14). In the last two rows, we compare our results with those obtained by AP (shown in Table II).

These changes have a favorable impact on the performance of the FDR proportions. For instance, the probability of misclassifying funds with an alpha of 2% per year is equal to 48% with the second specification (Parameter Relations), which represents a 28%-reduction relative to AP ((67-48)/67=0.28). Consistent with intuition, we also find that accounting for the relations between parameters materially improves the detection of funds with low alphas (i.e., those with α_{ann} between 1% and 2% per year).

Please insert Table IV here

A.2 The Choice of the interval I(λ)

We now revisit the choice of the interval I(λ) used by BSW and AP, and examine its impact on the FDR proportions. Choosing $\lambda \in (0,1)$ involves a trade-off between the mean and variance of the FDR proportions. To elaborate, a higher λ lowers the bias of $\hat{\pi}_0(\lambda)$ because $I(\lambda)$ includes the t-statistics of fewer non-zero alpha funds (i.e., $\delta(\lambda)$ grows small). However, it also increases the variance of $\hat{\pi}_0(\lambda)$ because $I(\lambda)$ contains a smaller number of fund t-statistics.

BSW account for this trade-off by using the method of Storey (2002) which chooses λ based on the estimated Mean Squared Error (MSE) of $\hat{\pi}_0(\lambda)$ (p. 189). In addition, BSW impose an upper bound λ at 0.7 to guarantee a conservative (biased) estimator of π_0. AP use exactly the same procedure to maintain consistency with BSW.

Using this particular procedure makes sense in the context of BSW because they use the FDR as a multiple testing approach (as discussed Section II.A). As a result, it is desirable to obtain a conservative estimator $\hat{\pi}_0(\lambda)$ to guarantee a strong control of the Type I error in the selection of funds—a key requirement for any testing procedure (e.g., Storey, Taylor, and Siegmund (2004)).[15] However, if the only interest lies in minimizing the bias of $\hat{\pi}_0(\lambda)$, imposing an upper bound at 0.7 is likely to be suboptimal. Formally, the bias is equal to

$$Bias[\hat{\pi}_0(\lambda)] = \delta(\lambda)\pi_A, \tag{15}$$

[15] Because the FDR literature primarily focuses on minimizing Type I errors in multiple testing, it is not surprising that it only proposes conservative estimators of π_0. In their seminal paper on the FDR, Benjamini and Hochberg (1995) simply set $\hat{\pi}_0$ equal to 1. The same approach is used by Efron (2010) in his book on the FDR.

which decreases if we set λ above the upper bound at 0.7.[16]

Importantly, choosing higher values for λ does not largely increase the variance of $\hat{\pi}_0(\lambda)$. If we assume for simplicity that the t-statistics are independent, we can write

$$Var[\hat{\pi}_0(\lambda)] = \frac{1}{N} \frac{P[I(\lambda)]\,(1 - P[I(\lambda)])}{(1-\lambda)^2}, \tag{16}$$

where $P[I(\lambda)] = \pi_0 P_0[I(\lambda)] + \pi_A P_A[I(\lambda)]$ (see BSW, Genovese and Wasserman (2004)).[17] With several thousand funds ($N = 2,076$ funds in the original sample), the variance term is close to zero and thus weakly impacted by changes in λ. To verify this point, Table V compares the MSE of $\hat{\pi}_0(\lambda)$, defined as

$$MSE(\hat{\pi}_0(\lambda)) = (Bias[\hat{\pi}_0(\lambda)])^2 + Var[\hat{\pi}_0(\lambda)], \tag{17}$$

across the 24 scenarios of AP (4 proportion vectors $\times$ 6 alphas) for λ equal to 0.5 and 0.95 (see the appendix for details). The results show that the MSE obtained with $\lambda = 0.95$ is lower or equal in all but two scenarios where the differences are marginal (3.0 and 2.4 for $\lambda = 0.5$ vs 3.1 and 2.6 for $\lambda = 0.95$). Consistent with intuition, the MSE reduction is strong when the bias in $\hat{\pi}_0(\lambda)$ is large (e.g., when $\pi_0 = 6\%$). In short, the large fund population size calls for a value for λ close to one.[18]

Please insert Table V here

Building on this insight, we recompute the misclassification probability using Equation (14) after increasing λ from 0.5 to 0.95. Table VI shows that a more careful choice of λ further improves the performance of the FDR estimated proportions. For instance, the misclassification probability of a fund with an alpha of 2% per year drops to 44%

[16] Storey (2002) shows that as $\lambda \to 1$ and $N \to \infty$, $\hat{\pi}_0(\lambda)$ converges to $\pi_0 + g'_A(1)\pi_A$, where $g'_A(1)) = \frac{dg_A(\lambda)}{d\lambda}\Big|_{\lambda=1}$, $g_A(\lambda) = 1 - P_A[I(\lambda)]$. This result implies that (i) the bias decreases with λ as long as $g_A(\lambda)$ is concave in λ; (ii) in some cases, we can completely eliminate the bias.

[17] See BSW (p. 5-7 (appendix)) for an extensive analysis of dependencies across t-statistics. In particular, BSW consider a specification in which the entire covariance matrix of the fund returns is directly taken from the data. Using this general specification, BSW find that cross-fund dependence does not largely increase $Var[\hat{\pi}_0(\lambda)]$.

[18] In practice, the value of λ that minimizes the true MSE is unknown because it depends on the parameters of the true DGP. However, we can modify the MSE-minimization method of Storey (2002) to make sure to pick up a large λ. First, we can simply enlarge the set of possible values for λ (e.g., $\lambda \in [0.3, ..., 0.95]$). Second, we can put more weight on the estimated bias versus the estimated variance in the computed approximation of the MSE of $\hat{\pi}_0$. The motivation for adding more weight on the bias arises because the approximation of the MSE underestimates the bias (i.e., it replaces the unknown proportion, π_0, with $\min_\lambda(\hat{\pi}_0)$).

under the second specification (Parameter Relations), which represents a 34%-decrease relative to AP ((67-44)/67=0.34).

Please insert Table VI here

B Updating the Evidence on the Bias

In order to assist current researchers, we next focus on an updated US equity mutual fund sample that begins in January 1975 and ends in December 2018 (2,291 funds with a minimum number of 60 monthly observations). As discussed above, the analysis of BSW and AP only cover the period 1975-2006 and omits the last 12 years of data. Their results, therefore, may have limited value for current researchers in mutual fund performance. To address this issue, we update the evidence to reach a more general conclusion about the usefulness of the FDR approach.

To begin, we compute the misclassification probability using the same approach as in BSW and AP. Specifically, we set the parameter values equal to the empirical mean across funds over 1975-2018 ($\sigma_e = 0.016$ and $T = 216$), and then compute $\delta(\lambda)$ using Equation (13). This analysis allows us to examine how $\delta(\lambda)$ changes when we simply increase the sample size, without making any changes to the framework of BSW and AP. The results in Table VII reveal a large decrease in the misclassification probability. For instance, a fund with an alpha of 2% per year is only misclassified 37% of the time—a number that is almost half of that documented by AP (67%).

Please insert Table VII here

Next, we reexamine the relation between the different parameters over the updated sample. Similar to the period 1975-2006, Table VIII provides supporting evidence that the fund parameters are related—lower values of α come with lower values of σ_e, and higher values of T.

Please insert Table VIII here

In Table IX, we report the updated misclassification probability after accounting for the changes in parameter values. In Panel A, we use the median values instead of the mean for the parameter values ($\sigma_e = 0.015$, $T = 191$, $\lambda = 0.5$), and then incorporate the parameter relations using Equation (14). In Panel B, we perform the same computations after replacing $\lambda = 0.5$ with $\lambda = 0.95$. The results reveal a further improvement in the performance of the FDR proportions. For instance, Panel B shows that the misclassification probabilities under the second specification (Parameter Relations) drops to 43%

and 32% for $\alpha_{ann} = 1.5\%$ and 2% per year (versus 80% and 67% for AP).[19]

Please insert Table IX here

Finally, we quantify the bias of the FDR proportions estimators across a wide range of scenarios. Specifically, we consider 10 values for the proportion vector $(\pi_0, \pi_A^-, \pi_A^+)'$, and 6 values for the true alpha of non-zero alpha funds, which provides a total of 60 scenarios.[20] We initially set π_0 equal to 95%, and then decrease it by increments of 10%, i.e., $\pi_0 = \{95\%, 85\%, ..., 5\%\}$. For each value of π_0, we then follow AP and determine π_A^- and π_A^+ such that the ratio π_A^-/π_A^+ is equal to 11.5.

Table X reports the average values of $\hat{\pi}_0(\lambda)$, $\hat{\pi}_A^-(\lambda)$, and $\hat{\pi}_A^+(\lambda)$ for each of the 60 scenarios using $\lambda = 0.95$. In about half of the scenarios (27 out of 60), the bias of $\hat{\pi}_0(\lambda)$ is lower than 10%, which implies that the FDR approach provides a relevant representation of performance in the mutual fund industry. The only scenarios in which the bias rises above 20% feature a large proportion of non-zero alpha funds with low alphas or, equivalently, funds that become increasingly similar to zero alpha funds (i.e., both π_0 and α are low). In other words, the bias can grow large in specific scenarios. But when it does, it also loses its economic significance—separating zero and non-zero alpha funds becomes less relevant when these funds become increasingly similar.

Please insert Table X here

C Summary of the Results

Our main findings are summarized in Figure 1. We plot the misclassification probability as a function of the alpha for the original sample (Panel A) and the updated sample (Panel B). We introduce each of our changes in a sequential manner to evaluate its marginal impact of $\delta(\lambda)$—that is, we (i) use the median residual volatility, (ii) account for the relations between parameters, and (iii) increase λ from 0.5 to 0.95.

Please insert Figure 1 here

Our analysis of the original sample (1975-2006) incorporates the main points raised by AP, but argues that additional changes are necessary. First, modifying the parame-

[19] In unreported results, we also compute $\delta(\lambda)$ using the five-factor model recently proposed by Fama-French (2015). In this case, the misclassification probability for an alpha of 2% per year is a mere 28% (with Equation (14) and $\lambda = 0.95$).

[20] Our analysis gives equal weight to the different values of π_0, and thus covers a larger set of scenarios than AP (60 versus 24).

ter values are important to capture the salient features of mutual fund data. Second, increasing the value of λ improves the performance of the FDR proportions. With these changes, the overall evidence documented here provides a more nuanced view relative to BSW and AP. Whereas we confirm that the initial analysis of BSW may be too optimistic, we do not find the high levels of bias documented by AP.

Our analysis of the updated sample provides a clearer picture of the performance of the FDR. The misclassification probability decreases significantly, and the bias of the FDR proportions is low across a wide range of scenarios. We therefore conclude that the FDR approach is useful for current researchers in the field of mutual fund performance. These results represent good news for the academic community which typically favors statistical methods that are simple and fast—two major advantages of the FDR approach. They also call for further research on the pros and cons of the different methods for estimating the cross-sectional distributions of fund alphas, which include the FDR as well as the parametric/Bayesian and nonparametric approaches proposed in the literature (e.g., Barras, Gagliardini, and Scaillet (2019), Harvey and Liu (2018), Jones and Shanken (2005)).

References

[1] Andrikogiannopoulou, A., and F. Papakonstantinou, 2019, Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power?, Journal of Finance 75, Replications and Corrigenda, 2667-2688

[2] Augustin, P., M. Brenner, and M. G. Subrahmanyan, 2019, Informed Options Trading prior to Takeover Announcements: Insider Trading?, Management Science, Forthcoming.

[3] Bajgrowicz, P., and O. Scaillet, 2012, Technical Trading Revisited: False Discoveries, Persistence Tests, and Transaction Costs, Journal of Financial Economics, 106, 473-491.

[4] Bajgrowicz, P., O. Scaillet, and A. Treccani, 2016, Jumps in High-Frequency Data: Spurious Detection, Dynamics, and News, Management Science 62, 2198-2217.

[5] Barras, L., 2019, A Large-Scale Approach for Evaluating Asset Pricing Models, Journal of Financial Economics, In Press.

[6] Barras, L., P.Gagliardini, and O. Scaillet, 2019, Skill and Value Creation in the Mutual Fund Industry, Working Paper.

[7] Barras, L., O. Scaillet, and R. Wermers, 2010, False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas, Journal of Finance 65, 179-216.

[8] Benjamini, Y., and Y. Hochberg, 1995, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society 57, 289-300.

[9] Berk, J.B., and R.C. Green, 2004, Mutual Fund Flows and Performance in Rational Markets, Journal of Political Economy 112, 1269-1295.

[10] Chen, Y., M. Cliff, and H. Zhao, 2017, Hedge Funds: The Good, the Bad, and the Lucky, Journal of Financial and Quantitative Analysis 52, 1081–1109.

[11] Cuthbertson, K., D. Nitzsche, and N. O'Sullivan, 2008, UK Mutual Fund Performance: Skill or Luck?, Journal of Empirical Finance 15, 613-634.

[12] Genovese, C., and L. Wasserman, 2004, A Stochastic Process Approach to False Discovery Control, Annals of Statistics 32, 1035-1061.

[13] Giglio, S., Y. Liao, and D. Xiu, 2018, Thousands of Alpha Tests, Working Paper.

[14] Efron, B., 2010, Large-Scale Inference, Cambridge University Press.

[15] Fama, E. F., French, K. R., 2015. A Five-Factor Asset Pricing Factor Model. Journal of Financial Economics 116, 1–22.

[16] Ferson, W., and Y. Chen, How Many Good and Bad Funds are There, Really?, 2019, in C.F. Lee, Editor, Handbook of Financial Economics, Mathematics, Statistics and Technology, Ch.108.

[17] Harvey, C. R., and Y. Liu, 2018, Detecting Repeatable Performance, Review of Financial Studies, Forthcoming.

[18] Harvey, C. R., Liu, Y., Zhu, H., 2016. ...and the cross-section of expected returns. Review of Financial Studies 29, 5–68.

[19] Jones, C.S., and J. Shanken, 2005, Mutual Fund Performance with Learning across Funds, Journal of Financial Economics 78, 507-552.

[20] Kosowski, R., A. Timmermann, H. White, and R. Wermers, 2006, Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis, Journal of Finance 61, 2551-2595.

[21] Scaillet, O., A. Treccani, and C. Trevisan, 2019,.High-Frequency Jump Analysis of the Bitcoin Market" Journal of Financial Econometrics, Forthcoming.

[22] Stambaugh, R.F, 2014, Presidential Address: Investment Noise and Trends, Journal of Finance 69, 1415-1453.

[23] Storey, J.D., 2002, A Direct Approach to False Discovery Rates, Journal of the Royal Statistical Society 64, 479-498.

[24] Storey, J.D., J.E. Taylor, and D. Siegmund, 2004, Strong Control, Conservative Point Estimation and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach, Journal of the Royal Statistical Society 66, 187-205.

[25] Treynor, J.L., and F. Black, 1973, How to Use Security Analysis to Improve Portfolio Selection, Journal of Business 46, 66–86.

VI Appendix

A The Bias of the Proportions on Non-Zero Alpha Funds

As discussed in Section II.B, the estimated proportions of funds with negative/positive alphas are given by

$$\hat{\pi}_A^-(\lambda) = \hat{P}[I_n^-(\lambda)] - \hat{\pi}_0(\lambda)P_0[I_n^-(\lambda)], \tag{A2}$$
$$\hat{\pi}_A^+(\lambda) = \hat{P}[I_n^+(\lambda)] - \hat{\pi}_0(\lambda)P_0[I_n^+(\lambda)]. \tag{A8}$$

Taking the expectations of $\hat{P}[I_n^-(\lambda)]$ and $\hat{P}[I_n^+(\lambda)]$, we obtain

$$\hat{P}[I_n^-(\lambda)] = P_A^-[I_n^-(\lambda)]\pi_A^- + P_0[I_n^-(\lambda)]\pi_0 + P_A^+[I_n^-(\lambda)]\pi_A^+, \tag{A3}$$
$$\hat{P}[I_n^+(\lambda)] = P_A^+[I_n^+(\lambda)]\pi_A^+ + P_0[I_n^+(\lambda)]\pi_0 + P_A^-[I_n^+(\lambda)]\pi_A^-, \tag{A4}$$

where $P_A^-[I_n^-(\lambda)]$ and $P_A^-[I_n^+(\lambda)]$ denote the probabilities that the t-statistic of a negative alpha fund falls in the interval $I_n^-(\lambda)$ and $I_n^+(\lambda)$. Similarly, $P_A^+[I_n^-(\lambda)]$ and $P_A^+[I_n^+(\lambda)]$ denote the probabilities that the t-statistic of a positive alpha fund falls in the interval $I_n^-(\lambda)$ and $I_n^+(\lambda)$. Inserting these expressions in Equations (A1)-(A2) and replacing $\mathrm{E}[\hat{\pi}_0(\lambda)]$ with $\pi_0 + \delta(\lambda)\pi_A$, we obtain

$$E[\hat{\pi}_A^-(\lambda)] = P_A^-[I_n^-(\lambda)]\pi_A^- - \delta(\lambda)P_0[I_n^-(\lambda)]\pi_A + P_A^+[I_n^-(\lambda)]\pi_A^+, \tag{A5}$$
$$E[\hat{\pi}_A^+(\lambda)] = P_A^+[I_n^+(\lambda)]\pi_A^+ - \delta(\lambda)P_0[I_n^+(\lambda)]\pi_A + P_A^-[I_n^+(\lambda)]\pi_A^-. \tag{A6}$$

Equations (A5)-(A6) reveal that both $\hat{\pi}_A^-(\lambda)$ and $\hat{\pi}_A^+(\lambda)$ are biased because their average values depend on all three proportions π_0, π_A^-, and π_A^+.

B Bias Computation for the Proportions on Non-Zero Alpha Funds

We explain how to extend the simple analytical approach to compute the bias for the proportions of funds with negative/positive alphas. We consider a population in which (i) a proportion π_A^- of funds yield a negative alpha $-\alpha$, (ii) a proportion π_A^+ of funds yield a positive alpha α. We consider the more general formulation where these funds can possibly have different values for the parameters σ_{e_j} and T_j ($j = 1, ..., J_\alpha$).

Building on Equation (14), we can write the misclassification probability as

$$\delta(\lambda) = \frac{\frac{1}{J_\alpha}\sum_{j=1}^{J_\alpha} F_{nc}(I(\lambda); \alpha, \sigma_{e_j}, T_j)}{1-\lambda}. \tag{A7}$$

Using the Student t-distributions, we can compute the remaining quantities:

$$P_A^-[I_n^-(\lambda)] = \frac{1}{J_\alpha}\sum_{j=1}^{J_\alpha} F_{nc}(I_n^-(\lambda); -\alpha, \sigma_{e_j}, T_j), \tag{A8}$$

$$P_A^-[I_n^+(\lambda)] = \frac{1}{J_\alpha}\sum_{j=1}^{J_\alpha} F_{nc}(I_n^+(\lambda); -\alpha, \sigma_{e_j}, T_j), \tag{A9}$$

$$P_A^+[I_n^-(\lambda)] = \frac{1}{J_\alpha}\sum_{j=1}^{J_\alpha} F_{nc}(I_n^-(\lambda); \alpha, \sigma_{e_j}, T_j), \tag{A10}$$

$$P_A^+[I_n^+(\lambda)] = \frac{1}{J_\alpha}\sum_{j=1}^{J_\alpha} F_{nc}(I_n^+(\lambda); \alpha, \sigma_{e_j}, T_j), \tag{A11}$$

$$P_0[I_n^-(\lambda)] = P_0(I_n^+(\lambda)) = \frac{1}{2}\lambda. \tag{A12}$$

Given values for the proportions π_A^- and π_A^+, we can then compute the bias of $\hat{\pi}_A^-(\lambda)$ and $\hat{\pi}_A^+(\lambda)$ by inserting the above quantities in Equations (A5)-(A6).

C Computation of the Mean Squared Error of the FDR Proportion

We explain how to compute the Mean Squared Error (MSE) of the proportion of zero alpha funds. We consider a population in which (i) a proportion π_A^- of funds yield a negative alpha $-\alpha$, (ii) a proportion π_A^+ of funds yield a positive alpha α. We consider the more general formulation where these fundscan possibly have different values for the parameters σ_{e_j} and T_j $(j = 1, ..., J_\alpha)$. The MSE of $\hat{\pi}_0(\lambda)$ depends on the bias and the variance terms:

$$Bias[\hat{\pi}_0(\lambda)] = \delta(\lambda)\pi_A, \tag{A13}$$

$$Var[\hat{\pi}_0(\lambda)] = \frac{1}{N}\frac{P[I(\lambda)]\,(1-P[I(\lambda)])}{(1-\lambda)^2}, \tag{A14}$$

where $P[I(\lambda)] = P_A^-[I(\lambda)]\pi_A^- + P_0[I(\lambda)]\pi_0 + P_A^+[I(\lambda)]\pi_A^+$, $P_A^-[I(\lambda)]$, and $P_A^-[I(\lambda)]$. For the bias, we simply need the misclassification probability given in Equation (A7). For the variance, we need to compute the following quantities from the Student t-distributions:

$$P_A^-[I_n(\lambda)] = \frac{1}{J_\alpha}\sum_{j=1}^{J_\alpha} F_{nc}(I_n(\lambda); -\alpha, \sigma_{e_j}, T_j), \tag{A15}$$

$$P_A^+[I_n(\lambda)] = \frac{1}{J_\alpha}\sum_{j=1}^{J_\alpha} F_{nc}(I_n(\lambda); \alpha, \sigma_{e_j}, T_j), \tag{A16}$$

$$P_0[I_n(\lambda)] = 1 - \lambda. \tag{A17}$$

Given values for the proportions π_0, π_A^- and π_A^+, and the number of funds N, we can then compute $Bias[\hat{\pi}_0(\lambda)]$ and $Var[\hat{\pi}_0(\lambda)]$ by inserting the above quantities in Equations (A13)-(A14).

Table I
Replication of BSW

Panel A reports the misclassification probability for the two values of alpha used by BSW (3.2% to 3.8% per year). Panel B shows the average values of the estimated proportions of funds with zero (0), negative (-), and positive (+) alphas for the calibration used by BSW in which 23% of the funds have a negative alpha of -3.2% per year and 2% of the funds have a positive alpha of 3.8% per year (DGP 1). In both Panels, we use the same parameter value as BSW (σ_e=0.021, T=384), and set λ equal to 0.5.

Panel A: Misclassification Probability

	3.2	3.8
BSW	7%	2%

Panel B: Bias in the FDR Proportions

			Negative α=-3.2%; Positive α=3.8%
DGP1	0	75	76%
	-	23	22%
	+	2	2%

Table II
Replication of AP

Panel A reports the misclassification probability for different values used by AP for the true alpha (from 1.0% to 3.5% per year). Panel B shows the average values of the estimated proportions of funds with zero (0), negative (-), and positive (+) alphas in the 24 scenarios considered by AP obtained with 4 values for the vector of proportions (DGP 1 to 4) and 6 different values for the true alpha α ranging from 1.0% to 3.5% per year. In both Panels, we use the same parameter value as AP (σ_e=0.021, T=150), and set λ equal to 0.5.

Panel A: Misclassification Probability

	Alpha (% per year)					
	1	1.5	2	2.5	3	3.5
AP	90%	80%	67%	53%	40%	29%

Panel B: Bias in the FDR Proportions

			Alpha (% per year)					
			1	1.5	2	2.5	3	3.5
DGP1	0	94	99%	99%	98%	97%	96%	96%
	-	6	1%	1%	2%	3%	4%	4%
	+	1	0%	0%	0%	0%	0%	0%
DGP2	0	75	98%	95%	92%	88%	85%	82%
	-	23	2%	5%	8%	12%	15%	18%
	+	2	0%	0%	0%	0%	0%	0%
DGP3	0	38	94%	87%	79%	71%	63%	55%
	-	58	6%	13%	21%	29%	37%	44%
	+	5	0%	0%	0%	0%	0%	0%
DGP4	0	6	91%	81%	69%	56%	44%	33%
	-	86	9%	19%	31%	44%	56%	66%
	+	8	0%	0%	0%	0%	0%	0%

Table III
Relations between Parameters (1975-2006)

This table reports the estimated median residual volatility and number of monthly return observations associated with different values for the true alpha α ranging from 1.0% to 3.5% per year. For each value of alpha, we select a total of J_α funds whose estimated alpha $\hat{\alpha}$ falls in the bin α±0.5%. Then, we report the median residual volatility and median number of observations among the J_α selected funds. The results are based on the original sample examined by BSW and AP which includes all open-end, active US equity funds over the period 1975-2006.

	Alpha (% per year)					
	1	1.5	2	2.5	3	3.5
Residual Volatility	0.016	0.016	0.017	0.017	0.018	0.018
Nb. Return Observations	147	145	143	147	137	121

Table IV
Misclassification Probability (1975-2006)
(with λ=0.5)

This table reports the misclassification probability in the population for different values for the true alpha α ranging from 1.0% to 3.5% per year, using a value for λ equal to 0.5. The second row (Residual Volatility) replaces the mean with the median values for the residual volatility (σ_e=0.018, T=135). The third row (Parameter Relations) uses the same specification as in the second row, but accounts for the existing relations between the fund parameters. The last two rows compare the results for the second specification (Parameter Relations) with those reported by AP (shown in Table II) in absolute and relative terms. The results are based on the original sample examined by BSW and AP which includes all open-end, active US equity funds over the period 1975-2006.

	Alpha (% per year)					
	1	1.5	2	2.5	3	3.5
The Two Changes						
Residual Volatility	88%	75%	60%	45%	32%	21%
Parameter Relations	77%	61%	48%	38%	33%	27%
Difference with AP						
Absolute	13%	19%	19%	15%	7%	2%
Relative	15%	23%	28%	28%	18%	6%

Table V
Mean Squared Error of the FDR Proportion (1975-2006)

This table shows the square root of the Mean Squared Error (MSE) for the estimated proportion of funds with zero alphas in the 24 scenarios considered by AP obtained with 4 values for the vector of proportions (DGP 1 to 4) and 6 different values for the true alpha α ranging from 1.0% to 3.5% per year. Panel A shows the results obtained with λ equal to 0.5. Panel B shows the results obtained with λ equal to 0.95. The results are based on the original sample examined by BSW and AP which includes all open-end, active US equity funds over the period 1975-2006,

Panel A: Lambda equal to 0.5

			Alpha (% per year)					
			1	1.5	2	2.5	3	3.5
DGP1	0	94	6.0	5.3	4.5	3.7	3.0	2.4
DGP2	0	75	22.6	20.0	16.7	13.3	10.1	7.3
DGP3	0	38	56.5	49.8	41.7	33.2	25.1	18.0
DGP4	0	6	84.7	74.7	62.5	49.7	37.6	27.0

Panel B: Lambda equal to 0.95

			Alpha (% per year)					
			1	1.5	2	2.5	3	3.5
DGP1	0	94	6.0	5.3	4.5	3.7	3.1	2.6
DGP2	0	75	22.3	19.3	15.7	12.1	8.9	6.2
DGP3	0	38	55.6	48.0	39.0	30.0	21.7	14.8
DGP4	0	6	83.3	71.9	58.5	44.9	32.5	22.1

Table VI
Misclassification Probability (1975-2006)
(with λ=0.95)

This table reports the misclassification probability in the population for different values for the true alpha α ranging from 1.0% to 3.5% per year, using a value for λ equal to 0.95. The second row (Residual Volatility) replaces the mean with the median values for the residual volatility (σ_e=0.018, T=135). The third row (Parameter Relations) uses the same specification as in the second row, but accounts for the existing relations between the fund parameters. The last two rows compare the results for the second specification (Parameter Relations) with those reported by AP (shown in Table II) in absolute and relative terms. The results are based on the original sample examined by BSW and AP which includes all open-end, active US equity funds over the period 1975-2006.

	Alpha (% per year)					
	1	1.5	2	2.5	3	3.5
The Two Changes						
Residual Volatility	86%	72%	56%	40%	27%	17%
Parameter Relations	75%	57%	44%	34%	29%	23%
Difference with AP						
Absolute	15%	23%	23%	19%	11%	6%
Relative	17%	28%	34%	36%	28%	20%

Table VII
Updated Misclassification Probability (1975-2018)
(with Parameter Values Chosen as in BSW and AP)

This table reports the misclassification probability in the population for different values for the true alpha α ranging from 1.0% to 3.5% per year. Similar to the framework of BSW and AP, we use the mean value across funds to determine the parameter values (σ_e=0.016, T=216), and set λ equal to 0.5. The results are obtained from the updated sample which includes all open-end, active US equity funds over the period 1975-2018.

	Alpha (% per year)					
	1	1.5	2	2.5	3	3.5
Baseline (BSW and AP)	78%	57%	37%	21%	10%	4%
Difference with AP						
Absolute	12%	23%	30%	32%	30%	25%
Relative	14%	28%	44%	60%	75%	85%

Table VIII
Updated Relations between Parameters (1975-2018)

This table reports the estimated median residual volatility and number of monthly return observations associated with different values for the true alpha α ranging from 1.0% to 3.5% per year. For each value of alpha, we select a total of J_α funds whose estimated alpha $\hat{\alpha}$ falls in the bin α±0.5%. Then, we report the median residual volatility and median number of observations among the J_α selected funds. The results are obtained from the updated sample which includes all open-end, active US equity funds over the period 1975-2018.

	Alpha (% per year)					
	1	1.5	2	2.5	3	3.5
Residual volatility	0.014	0.014	0.015	0.015	0.016	0.017
Nb. Return Observations	217	216	195	177	157	139

Table IX
Updated Misclassification Probability (1975-2018)
(with Parameter Changes)

This table reports the misclassification probability in the population for different values for the true alpha α ranging from 1.0% to 3.5% per year. The second row (Residual Volatility) replaces the mean with the median values for the residual volatility (σ_e=0.015, T=191). The third row (Parameter Relations) uses the same specification as in the second row, but accounts for the existing relations between the fund parameters. The last two rows compare the results for the second specification (Parameter Relations) with those reported by AP (shown in Table II) in absolute and relative terms. Panel A shows the results obtained with λ equal to 0.5. Panel B shows the results obtained with λ equal to 0.95. The results are obtained from the updated sample which includes all open-end, active US equity funds over the period 1975-2018.

Panel A: Lambda equal to 0.5

	Alpha (% per year)					
	1	1.5	2	2.5	3	3.5
The Two Changes						
Residual Volatility	77%	56%	36%	20%	10%	4%
Parameter Relations	65%	47%	36%	26%	23%	21%
Difference with AP						
Absolute	25%	33%	31%	27%	17%	8%
Relative	28%	41%	46%	51%	43%	27%

Panel A: Lambda equal to 0.95

	Alpha (% per year)					
	1	1.5	2	2.5	3	3.5
The Two Changes						
Residual Volatility	74%	51%	30%	15%	7%	2%
Parameter Relations	62%	43%	32%	23%	20%	18%
Difference with AP						
Absolute	28%	37%	35%	30%	20%	11%
Relative	31%	46%	52%	57%	50%	37%

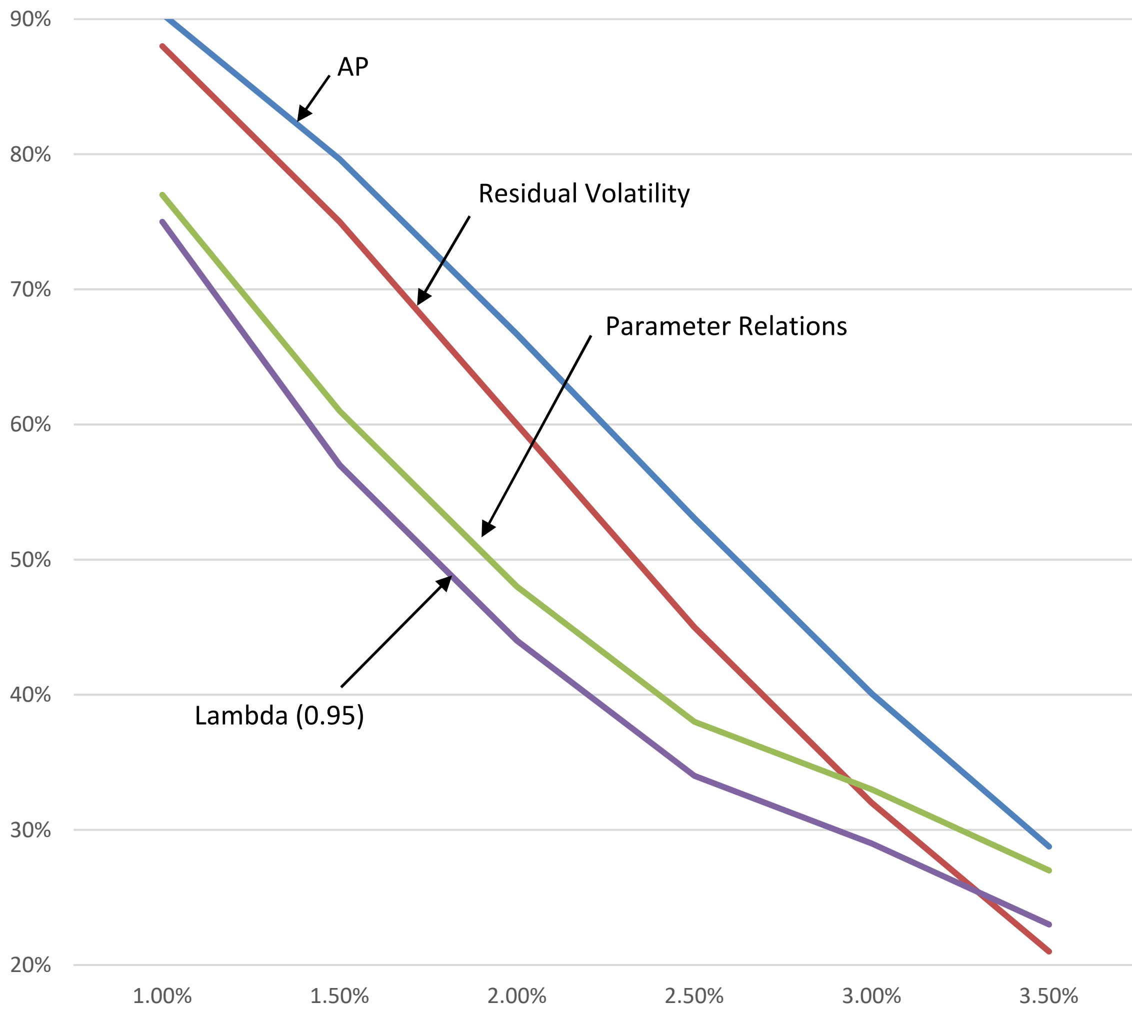
Table X
Average Values of the Estimated Proportions (1975-2018)

This table shows the average values of the estimated proportions of funds with zero (0), negative (-), and positive (+) alphas in 60 scenarios obtained with 10 values for the vector of proportions (DGP 1 to 10) and 6 different values for the true alpha α ranging from 1.0% to 3.5% per year. The computations are based on the second specification (Parameter Relations) and a value for λ equal to 0.95. The results are obtained from the updated sample which includes all open-end, active US equity funds over the period 1975-2018.

			Alpha (% per year)					
			1	1.5	2	2.5	3	3.5
DGP1	0	95	98%	97%	97%	96%	96%	96%
	-	5	2%	3%	3%	4%	4%	4%
	+	0	0%	0%	0%	0%	0%	0%
DGP2	0	85	94%	91%	90%	88%	87%	87%
	-	14	6%	8%	10%	11%	12%	13%
	+	1	0%	0%	0%	0%	1%	1%
DGP3	0	75	91%	86%	83%	80%	79%	78%
	-	23	9%	14%	17%	19%	20%	21%
	+	2	0%	0%	0%	1%	1%	1%
DGP4	0	65	87%	80%	76%	73%	71%	69%
	-	32	13%	20%	24%	26%	28%	30%
	+	3	0%	0%	1%	1%	1%	2%
DGP5	0	55	83%	74%	69%	65%	62%	60%
	-	41	17%	25%	31%	34%	36%	38%
	+	4	0%	0%	1%	1%	2%	2%
DGP6	0	45	79%	69%	62%	57%	54%	51%
	-	51	20%	31%	37%	41%	44%	47%
	+	4	0%	0%	1%	2%	2%	3%
DGP7	0	35	76%	63%	55%	49%	46%	42%
	-	60	24%	36%	44%	49%	52%	55%
	+	5	0%	1%	1%	2%	2%	3%
DGP8	0	25	72%	57%	48%	41%	37%	33%
	-	69	28%	42%	51%	57%	60%	64%
	+	6	0%	1%	1%	2%	3%	4%
DGP9	0	15	68%	52%	41%	33%	29%	24%
	-	78	31%	48%	58%	64%	68%	72%
	+	7	0%	1%	2%	3%	3%	4%
DGP10	0	5	65%	46%	34%	26%	21%	15%
	-	87	35%	53%	65%	72%	76%	80%
	+	8	0%	1%	2%	3%	4%	5%

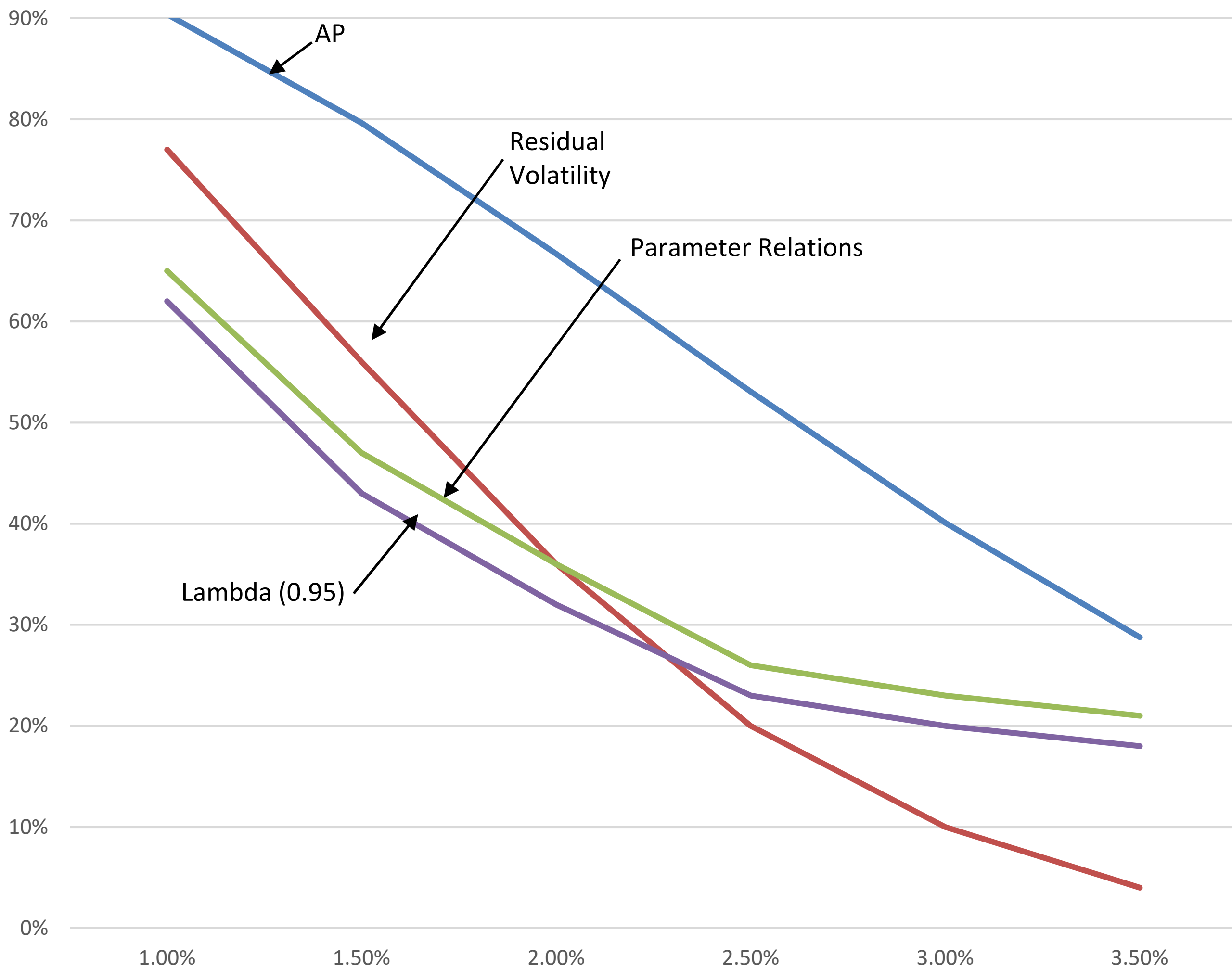
Figure 1
Summary of the Results: Misclassification Probability

This figure plots the misclassification probability for different values for the true alpha α ranging from 1.0% to 3.5% per year. The first line (AP) uses the parameters that replicates the analysis of AP (σ_e=0.021, T=150, λ=0.5). The second line (Residual Volatility) replaces the mean with the median values of the fund residual volatility. The third line (Parameter Relations) accounts for the existing relations between the fund parameters. The fourth line (Lambda (0.95)) uses the same specification as in the third line, but reduces the width of the interval I(λ) by increasing λ from 0.5 to 0.95. Panel A shows the results for the original sample examined by BSW and AP (1975-2006). Panel B shows the results for the updated sample (1975-2018).



Panel A: Original Sample (1975-2006)

Figure 1
Summary of the Results: Misclassification Probability (Continued)



Panel B: Updated Sample (1975-2018)