Accepted Manuscript

In search of the optimal number of fund subgroups

Cheng Yan, Tingting Cheng

PII: S0927-5398(18)30086-0
DOI: https://doi.org/10.1016/j.jempfin.2018.12.002
Reference: EMPFIN 1084

To appear in: *Journal of Empirical Finance*

Received date : 11 September 2018
Revised date : 5 December 2018
Accepted date : 24 December 2018

Please cite this article as: C. Yan and T. Cheng, In search of the optimal number of fund subgroups. *Journal of Empirical Finance* (2019), https://doi.org/10.1016/j.jempfin.2018.12.002

In Search of the Optimal Number of Fund Subgroups☆

Cheng Yan[a], Tingting Cheng[b]

[a]*Essex Business School, Colchester, CO4 3SQ, UK.*

[b]*School of Finance, Nankai University, Tianjin, 300350, PRC.*

Abstract

The idea of determining the number of fund subgroups is of central importance in the currently popular academic field of Risk Parity Portfolio Theory, and especially for practitioners' direct use of Funds-of-Funds (FoF) managers. Can the Gaussian Mixture Distributions plug-in approach via traditional procedures select the right number of fund subgroups? Probably not. According to our in-sample/out-of-sample likelihood score analysis, the actual locations of subgroups in real data (of both U.S. mutual funds and hedge funds) are too close to each other. The information loss incurred by parameter uncertainty outweigh those incurred by mis-specification, and can only be slightly alleviated using the nonparametric density estimators. An arbitrary choice of two subgroups only causes affordable information loss relative to more fund subgroups. These findings challenge the reliability of the Gaussian Mixture Distributions plug-in approach via traditional procedures (e.g., BIC, Likelihood Ratio and Chi-square

☆Corresponding author: Tingting Cheng. Email: tingting.cheng@nankai.edu.cn. We are grateful to David Blake, Yong Chen, Wayne Ferson, Ken French, Jiti Gao, Robert Kosowski, Yan Liu, Lubos Pastor and participants in the workshops in University of Durham, University of Glasgow, University of Leicester, University of Kent, University of Sussex, and University of Sydney for helpful comments. We are grateful to Xiangdong Wang for his excellent research assistance, and a grant from National Natural Science Foundation of China (NNSFC Grant No.71703043, 71803091, 71801117, 71803196). Tingting Cheng is also grateful to the Ministry of Education in China Project of Humanities and Social Sciences (Project No. 18YJC790015) and the National Social Science Foundation of China (Project No. 18CJY061). We alone are responsible for all errors.

statistics) in selecting the correct number of subgroups.

Keywords: Performance evaluation, Fund subgroups, Gaussian mixture distribution, Parameter uncertainty, Mis-specification

JEL Classification: C15, G11, G12, G23

HIGHLIGHTS

- Investigate the two potential concerns on the validity and reliability of GMD.
- Parameter uncertainty has a larger effect than mis-specification.
- Nonparametric kernel approaches slightly alleviate information loss.
- An arbitrary choice of two subgroups causes affordable information loss.
- Challenge the GMD procedures in selecting the correct number of subgroups.

In Search of the Optimal Number of Fund Subgroups☆

Abstract

The idea of determining the number of fund subgroups is of central importance in the popular academic field of risk parity portfolio theory, and especially for practitioners' direct use of fund-of-funds managers. Can the Gaussian Mixture Distributions plug-in approach via traditional procedures select the correct number of fund subgroups? Probably not. According to our in-sample/out-of-sample likelihood score analysis, the actual locations of subgroups in real data (of both U.S. mutual funds and hedge funds) are too close to each other. The information loss incurred by parameter uncertainty outweighs that incurred by misspecification, and can only be slightly alleviated using the nonparametric density estimators. An arbitrary choice of two subgroups only causes affordable information loss relative to more fund subgroups. These findings challenge the reliability of the Gaussian Mixture Distributions plug-in approach via traditional procedures (e.g., Bayesian Information Criterion, Likelihood Ratio and Chi-squared statistics) in selecting the correct number of subgroups.

Keywords: Performance evaluation; Fund subgroups; Gaussian Mixture Distribution; Parameter uncertainty; Misspecification

JEL Classification: C15; G11; G12; G23

1. Introduction

It is a common practice in fund performance evaluation literature, to categorize funds into subgroups based on their performance indicators (e.g., fund alphas). So, how many subgroups of funds are there? Theoretically, two subgroups (skilled and unskilled) could coexist but it is also reasonable to conceptualize three subgroups, with zero, negative, and positive alphas, respectively. To determine the number of fund subgroups prior to performance evaluation, most studies use the Gaussian Mixture Distributions (hereafter GMD) plug-in approach (i.e., estimating alphas first through factor model regressions and then fitting a GMD model on the estimated alphas) via traditional procedures, such as the Bayesian Information Criterion (BIC), Likelihood Ratio (LR) test, and the Chi-squared test. Employing the same sample of mutual fund monthly returns, Harvey and Liu (2018) find that two subgroups are optimal while Ferson and Chen (2017) obtain three. Using hedge fund monthly returns over the period from 1994 to 2011, Ferson and Chen (2017) find two subgroups, while Chen et al. (2017) identify four. Since these authors rely on slightly different variants of the same GMD model, it seems difficult to reconcile their mixed results, and it is surprising that few rigorous statistical analyses have been conducted to examine the associated information loss. Given the popularity of the GMD approach, such analyses are essential to the understanding of the literature and appropriate applications in future research. Our paper fills this gap using Monte Carlo simulations as well as two real data samples identical to these of Cai et al. (2018) and similar to these of Ferson and Chen (2017) and Harvey and Liu (2018) covering both U.S. mutual funds and hedge funds.

The idea of determining the number of fund subgroups is of central importance in the popular academic field of risk parity portfolio theory, and especially for practitioners' direct use of fund-of-funds managers. For instance, in fund performance evaluation studies, our question is of essential importance for at least two recent strands of literature, including the studies related to the False Discovery Rate (e.g., Barras et al. (2010); Bajgrowicz and Scaillet (2012);

Criton and Scaillet (2014); Bajgrowicz et al. (2015); Ferson and Chen (2017); Scaillet et al. (2018)) and Expectation-Maximization algorithm (e.g., Chen et al. (2017); Harvey and Liu (2018)). The conclusions of these two strands of literature hinge on the correct presumption of the number of subgroups, according to traditional wisdom. For instance, if, in reality, there are only two subgroups of mutual fund managers, a three-subgroup approach will inevitably cause one entire group to be a false discovery. Hereafter, we refer to this misconception as model misspecification.

Unfortunately, in reality, the true model specification is not observable, making the determination of the true number of fund subgroups especially difficult. In particular, a model with more fund subgroups could fit the empirical data better than a counterpart model with fewer fund subgroups, given the well-known overfitting problem in econometrics.[1] Even if we ignore the overfitting problem, a true fund population with more subgroups, if approximated parametrically, requires more parameters to be estimated, as in most of the finance literature, which will inevitably exacerbate the estimation errors (i.e., estimation risk issues). In other words, since the fund alphas are estimates, the estimation error creates an errors-in-variables problem, which, following the portfolio selection literature we refer to as parameter uncertainty hereafter in this paper (e.g., Brown (1979); Jobson and Korkie (1980); Jorion (1986); Garlappi et al. (2006); DeMiguel et al. (2007); Yan and Zhang (2017)).[2] Hence, the benefits of a better-specified model with more fund subgroups could be offset by the disadvantages incurred by the increased estimation errors brought about by the greater number of parameters

[1]For instance, Harvey and Liu (2018, page 29) explicitly state that "A central issue is how we choose the number of components for the GMD that models the alpha distribution in the cross-section. A more complex model (i.e., a model with more component distributions) can potentially provide a better approximation to the underlying alpha distribution, but may overfit, leading to a model that has inferior forecasts out of sample."

[2]For brevity, we do not fully review the portfolio selection papers regarding parameter uncertainty here.

to be estimated.

We explicitly investigate the effects of model misspecification and/or parameter uncertainty on three aspects of the GMD via Monte Carlo simulations: mean validity (i.e., bias of the estimated mean of alphas), density validity (i.e., Mean Integrated Squared Errors) and reliability (i.e., Standard Deviation of Integrated Squared Errors). To do so, we first generate fund returns series from a CAPM-type of single-factor market model with the individual fund alphas randomly drawn from one of three scenarios: a mixture of two Gaussian distributions (GMD(2) as in Harvey and Liu (2018), Ferson and Chen (2017)), a mixture of three Gaussian distributions (GMD(3) as in Ferson and Chen (2017)) and a mixture of four Gaussian distributions (GMD(4) as in Chen et al. (2017)). In every scenario, we estimate the market model via fund-by-fund Ordinary Least Squares (OLS) and then employ a GMD model to match the distribution of the estimated individual fund alphas. If model specification/misspecification is more important, we expect to observe that both the validity loss and reliability loss are lowest when the number of Gaussian components in the GMD model we use to capture the distribution of fund alphas equals the number of Gaussian components in the data generating processes. Alternatively, if parameter uncertainty is more important, we conjecture that there is a monotonic increase in both the validity loss and reliability loss when the number of mixed Gaussian distributions grows. We find that, of all the GMD estimators, GMD (2) has the lowest mean validity loss, density validity loss and reliability loss in every scenario while the GMD (4) has the largest. In other words, parameter uncertainty is more important than model specification/misspecification, and the traditional procedures do not always select the correct number of fund subgroups.[3] This challenge the use of GMD plug-in approach in selecting the

[3]Admittedly, if the distance between different Gaussian mixture components is large enough, the traditional procedures could still work. However, the fund manager performances from different subgroups are so close to each other that it is technically challenging to make the traditional methods to work properly. This could change

optimal number of fund subgroups which subsequently motivates us to evaluate the potential information loss incurred by using a wrong number of fund subgroups.

To what extent can we improve validity and reliability using alternative estimators to alleviate the parameter uncertainty problem? To answer this question, we minimize the parameter uncertainty problem by employing two popular nonparametric (i.e., global and adaptive) kernel density estimators that require least parameter estimation (bandwidth being the only one parameter) and hence are affected the least by the parameter uncertainty problem (see, e.g., Breiman et al. (1977), Abramson (1982), Silverman (1986)). To our surprise, these two nonparametric kernel estimators can add only limited value to the GMD approach.[4]

In terms of real data, we apply the traditional procedures (i.e., BIC, LR, and Chi-squared test) to two real data samples, similar to Ferson and Chen (2017) and Harvey and Liu (2018), covering both U.S. mutual funds and hedge funds, and find empirical evidence that supports our prior results. In general, we find that the traditional procedures do not provide consistent estimation of the number of fund subgroups, no matter whether we focus on the mutual fund or hedge fund sample. Since the true density of fund returns is unknown in reality, we employ the scoring rule proposed by Amisano and Giacomini (2007) to examine both the in-sample and out-of-sample performance of GMD (1) (a single normal density), GMD (2), GMD (3) and GMD (4). We find that both the in-sample and out-of-sample likelihood scores for GMD (2), GMD (3) and GMD (4) do not differ much from each other, whereas all of them differ substantially from those of GMD (1). This result reaffirms our previous simulation results that

in the future when the performance of fund managers becomes clearly categorized. For instance, if there are four modes (say, highly skilled, moderately skilled, just skilled and poorly skilled) in the density function, then GMD (4) could be preferred in this case.

[4]However, it is possible to enhance the performance of the kernel estimator by correcting or smoothing the errors-in-variables biases. We thank an anonymous referee for pointing this out.

the information loss is trivial, whether we use GMD (2), GMD (3) or GMD (4). We apply two nonparametric (i.e., global and adaptive) kernel density estimators and the scoring rule of Amisano and Giacomini (2007) to determining the optimal number of fund subgroups, which constitutes a methodological novelty.

Overall, our findings challenge the reliability of traditional procedures in selecting the correct number of subgroups, and the traditional wisdom that the results of fund performance evaluation hinge on the presumption of the correct number of subgroups. As the number of fund subgroups increases, there is a tradeoff between precisely estimating the model parameters and a better empirical fit to the data. Therefore, an arbitrary choice of two fund subgroups might only cause affordable information loss relative to an alternative assumption of more fund subgroups, even when the latter fits the empirical data better than the former. Although the importance of parameter uncertainty has been underscored in the topic of portfolio selection, we extend this idea to an entirely new area that has been largely neglected by previous research. We find that parameter uncertainty plays a dominating role in fund subgroup selection, suggesting that the effects of parameter uncertainty in Finance could be greater than people currently think and may not be limited to one or two particular topics.

Our results provide both academic researchers and practitioners with guidance in selecting the optimal number of fund subgroups in their simulations and real data analysis, intuition in regards to the potential deficiencies in their specifications, and new insights for generating alternative fund evaluation approaches. For this, methods that take into account the estimation uncertainty of alphas and complement the traditional procedures by feeding both the estimated alphas and their standard errors into the EM algorithm (i.e., Chen et al. (2017)) can represent a promising direction to pursue, as does specifying the likelihood function for the factor model regression for each fund (i.e., Harvey and Liu (2018)).

The remainder of the paper proceeds as follows. Sections 2 and 3 present the results from Monte Carlo simulations and real data sets (of both mutual funds and hedge funds), respectively. Section 4 concludes. For brevity, we delegate all technical contents and additional results to the Appendices.

2. Simulation

In this section we simulate the effects of misspecification and parameter uncertainty on the performance of a GMD estimator, we examine the reliability of traditional procedures to select an optimal number of fund subgroups, and we compare the performance of the aforementioned density estimators for the cross-sectional fund alpha density from three aspects: mean validity, density validity and reliability. According to the Occam's Razor, we focus on the Euclidian distance although we are aware of many other more complex density comparing measures (e.g., Kullback-Leibler divergence (Kullback and Leibler (1951)), Bhattacharyya distance (Bhattacharyya (1943)), etc.) in the literature. In untabulated results which are available from the authors upon request, we find supportive evidence using the Kullback-Leibler divergence as well as the Bhattacharyya distance, and we choose to omit the specific results here for brevity. We conservatively implement our GMD estimators via the standard expectation-maximization (EM) algorithm rather than more advanced EM algorithms such as the ones in Harvey and Liu (2018) and Chen et al. (2017) to stay on the safe side.

2.1. Data Generating Processes

Following Cheng and Yan (2017), Zhang and Yan (2018) and other extant studies, for simplicity we generate fund returns (r_{it}) from the CAPM-type of single-factor market model using the following Data Generating Processes (DGP):

$$r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it}, \text{ for } i = 1, \cdots, N, t = 1, \cdots, T, \tag{1}$$

where β_i is generated from a uniform distribution over the support $[0.5, 1.5]$. Market return, r_{mt}, is generated from a Gaussian distribution with a mean of $0.08/12$ and a standard deviation of $\sqrt{0.15^2/12}$. The disturbance term, e_{it} is generated from a Gaussian distribution with mean 0 and standard deviation 0.02 denoted as $N(0, 0.02^2)$. We consider the following representative scenarios for the generation of α_i.

Scenario 1: Following Table 1 of Harvey and Liu (2018), we generate α_i from a mixture of two Gaussian distributions: $0.283N(-0.02277/12, 1.513^2/12) + 0.717N(-0.00685/12, 0.586^2/12)$.

Scenario 2: Following Panel C of Table IV of Ferson and Chen (2017), we generate α_i from a mixture of three Gaussian distributions[5]: $0.507N(0, 0.2^2) + 0.069N(-0.0003, 0.2^2) + 0.424N(-0.002, 0.2^2)$.

Scenario 3: Following Table A1 of Chen et al. (2017), we generate α_i from a mixture of four Gaussian distributions: $0.1N(0.01, 0.7^2) + 0.4N(0.003, 0.7^2) + 0.4N(0, 0.7^2) + 0.1N(-0.01, 0.7^2)$.

We randomly generate $R = 500$ replications in each scenario with $(N, T) = (400, 200)$, after consulting with the size of actual fund samples in the literature (Harvey and Liu (2018), Ferson and Chen (2017), Chen et al. (2017)). Although we have obtained qualitatively similar results using a variety of alternative parameter values suggested by earlier literature such as Barras et al. (2010), we omit them for brevity.

The estimation procedure is as follows.

- For the rth replication ($r = 1, 2, \cdots, 500$), we estimate model (1) via fund-by-fund Ordinary Least Squares (OLS) and denote estimated fund alpha as $\widehat{\alpha}_{i,r}$, for $i = 1, 2, \cdots, N$.
- We then estimate the cross-sectional distribution of alpha based on $\left\{\widehat{\alpha}_{i,r}\right\}_{i=1}^{N}$ using GMD

[5] Strictly speaking, Ferson and Chen (2017) did not rely on normality in their main text but rather extensively in their (especially internet) appendices. For instance, Ferson and Chen (2017, page 35) end their manuscript with "We find that the use of an asymptotic normal approximation in these calculations provides improved finite sample performance for the standard errors. The Internet Appendix provides the details."

(2), GMD (3), GMD (4), global and adaptive kernel estimator, and denote the density estimates as $\widehat{f}_{GMD(2),r}(\alpha)$, $\widehat{f}_{GMD(3),r}(\alpha)$, $\widehat{f}_{GMD(4),r}(\alpha)$, $\widehat{f}_{global,r}(\alpha)$ and $\widehat{f}_{adaptive,r}(\alpha)$, respectively. The details of global and adaptive kernel estimators are presented in Appendix A.

2.2. Do traditional procedures such as BIC select the right number of fund subgroups?

Another vital question pertaining to our analysis is whether the prevailing traditional procedures select the right number of fund subgroups. Given the popularity of using the traditional procedures to select the number of fund subgroups (for recent examples, see, e.g., Chen et al. (2017); Ferson and Chen (2017); Harvey and Liu (2018)), no rigorous statistical analysis has been conducted to examine this question, to the best of our knowledge. Although the traditional procedures feasible and arguably grounded in the literature, they are mostly designed for the choice of best empirical fit, and do not necessarily select the right number of fund subgroups, given the concerns (e.g., parameter uncertainty) we raised in our previous analysis. Admittedly, it is difficult, if not impossible, to use the real data to examine this question, as the true fund population in reality is unobservable and hence the true number of fund subgroups, in reality, is unknown. Fortunately, our Monte Carlo simulation design offers us an opportunity to answer this question, as we know the true number of fund subgroups is 2, 3, and 4 for Scenarios 1, 2 and 3, respectively.

In this subsection, we follow Chen et al. (2017) and focus on the popular Bayesian Information Criterion (BIC), for brevity. The results from other traditional procedures such as Likelihood ratio statistics (see, e.g., Harvey and Liu (2018)) and Chi-squared statistics (Ferson and Chen (2017)) are qualitatively similar and hence we delegate them to Section 4.

We present our results of BIC values over 500 replications for Scenarios 1, 2 and 3 in Table 3. To be specific, we report the mean of BIC values over 500 replications when we specify two subgroups, three subgroups and four subgroups using GMD (2), GMD (3) and GMD (4) estimators with the corresponding frequency of being selected in parentheses. It is clear that, the mean of BIC values over 500 replications for GMD (2), GMD (3) and GMD (4) do not differ much from each other, no matter which scenario we look into. Interestingly and surprisingly, in all three scenarios, the BIC suggests us to select GMD (2) as it generates the smallest BIC value with a probability of at least 96.2% (=481/500 in Scenario 2). In other words, the BIC does not always select the right number of fund subgroups, which may upset many financial economists who rely on traditional procedures to select the optimal number of fund subgroups and motivates us to evaluate the potential information loss incurred by using a wrong number of fund subgroups in the next subsection.

2.3. Information loss evaluation

We first investigate the mean validity by comparing the bias in mean alpha μ_α in each scenario, which is defined as the average distance between the true value of mean alpha and the estimated value of mean alpha obtained from different density estimators.

$$Bias = \frac{1}{R}\sum_{r=1}^{R}\left|\widehat{\mu}_{\alpha,r} - \mu_\alpha\right|, \tag{2}$$

where μ_α is the true value and $\widehat{\mu}_{\alpha,r}$ is the estimated value of mean based on the rth replication.

The results are presented in Panel A of Table 4, from which we can see that i) the mean validity loss is negligible in magnitude (the maximum is 2.6434982/1000 $<$ 0.3% for GMD (4) in scenario 3) for all estimators; ii) nonparametric kernel estimators only alleviate mean validity loss by $0.755050/0.751167 - 1 = 0.5\%$; iii) parameter uncertainty is more important than model misspecification in using GMD. As in all three scenarios, the mean validity loss monotonically increases at the number of unknown parameters and is least when there are the fewest parameters to be estimated, not when the GMD is correctly specified.

We then look at the density validity and reliability in general, using the Mean Integrated

Squared Errors (MISE) as the criteria for density validity loss, and the Standard Deviation of Integrated Squared Errors (SDISE) as the criteria for reliability loss, respectively. Before calculating MISE and SDISE, we choose a sequence of grid points $\{\widehat{\alpha}_{min,r}, \cdots, \widehat{\alpha}_{max,r}\}$ for $m = 200$ and approximate Integrated Squared Errors (ISE) based on the r^{th} replication as follows.

$$ISE(\widehat{f}_{GMD(2),r}) \approx \frac{1}{m}\sum_{i=1}^{m}\left(\widehat{f}_{GMD(2),r}(\alpha_i) - f(\alpha_i)\right)^2, \quad ISE(\widehat{f}_{GMD(3),r}) \approx \frac{1}{m}\sum_{i=1}^{m}\left(\widehat{f}_{GMD(3),r}(\alpha_i) - f(\alpha_i)\right)^2$$

$$ISE(\widehat{f}_{GMD(4),r}) \approx \frac{1}{m}\sum_{i=1}^{m}\left(\widehat{f}_{GMD(4),r}(\alpha_i) - f(\alpha_i)\right)^2, \quad ISE(\widehat{f}_{global,r}) \approx \frac{1}{m}\sum_{i=1}^{m}\left(\widehat{f}_{global,r}(\alpha_i) - f(\alpha_i)\right)^2$$

$$ISE(\widehat{f}_{adaptive,r}) \approx \frac{1}{m}\sum_{i=1}^{m}\left(\widehat{f}_{adaptive,r}(\alpha_i) - f(\alpha_i)\right)^2 \quad ,$$

where $f(\alpha_i)$ denote the true density of alpha evaluated at point α_i.

Upon the obtained ISE values, it is straightforward to compute MISE and SDISE. We use the GMD (2) estimator as an example and the steps for other estimators are likewise.

$$MISE(\widehat{f}_{GMD(2)}) = \frac{1}{R}\sum_{r=1}^{R} ISE(\widehat{f}_{GMD(2),r}), \quad R = 500 \tag{3}$$

$$SDISE(\widehat{f}_{GMD(2)}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(ISE(\widehat{f}_{GMD(2),r}) - MISE(\widehat{f}_{GMD(2)})\right)^2}, \quad R = 500. \tag{4}$$

The MISE values are presented in Panel B of Table 4, from which we find a monotonic pattern similar to mean validity and the magnitude of density validity loss is tiny (the maximum is 2.6933% for GMD (4) in scenario 2 for all estimators. In all the scenarios, GMD (2) underperforms the adaptive kernel estimator but outperforms the global kernel estimator, while the global kernel estimator performs better than GMD (3) and GMD (4), reaffirming the previous finding that the validity loss of using GMD estimators (at least GMD (2)) is small.

With regards to reliability and SDISE values in Panel C of Table 4, we also find a monotonic pattern similar to mean and density validity, and the magnitude of reliability loss is subtle (the maximum is 2.7509% for GMD (4) in scenario 1. The GMD estimators no longer outperform nonparametric estimators, and the SDISE values of nonparametric estimators can be one order of magnitude smaller than the ones of the GMD (4).

To visualize our results, we plot the corresponding density estimates in Figure 1 and find that the density estimates obtained from various parametric and nonparametric estimators are very close to each other (The only exception is GMD (4), which performs slightly less reliably than the others, probably due to the classical parameter uncertainty problem, which collaborates with our previous analysis), and reasonably close to the true density of alpha. Overall, although the traditional procedures may not be a reliable indicator to select the optimal number of fund subgroups, it will not incur too severe potential information loss using an incorrect (but not too far from the correct one, of course) number of fund subgroups, which to some extent comforts the financial economists who rely on traditional procedures to select the optimal number of fund subgroups.

3. Real data analysis: Mutual funds and hedge funds

In this section, we evaluate the performance of GMD (2), GMD (3) and GMD (4) in real data of the fund returns net of all management expenses and 12b-fees in two scenarios: for all U.S. mutual funds and hedge funds, respectively. We first introduce our fund data sets, employ the traditional procedures to select fund subgroups, and use both in-sample and out-of-sample likelihood scores to forecast the density of real data of fund returns afterward. As the true density of fund returns, in reality, is unknown, we employ the scoring rule proposed by Amisano and Giacomini (2007) to examine both the in-sample and out-of-sample performance of GMD (1), GMD (2), GMD (3) and GMD (4). Using this scoring rule, we should be able to decide the best performer among GMD (1), GMD (2), GMD (3) and GMD (4).

3.1. Data and descriptive statistics

Our data set has been used in a companion paper (i.e., Cai et al. (2018)), and hence the descriptive statistics necessarily follow Cai et al. (2018). Our cross-section sample of mutual funds is similar to Ferson and Chen (2017) and Harvey and Liu (2018). To be specific, we obtain active U.S. equity mutual funds data from the Center for Research in Security Prices (CRSP) Survivor-Bias-Free Mutual Fund database for the 1984-2011 period. Our sample period is the same as that of Harvey and Liu (2018) and Ferson and Chen (2017) for comparison reasons. We exclude the index funds. To mitigate omission bias (Elton et al. (2001)) and incubation and back-fill bias (Evans (2010)), we exclude observations prior to the reported year when the mutual funds were first entered into the database, and the funds which do not report a year of organization. We only include the funds which have initial total net assets (TNA) above $10 million and more than 80% of their holdings in equity markets. To avoid the look-ahead bias, we do not exclude funds whose TNA subsequently fall below $10 million. These screens leave us with a sample of 3026 (2557) mutual funds with at least 8 (30) months of returns data for the 1984-2011 period.[6]

Our cross-section sample of hedge funds is similar to Ferson and Chen (2017). To be specific, we obtain U.S. equity-oriented hedge fund data from Lipper TASS for the 1994-2011 period. Our sample period is the same as that of Ferson and Chen (2017) for comparison reasons. To mitigate back-fill bias, we remove the first 24 months of returns and returns dated before funds were first entered into the database, and funds with missing values in the field for the add date (Ferson and Chen (2017)). We only include those categorized for a given month as either Dedicated short bias, Event-driven, Equity market neutral, Fund-of-Funds or Long/short

[6]Similarly, Harvey and Liu (2018) and Ferson and Chen (2017) obtained a sample of 3619 and 3716 mutual funds with at least 8 months of returns over the same period, respectively. We follow Hunter et al. (2014) by using 30 months as our threshold as it adds robustness to our results.

equity hedge. Similar to the mutual fund sample, we require that a fund has initial total net assets (TNA) above \$10 million as of the first date. These screens leave us with a sample of 3533 (2072) hedge funds with at least 8 (30) months of returns data for the 1994-2011 period.

Table 5 presents summary statistics of the mutual fund and hedge fund data in our study. We find that they share similar characteristics with the data sample used in Ferson and Chen (2017). The main characteristics are listed as follows.

- The range of average returns across funds is much greater in the hedge fund sample ($-0.114 \sim 0.173$) than that in the mutual fund sample ($-0.09 \sim 0.06$).
- The median of estimated alpha from our linear model for the hedge funds is positive, while for the mutual funds it is slightly negative. The tails of the cross-sectional alpha distributions extend to larger values for the hedge funds. For example, the upper 5% tail value for the alphas in the hedge fund sample is 1.2% per month, while for the mutual funds it is only 0.4%. In the left tails the two types of funds also present different alpha distributions, with a thicker lower tail for the alphas in the hedge fund sample.
- The sample volatility of the median hedge fund return (2.8% per month) is smaller than for the median mutual fund (5.3%). The range of volatilities across the hedge funds is greater, with more mass in the lower tail. For example, between the 10% and 90% quantiles of hedge funds the volatility range is 1.2% - 7.5% (1.2% - 6.7% in Ferson and Chen (2017)), while for the mutual funds it is 3.6% - 7.8% (4.2% - 7.0%) in Ferson and Chen (2017).
- The autocorrelations of the returns are slightly higher for the hedge funds. The median autocorrelation for the hedge funds is 0.127, compared with 0.121 for the mutual funds, and some of the hedge funds have substantially higher autocorrelations. The 5% left tail for the autocorrelations is -0.304 for the hedge funds, versus only -0.121 for the mutual funds.

3.2. Estimating fund alphas via a linear factor model

Before using the likelihood score methodology in Amisano and Giacomini (2007) to examine the in-sample performance of GMD (2), GMD (3) and GMD (4), we have to estimate the unobservable fund skills. For brevity, we use the one-factor market model, i.e., model (1) in the previous simulation section to illustrate our idea, but one factor may not be enough for empirically analyzing the real data sets, according to the vast strand of asset pricing and fund performance evaluation literature. As a result, for real data sets, we estimate fund alpha as the measure of fund skills via various linear factor models. To be specific, for mutual funds, we consider the Fama-French-Carhart four-factor model as well as its two most famous special cases: the one-factor market model (i.e., model (1)) and the Fama-French three-factor model. The Fama-French-Carhart four-factor model can be written as below:

$$r_{it} = \alpha_i + \beta_{i1}MKT_t + \beta_{i2}SMB_t + \beta_{i3}HML_t + \beta_{i4}MOM_t + \varepsilon_{it},\ t = 1, \cdots, T \tag{5}$$

where r_{it} denotes the excess return of fund i at time t. MKT_t, SMB_t, HML_t and MOM_t are the Fama-French-Carhart four factors, which to be specific denote the Market excess return (MKT) factor, the Small-Minus-Big (SMB) size factor, the High-Minus-Low (HML) value factor and the Momentum (MOM) factor at time t, respectively. Different from the Fama-French-Carhart four-factor model, the Fama-French three-factor model does not include the Momentum (MOM) factor, while the one-factor market model (i.e., model (1)) excludes the Small-Minus-Big (SMB) size factor, the High-Minus-Low (HML) value factor as well as the Momentum (MOM) factor.

For hedge funds, we present the results from the analogous Fung-Hsieh seven-factor model (c.f., Fung and Hsieh (1997, 2001)), instead of the Fama-French-Carhart four-factor model. These seven factors (i.e., Bond Trend-Following Factor, Currency Trend-Following Factor, Commodity Trend-Following Factor, Equity Market Factor, Size Spread Factor constructed from Russell 2000 index and S&P500, Bond Market Factor and Credit Spread Factor) proposed by Fung and Hsieh (1997, 2001) are arguably more suitable for the hedge funds than the Fama-French-Carhart four factors(for a recent example, see, e.g., Criton and Scaillet (2014)). For robustness, we have also tried the Fama-French-Carhart four-factor model and the results are available upon request.

3.3. Using traditional procedures to select the optimal number of fund subgroups from real data

Following previous simulation analysis, we follow Chen et al. (2017) and use the BIC to select the optimal number of fund subgroup from the real data sets of US mutual funds and hedge funds. The results from other traditional procedures such as Likelihood ratio statistics (see, e.g., Harvey and Liu (2018)) and Chi-squared statistics Ferson and Chen (2017)) are qualitatively similar and hence we delegate them to Appendix B. We present our results of BIC values over 500 replications for Scenarios 1, 2 and 3 in Table 6. To be specific, we report the BIC values when we specify two subgroups, three subgroups and four subgroups using GMD (2), GMD (3) and GMD (4) estimators. In all panels, for robustness, we use two thresholds to filter out our final sample: at least 8 months of returns, and at least 30 months of returns. We use the one-factor market model, the Fama-French three-factor model, the Fama-French-Carhart four-factor model and the Fung-Hsieh seven-factor model to estimate fund alpha in Panels A, B, C, and D, respectively.

Interestingly and surprisingly, BIC does not provide a consistent suggestion on the number of fund subgroups, as its suggestions vary from one underlying asset pricing model to another, and from one sample filtering method to another. In Panels B, C, and D, the BIC suggests the optimal number of fund subgroups is three if we filter our sample with at least 30 months of returns, but suggests at least four subgroups when we filter out our sample with at least 8 months of returns. In Panel A, the suggested number of subgroups is reversed, as the BIC value of -21729.28 (-24721.70) is smallest for four[7] (three) subgroups when we filter our sample with at least 30 (8) months of returns. Overall, BIC is not a reliable indicator to select the optimal number of fund subgroups. Consistent with our simulation results, the BIC values for GMD (2), GMD (3) and GMD (4) do not differ much from each other but greatly differ from that of GMD (1), no matter which scenario we look into. We further investigate the potential consequences brought by a not-so-wrong number of fund subgroups via both in-sample and out-of-sample likelihood score analysis.

3.4. *In-sample likelihood score analysis*

We first apply the likelihood score methodology in Amisano and Giacomini (2007) to examine the in-sample performance of GMD (2), GMD (3) and GMD (4).[8] For comparison reasons, we also considered the GMD (1) method (a single normal density). The in-sample likelihood score for a given density estimator $\widehat{f}(\alpha)$ can be computed by

$$LS_{is}(\widehat{f}(\alpha)) = \frac{1}{n}\sum_{i=1}^{n}\widehat{f}(\widehat{\alpha}_i), \tag{6}$$

where $\widehat{\alpha}_i$ is the OLS estimator of fund alpha for the ith fund, n is the total number of funds. In our case, $n = 3026$ or 2557 for mutual funds and $n = 3533$ or 2072 for hedge funds.

Table 7 presents the results of the in-sample likelihood score to two real data samples

[7] Of course, in this case, the number of fund subgroups suggested by BIC might be larger than four if we try GMD (5), GMD (6), etc. However, this issue is arguably trivial as more subgroups are not grounded in the literature and are hard to interpret with economic meaning. Hence, we do not consider this possibility in this paper.

[8] Some may argue that in-sample analysis is not necessary here, given our strong results from latter out-of-sample analysis. We agree to disagree as many existing studies on this topic including Harvey and Liu (2018) include both in-sample and out-of-sample analyses as well.

similar to Harvey and Liu (2018) and Ferson and Chen (2017) which cover both U.S. mutual funds and hedge funds, respectively. We present the values of likelihood score when we specify one subgroup, two subgroups, three subgroups and four subgroups using GMD (1), GMD (2), GMD (3) and GMD (4), respectively.[9] It is clear in Table 7 that the likelihood scores for GMD (2), GMD (3) and GMD (4) do not differ much from each other while all of them differ substantially from that of GMD (1), which reaffirms our previous simulation results that the information loss is trivial no matter whether using GMD (2), GMD (3) or GMD (4).

For instance, after estimating alphas via the Fama-French-Carhart four-factor model from the sample of mutual funds with at least 8 (30) months of return data, in Panel C we find that the likelihood scores of GMD (2), GMD (3) or GMD (4) (i.e., 108.1890, 109.2931 and 109.4282, respectively) are much larger than that of GMD (1), being 88.2091. The likelihood only increases by about 1% (i.e., 109.2931/108.1890-1) when we specify three subgroups (i.e., using GMD (3)) instead of two subgroups (i.e., using GMD (2)), and only further increases by less than 0.1% (i.e., 109.4282/109.2931-1) when we specify four subgroups (i.e., using GMD (4)) instead of three subgroups (i.e., using GMD (3)). This finding becomes stronger if we use the threshold of 8 months instead of 30 months to filter out our final sample, and the results obtained from different factor models in Panels A and B are qualitatively similar to the ones in Panel C.

[9]Different from the simulation analysis, we add the special case of one subgroup via GMD (1) for two reasons. i) to have something to compare with the results obtained from GMD (2), GMD (3) or GMD (4), although it is not the focus of this paper. ii) Ferson and Chen (2017) note that "The approach here also generalizes studies such as Kosowski et al. (2006); Fama and French (2010), who bootstrap the cross-section of mutual fund alphas. In those studies, all of the inferences are conducted under the null hypothesis of zero alphas, so there is only one group of funds. The analysis is directed at the hypothesis that all funds have zero alphas, accounting for the multiple hypothesis tests. The current approach also accounts for multiple hypothesis tests, but allows that some of the funds have nonzero alphas."

Despite a much shorter sample length, the results in Panel D obtained from our sample of U.S. hedge funds are also qualitatively similar to the ones obtained from our sample of U.S. mutual funds above. In general, if we use 30 months as the threshold, the likelihood scores of GMD (2), GMD (3) or GMD (4) (i.e., 62.7366, 62.7479 and 64.6349, respectively) are much larger than that of GMD (1), which is 50.3654. The likelihood increases by only about 0.01% (i.e., 62.7479/62.7366-1) when we specify three subgroups (i.e., using GMD (3)) instead of two subgroups (i.e., using GMD (2)), and further increases by only 3% (i.e., 64.6349/62.7479-1) when we specify four subgroups (i.e., using GMD (4)) instead of three subgroups (i.e., using GMD (3)). Again, this finding becomes much stronger if we use the threshold of 8 months instead of 30 months to filter out our final sample.

It is noteworthy that to stay on the conservative side we have not used any penalty factor to deal with the parameter uncertainty problem in the above analysis, i.e., we did not take the increased number of estimable parameters (parameter uncertainty) into consideration. The benefit regarding increased likelihood score brought by a specification of more fund subgroups, should be of a smaller magnitude if we take into account the parameter uncertainty (i.e., the increased number of estimable parameters). In untabulated results which are available from the authors upon request, we find supportive evidence using the cross-validation method.

3.5. Out-of-sample likelihood score analysis

To evaluate the out-of-sample (oos) performance of those above-mentioned density estimators, we first divide the whole sample into two sub-samples. The first sub-sample (in-sample) contains n_1 observations, which are used to estimate the density. The second sub-sample (out-of-sample) contains the rest of $n - n_1$ observations to validate the estimated density. According to Cheng et al. (2017), for a given density estimator $\widehat{f}(\alpha)$, the out-of-sample likelihood score of Amisano and Giacomini (2007) can be written as below:

$$LS_{oos}(\widehat{f}(\alpha)) = \frac{1}{n-n_1}\sum_{i=1}^{n-n_1}\widehat{f}(\widehat{\alpha}_i), \tag{7}$$

where n_1 is the size of in-sample for density estimation, $\widehat{\alpha}_i$ the ordinary least squares estimator of fund alpha for the ith fund, and n the total number of funds. For brevity we only report three choices of n_1, which are respectively $n/3$, $n/2$ and $2n/3$. For instance, when we let n_1 equal $n/3$, $n/2$ and $2n/3$ and use 30 months as the threshold, the out-of-sample length is 1705 (1381), 1279 (1036) and 852 (691) for mutual (hedge) funds, respectively.

Table 8 presents the results from applying the aforementioned out-of-sample likelihood scores to the mutual funds and hedge funds data sets. We can see that the out-of-sample performance of GMD (1), GMD (2), GMD (3) and GMD (4) has a similar pattern to their in-sample performance. For all our samples of mutual funds and hedge funds, no matter what value n_1 takes ($n/3$, $n/2$ or $2n/3$), GMD (2), GMD (3) and GMD (4) have similar performance, far better than that of GMD (1). Taking the sample of hedge funds with at least 30 months of returns data as an example, when n_1 equals $n/3$ and the out-of-sample length is 1381, the out-of-sample likelihood scores for GMD (1), GMD (2), GMD (3) and GMD (4) are 46.6932, 62.3785, 63.0389 and 63.3135, respectively. When n_1 equals $2n/3$ and the out-of-sample length is 691, the out-of-sample likelihood scores for GMD (1), GMD (2), GMD (3) and GMD (4) are 49.0155, 60.4667, 61.9320 and 61.9059, respectively.

4. Concluding remarks

The Gaussian Mixture Distribution (GMD) approach has recently become increasingly popular in determining the number of fund subgroups prior to performance evaluation, while it is surprising that no rigorous statistical analysis has been conducted to examine the information loss of this model. This paper presents evidence that the traditional procedures do not always select the correct number of fund subgroups, and we use both Monte Carlo simulations and real data sets to evaluate the possible information loss in the GMD approach due to model misspecification and parameter uncertainty. We find that parameter uncertainty is more important than model misspecification when using GMD, since the information loss (in terms of mean validity, density validity and reliability under three scenarios) is smallest when there are the fewest parameters to be estimated, not when the number of GMD components is correctly specified. Our results stress the importance of parameter uncertainty, which echoes the portfolio selection literature on the same problem (Brown (1979); Jobson and Korkie (1980); Jorion (1986); Garlappi et al. (2006); DeMiguel et al. (2007); Yan and Zhang (2017)).

There are, of course, caveats to our analysis. Ideally, we need an economic or financial theory that captures the evolution of fund subgroups. In the absence of such a generally accepted theory, we rely on statistical and econometric tools to gauge the number of fund subgroups within both simulated and real data. Regarding real data, we follow the mainstream literature and focus on U.S. equity mutual funds and U.S. equity-oriented hedge funds given the lion's share of their market size. In this paper, we consider neither pension funds, bond mutual funds, nor funds in other developed and emerging markets, which is a fruitful direction for future research. Although we did not look into every possible data generating process and econometric tool, we believe that our results cannot be easily qualitatively altered given their current robustness and the information loss in the GMD model is less than theories lead one to believe. Albeit of importance, in this paper, we did not try to introduce the concepts of Order Statistics or the bootstrap methodology to differentiate fund subgroups by luck and skill, since these have been investigated in companion studies (Cheng and Yan (2017); Cai et al. (2018); Zhang and Yan (2018)). There are surely other exciting related questions to ask, since this area is of obvious importance and far from completed.

Reference

Abramson, I. S. (1982), 'On bandwidth variation in kernel estimates-a square root law', *Annals of Statistics* **10**(4), 1217–1223.

Amisano, G. and Giacomini, R. (2007), 'Comparing density forecasts via weighted likelihood ratio tests', *Journal of Business & Economic Statistics* **25**(2), 177–190.

Bajgrowicz, P. and Scaillet, O. (2012), 'Technical trading revisited: False discoveries, persistence tests, and transaction costs', *Journal of Financial Economics* **106**(3), 473–491.

Bajgrowicz, P., Scaillet, O. and Treccani, A. (2015), 'Jumps in high-frequency data: Spurious detections, dynamics, and news', *Management Science* **62**(8), 2198–2217.

Barras, L., Scaillet, O., Wermers, R. et al. (2010), 'False discoveries in mutual fund performance: Measuring luck in estimated alphas', *Journal of Finance* **65**(1), 179–216.

Bhattacharyya, A. K. (1943), 'On a measure of divergence between two statistical populations defined by their probability distributions', *Bulletin of the Calcutta Mathematical Society* **35**, 99–109.

Breiman, L., Meisel, W. and Purcell, E. (1977), 'Variable kernel estimates of multivariate densities', *Technometrics* **19**(2), 135–144.

Brown, S. (1979), 'The effect of estimation risk on capital market equilibrium', *Journal of Financial and Quantitative Analysis* **14**(2), 215–220.

Cai, B., Cheng, T. and Yan, C. (2018), 'Time-varying skills (versus luck) in U.S. active mutual funds and hedge funds', *Journal of Empirical Finance* **49**(12), 81–106.

Carhart, M. M. (1997), 'On persistence in mutual fund performance', *Journal of Finance* **52**(1), 57–82.

Chen, X., Ponomareva, M. and Tamer, E. (2014), 'Likelihood inference in some finite mixture models', *Journal of Econometrics* **182**(1), 87–99.

Chen, Y., Cliff, M. and Zhao, H. (2017), 'Hedge funds: The good, the bad, and the lucky', *Journal of Financial and Quantitative Analysis* **52**(3), 1081–1109.

Cheng, T., Gao, J. and Zhang, X. (2017), 'Semiparametric localized bandwidth selection in kernel density estimation', *Econometric Reviews* **forthcoming**.

Cheng, T. and Yan, C. (2017), 'Evaluating the size of the bootstrap method for fund performance evaluation', *Economics Letters* **156**(7), 36–41.

Criton, G. and Scaillet, O. (2014), 'Hedge fund managers: Luck and dynamic assessment', *Bankers, Markets & Investors* **129**(3-4), 28–38.

DeMiguel, V., Garlappi, L. and Uppal, R. (2007), 'Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?', *Review of Financial Studies* **22**(5), 1915–1953.

Elton, E. J., Gruber, M. J. and Blake, C. R. (2001), 'A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases', *Journal of Finance* **56**(6), 2415–2430.

Evans, R. B. (2010), 'Mutual fund incubation', *Journal of Finance* **65**(4), 1581–1611.

Fama, E. F. and French, K. R. (2010), 'Luck versus skill in the cross-section of mutual fund returns', *Journal of Finance* **65**(5), 1915–1947.

Ferson, W. and Chen, Y. (2017), 'How many good and bad fund managers are there, really?', *University of Southern California working paper* .

Fung, W. and Hsieh, D. A. (1997), 'Empirical characteristics of dynamic trading strategies: The case of hedge funds', *Review of Financial Studies* **10**(2), 275–302.

Fung, W. and Hsieh, D. A. (2001), 'The risk in hedge fund strategies: Theory and evidence from trend followers', *Review of Financial Studies* **14**(2), 313–341.

Garlappi, L., Uppal, R. and Wang, T. (2007), 'Portfolio selection with parameter and model uncertainty: A multi-prior approach.', *Review of Financial Studies* **20**(1), 41–81.

Hansen, B. E. (2008), 'Uniform convergence rates for kernel estimation with dependent data', *Econometric Theory* **24**(03), 726–748.

Harvey, C. R. and Liu, Y. (2018), 'Detecting repeatable performance', *Review of Financial Studies* **31**(7), 2499–2552.

Hunter, D., Kandel, E., Kandel, S. and Wermers, R. (2014), 'Mutual fund performance evaluation with active peer benchmarks', *Journal of Financial Economics* **112**(1), 1–29.

Jobson, J. D. and Korkie, B. (1980), 'Estimation for markowitz efficient portfolios', *Journal of the American Statistical Association* **75**(371), 544–554.

Jorion, P. (1986), 'Bayes-stein estimation for portfolio analysis', *Journal of Financial and Quantitative Analysis* **21**(3), 279–292.

Kosowski, R., Timmermann, A., Wermers, R. and White, H. (2006), 'Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis', *Journal of Finance* **61**(6), 2551–2595.

Kullback, S. and Leibler, R. A. (1951), 'On information and sufficiency', *The annals of mathematical statistics* **22**(1), 79–86.

Marron, J. and Wand, M. (1992), 'Exact mean integrated squared error', *Annals of Statistics* **20**(2), 712–736.

Pittau, M. G., Zelli, R. and Massari, R. (2016), 'Evidence of convergence clubs using mixture models', *Econometric Reviews* **35**(7), 1317–1342.

Scaillet, O., Treccani, A. and Trevisan, C. (2018), 'High-frequency jump analysis of the bitcoin market', *Journal of Financial Econometrics* **forthcoming**.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall/CRC, London.

Yan, C. and Zhang, H. (2017), 'Mean-variance versus naïve diversification: The role of mispricing', *Journal of International Financial Markets, Institutions and Money* **48**(5), 61–81.

Zhang, H. and Yan, C. (2018), 'A skeptical appraisal of the bootstrap approach in fund performance evaluation', *Financial Markets, Institutions and Instruments* **27**(2), 49–86.

Figure 1: **Graphical performance of density estimators.** Density estimates by adaptive kernel estimator, global kernel estimator, GMD (2), GMD (3) and GMD (4) in Scenario 1, Scenario 2 and Scenario 3 based on one replication path.

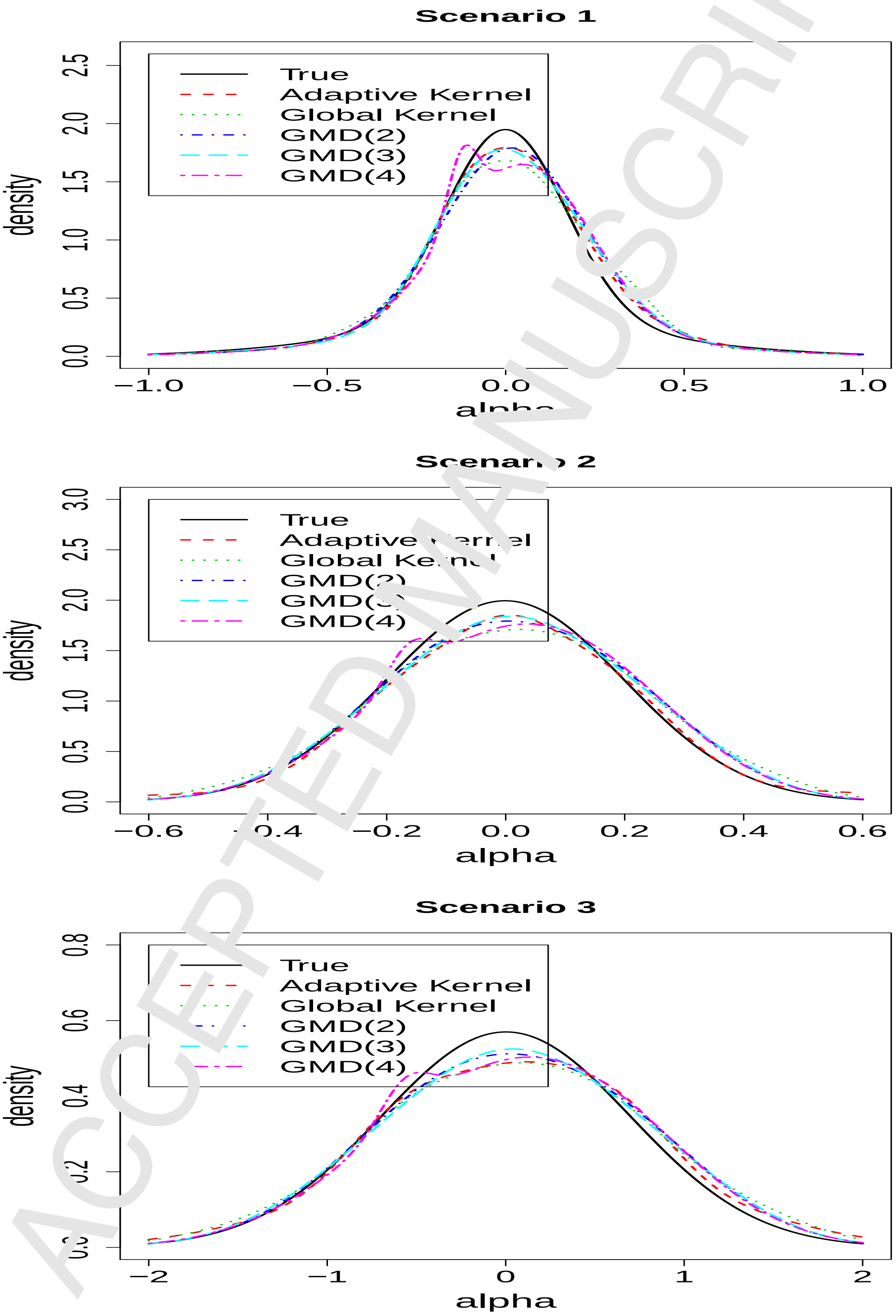

Table 1: **Summary of Density estimators.** The first column presents the five density estimators used in this paper, including adaptive kernel density estimator, global kernel density estimator, GMD (2), GMD (3) and GMD (4). The second column summarizes the unknown estimable parameters involved in each estimator.

Density estimators	Unknown parameters
GMD (2)	$\pi_1, \mu_1, \mu_2, \sigma_1, \sigma_2$
GMD (3)	$\pi_1, \pi_2, \mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3$
GMD (4)	$\pi_1, \pi_2, \pi_3, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1, \sigma_2, \sigma_3, \sigma_4$
Global kernel estimator with fixed bandwidth	h
Adapative kernel estimator with variable bandwidth	h_i

Table 2: **Generation of alpha in the simulation study.** Based on a CAPM-type of single factor model, we considered the following three representative scenarios for the generation of α_i, for $i = 1, 2, \cdots, N$. In Scenario 1, we generate alpha from a mixture of two Gaussian distributions, which follows Table 1 of Harvey and Liu (2018). In Scenario 2, we generate alpha from a mixture of three Gaussian distributions, which follows Panel C of Table IV of Ferson and Chen (2017). In Scenario 3, we generate alpha from a mixture of four Gaussian distributions, which follows Table A1 of Chen et al. (2017).

Scenario	Simulated distribution for alpha
Scenario 1	$0.283N(-0.02277/12, 1.513^2/12) + 0.717N(-0.00685/12, 0.586^2/12)$
Scenario 2	$0.507N(0, 0.2^2) + 0.069N(-0.0003, 0.2^2) + 0.424N(-0.002, 0.2^2)$
Scenario 3	$0.1N(0.01, 0.7^2) + 0.4N(0.003, 0.7^2) + 0.4N(0, 0.7^2) + 0.1N(-0.01, 0.7^2)$

Table 3: **BIC model selection results for simulated data.** This table reports means of BIC values over 500 replications for GMD (2), GMD (3) and GMD (4) estimators with the corresponding frequency of being selected in parentheses.

Scenario	GMD (2)	GMD (3)	GMD (4)
Scenario 1	691.4137 (489/500)	710.5049 (8/500)	730.4898 (3/500)
Scenario 2	369.3813 (481/500)	390.5060 (18/500)	412.2884 (1/500)
Scenario 3	811.6118 (492/500)	831.3073 (8/500)	851.5021 (0/500)

Table 4: **Performance of density estimators.** In Panel A, we present the results of mean validity, that is, the bias($\times 10^3$) of estimated mean parameter of alpha resulted from adaptive kernel estimator, global kernel estimator, GMD (2), GMD (3) and GMD (4) based on 200 grid points when $(N, T) = (400, 200)$. In Panel B and Panel C, we respectively present the results of density validity in terms of MISE and reliability in terms of SDISE resulted from the above five density estimators. In Panel A, we report the value of bias$\times 10^3$ to be reader-friendly.

Panel A: Mean Validity						
Scenario	Criteria	Adaptive Kernel	Global Kernel	GMD (2)	GMD (3)	GMD (4)
Scenario 1	Bias	0.998830	0.998830	0.999095	0.999249	1.002212
Scenario 2	Bias	0.751167	0.751167	0.751628	0.751840	0.755070
Scenario 3	Bias	2.630384	2.630384	2.630893	2.634982	2.640672
Panel B: Density Validity						
Scenario		Adaptive Kernel	Global Kernel	GMD (2)	GMD (3)	GMD (4)
Scenario 1	MISE	0.003435	0.00[illegible]141	0.003437	0.007097	0.013260
Scenario 2	MISE	0.006366	0.009401	0.008427	0.016075	0.026933
Scenario 3	MISE	0.000716	0.000[illegible]02	0.000721	0.001379	0.002320
Panel C: Reliability						
Scenario		Adaptive Kernel	Global Kernel	GMD (2)	GMD (3)	GMD (4)
Scenario 1	SDISE	0.002569	0.003728	0.004004	0.007079	0.027509
Scenario 2	SDISE	0.004053	0.006106	0.010818	0.014136	0.025306
Scenario 3	SDISE	0.00047[illegible]	0.000522	0.001023	0.001227	0.002160

Table 5: **Summary statistics.** Monthly returns are summarized for mutual funds (top panel) and hedge funds (bottom panel), measured in excess of the one-month return of a three-month Treasury bill. The values at the cutoff points for various quantiles of the cross-sectional distributions of the sample of funds are reported. Each column is sorted on the statistic shown. Nobs is the number of available monthly returns, where for the left top (and left bottom) panel, there is no restriction while a minimum of 30 are required for the right top (and right bottom) panel. Mean is the sample mean return, Std the sample standard deviation of return, and Rho1 the first order sample autocorrelation. The alpha estimates are based on OLS regressions using the Fama-French-Carhart four factors (Carhart (1997)) for mutual funds, while the Fung-Hsieh seven factors (Fung and Hsieh (1997, 2001) are used for the hedge funds.

Quantiles	Mutual funds (full sample)					Mutual funds (minimum 30 obs)				
	Nobs	Mean	Std	Rho1	$\widehat{\alpha}_{ols}$	Nobs	Mean	Std	Rho1	$\widehat{\alpha}_{ols}$
Top	335	0.060	0.512	0.688	0.032	335	0.060	0.512	0.688	0.024
1%	333	0.021	0.117	0.406	0.008	325	0.018	0.114	0.361	0.008
5%	263	0.013	0.088	0.303	0.004	277	0.012	0.087	0.284	0.004
10%	223	0.010	0.078	0.254	0.003	232	0.010	0.077	0.243	0.003
20%	178	0.008	0.069	0.207	0.001	190	0.007	0.068	0.205	0.001
30%	149	0.006	0.062	0.172	0.001	163	0.006	0.062	0.173	0.001
Median	97	0.004	0.053	0.121	-0.000	118	0.004	0.054	0.127	-0.000
30%	53	0.002	0.046	0.063	-0.001	76	0.002	0.047	0.079	-0.001
20%	38	-0.000	0.042	0.020	-0.002	58	0.001	0.043	0.049	-0.002
10%	22	-0.003	0.036	-0.057	-0.003	44	-0.002	0.038	0.000	-0.003
5%	13	-0.008	0.030	-0.121	-0.005	38	-0.004	0.034	-0.052	-0.005
1%	9	-0.023	0.018	-0.287	-0.010	32	-0.010	0.022	-0.149	-0.009
Bottom	8	-0.090	0.004	-0.627	-0.141	31	-0.035	0.004	-0.551	-0.049
Quantiles	Hedge funds (full sample)					Hedge funds (minimum 30 obs)				
	Nobs	Mean	Std	Rho1	$\widehat{\alpha}_{ols}$	Nobs	Mean	Std	Rho1	$\widehat{\alpha}_{ols}$
Top	192	0.173	0.695	0.814	0.868	192	0.051	0.324	0.814	0.045
1%	172	0.026	0.173	0.579	0.024	182	0.021	0.156	0.584	0.020
5%	126	0.014	0.098	0.457	0.012	147	0.012	0.090	0.479	0.011
10%	102	0.009	0.075	0.390	0.008	124	0.009	0.071	0.409	0.008
20%	73	0.006	0.053	0.296	0.005	96	0.006	0.052	0.323	0.005
30%	56	0.004	0.042	0.234	0.004	78	0.004	0.042	0.265	0.004
Median	38	0.001	0.028	0.127	0.002	57	0.002	0.029	0.170	0.002
30%	22	-0.002	0.020	0.009	-0.000	46	-0.000	0.022	0.078	0.000
20%	16	-0.005	0.016	-0.072	-0.001	40	-0.002	0.018	0.021	-0.001
10%	11	-0.011	0.012	-0.188	-0.004	36	-0.005	0.014	-0.071	-0.003
5%	8	-0.018	0.009	-0.304	-0.009	33	-0.008	0.010	-0.133	-0.006
1%	3	-0.043	0.005	-0.518	-0.024	31	-0.018	0.007	-0.284	-0.014
Bottom	1	-0.114	0.000	-0.794	-1.513	31	-0.038	0.001	-0.492	-0.031

Table 6: **BIC model selection results for real data.** Panel A, B and C report the BIC values to the real data set of U.S. mutual funds with a sample similar to Ferson and Chen (2017); Harvey and Liu (2018), while Panel D presents the counterpart results from the real data set of U.S. hedge funds with a sample similar to Ferson and Chen (2017). In all panels, for robustness we use two thresholds to filter out our final sample: at least 8 months of returns, and at least 30 months of returns. We use the one-factor market model, the Fama-French three-factor model, the Fama-French-Carhart four-factor model and the Fung-Hsieh seven-factor model to estimate fund alpha in Panel A, B, C and D, respectively. We present the BIC values when we specify one subgroup, two subgroups, three subgroups and four subgroups using GMD (1), GMD (2), GMD (3) and GMD (4), respectively. The selected specifications are in bold.

Real data sample	n	GMD (1)	GMD (2)	GMD (3)	GMD (4)
Panel A: Mutual funds via the one-factor market model					
Mutual funds with at least 8 obs	3026	−23382.27	−24633.67	**−24721.70**	−24715.99
Mutual funds with at least 30 obs	2557	−21054.07	−21673.89	−21666.05	**−21729.28**
Panel B: Mutual funds via the Fama-French three-factor model					
Mutual funds with at least 8 obs	3026	−23386.16	−25225.93	−25240.73	**−25383.28**
Mutual funds with at least 30 obs	2557	−21480.77	−22135.80	**−22188.04**	−22167.64
Panel C: Mutual funds via the Fama-French-Carhart four-factor model					
Mutual funds with at least 8 obs	3026	−22538.89	−25123.16	−25367.55	**−25384.46**
Mutual funds with at least 30 obs	2557	−21587.96	−22246.82	**−22282.35**	−22263.22
Panel D: Hedge funds via the Fung-Hsieh seven-factor model					
Hedge funds with at least 8 obs	35[illegible]	−21839.24	−21881.95	−22651.32	**−22772.75**
Hedge funds with at least 30 obs	2072	−15143.04	**−15725.49**	−15719.51	−15719.47

Table 7: **In-sample likelihood score comparison for real data.** Panel A, B and C present the results when apply the well-known likelihood score methodology in Amisano and Giacomini (2007) to the real data set of U.S. mutual funds with a sample similar to Ferson and Chen (2017) and Harvey and Liu (2018), while Panel D presents the counterpart results from the real data set of U.S. hedge funds with a sample similar to Ferson and Chen (2017). In all panels, for robustness we use two thresholds to filter out our final sample: at least 8 months of returns, and at least 30 months of returns. We use the one-factor market model, the Fama-French three-factor model, the Fama-French-Carhart four-factor model and the Fung-Hsieh seven-factor model to estimate fund alpha in Panel A, B, C and D, respectively. We present the values of in-sample likelihood scores when we specify one subgroup, two subgroups, three subgroups and four subgroups using GMD (1), GMD (2), GMD (3) and GMD (4), respectively.

Real data sample	n	GMD (1)	GMD (2)	GMD (3)	GMD (4)
Panel A: Mutual funds via the one-factor market model					
Mutual funds with at least 8 obs	3026	63.4266	84.8062	86.8392	87.2844
Mutual funds with at least 30 obs	2557	79.3837	95.3619	96.1793	97.9721
Panel B: Mutual funds via the Fama-French three-factor model					
Mutual funds with at least 8 obs	3026	65.3312	94.3219	97.1321	98.5232
Mutual funds with at least 30 obs	2557	8[illegible].5160	106.1217	107.9068	108.0082
Panel C: Mutual funds via the Fama-French-Carhart four-factor model					
Mutual funds with at least 8 obs	3026	59.3210	94.7754	98.5811	99.2925
Mutual funds with at least 30 obs	2557	88.2091	108.1890	109.2931	109.4282
Panel D: Hedge funds via the Fung-Hsieh seven-factor model					
Hedge funds with at least 8 obs	3583	8.4653	37.1801	43.4041	43.8901
Hedge funds with at least 30 obs	2072	50.3654	62.7366	62.7479	64.6349

Table 8: **Out-of-sample likelihood score comparison for real data.** Panel A, B, and C present the results when apply the well-known likelihood score methodology in Amisano and Giacomini (2007) to the real data set of U.S. mutual funds with a sample similar to Ferson and Chen (2017) and Harvey and Liu (2018), while Panel D presents the counterpart results from the real data set of U.S. hedge funds with a sample similar to Ferson and Chen (2017). In all panels, for robustness, we use two thresholds to filter out our final sample: at least 8 months of returns, and at least 30 months of returns. We use the one-factor market model, the Fama-French three-factor model, the Fama-French-Carhart four-factor model and the Fung-Hsieh seven-factor model to estimate fund alpha in Panel A, B, C, and D, respectively. We present the values of out-of-sample likelihood score when we specify one subgroup, two subgroups, three subgroups and four subgroups using GMD (1), GMD (2), GMD (3) and GMD (4), respectively.

Real data sample	$n-n_1$	GMD (1)	GMD (2)	GMD (3)	GMD (4)
Panel A: Mutual funds via the one-factor market model					
Mutual funds with at least 8 obs	2017	66.3361	86.7579	88.4501	88.5952
	1513	66.8874	85.0836	86.6306	86.9012
	1009	66.4014	82.7363	82.9075	84.1940
Mutual funds with at least 30 obs	1705	72.6077	96.0531	98.0006	97.6183
	1279	76.0094	95.0133	97.3747	97.7754
	852	76.1577	94.9824	96.8780	97.6143
Panel B: Mutual funds via the Fama-French three-factor model					
Mutual funds with at least 8 obs	2017	73.3779	98.4090	99.2168	99.3773
	1513	75.6775	96.0021	97.9590	97.9618
	1009	74.0500	92.3150	94.3411	94.3026
Mutual funds with at least 30 obs	1705	80.3043	107.2863	107.3153	108.3281
	1279	83.4818	106.0315	107.4896	107.0795
	852	84.9855	104.5311	106.2564	105.8741
Panel C: Mutual funds via the Fama-French-Carhart four-factor model					
Mutual funds with at least 8 obs	2017	73.5300	99.7986	101.0600	101.0935
	1513	75.9806	97.4218	98.8737	98.9133
	1009	73.4680	92.7965	94.6067	94.6055
Mutual funds with at least 30 obs	1705	82.8348	109.2539	110.1331	110.1930
	1279	85.8566	107.7597	109.0251	109.0295
	852	87.3349	105.8908	107.3995	107.4310
Panel D: Hedge funds via the Fung-Hsieh seven-factor model					
Hedge funds with at least 8 obs	2355	11.9880	38.4978	44.2847	44.5295
	1767	10.7027	38.6438	43.1439	43.4138
	1178	10.3157	36.2571	39.6653	39.8822
Hedge funds with at least 30 obs	1381	46.6932	62.3785	63.0389	63.3135
	1036	48.6866	62.1095	62.8484	62.9377
	691	49.0155	60.4667	61.9320	61.9059

Appendix A

This appendix briefly introduces three density estimators: the Gaussian Mixture Distribution (GMD) estimator, the global kernel density estimator, and the adaptive kernel density estimator.

GMD is one of the most common parametric estimators in Economics, and it can provide a natural representation of heterogeneity in a finite number of latent classes (see Chen et al., 2014). If fund alpha α is deemed to be generated from a mixture (weighted average) of k Gaussian distributions, the corresponding GMD (k) estimator of x is:

$$\widehat{f}_{GMD(k)}(\alpha) = \sum_{i=1}^{k} \pi_i f_i(\alpha), \tag{8}$$

where $f_i(\alpha) = (2\pi\sigma_i)^{-1/2} \exp\left\{-\frac{(\alpha-\mu_i)^2}{2\sigma_i^2}\right\}$, $\sigma_i > 0$, π_i are weights satisfying $\pi_i > 0$ and $\sum_{i=1}^{k} \pi_i = 1$. Empirically, GMD is almost always an approximation of the true density as the components are unlikely to be normal, especially in Finance. Although a few papers argue that GMD can capture a large domain of non-normal complex distributions even when the number of Gaussian components, k, is small (e.g., Marron and Wand (1992)), the use of GMD suffers from two aspects regarding the choice of k: model misspecification and parameter uncertainty. On the one hand, since the true value of k is unobservable and may not be an integer (if not non-existent) in real data, it is not uncommon to mis-specify the GMD estimator with a wrong choice of k, since the extant studies mainly rely on traditional procedures to pick up a reasonably small integer. For instance, Harvey and Liu (2018), Ferson and Chen (2017) and Chen et al. (2017) only rely on a Likelihood Ratio (LR) test, a Chi-squared test and a Bayesian Information Criterion (BIC), respectively. Recently, Pittau et al. (2016) proposed a kernel-based test to determine the number of components k. However, this test does not have sufficient power to reject the hypothesis of the number of components smaller than that in the true model when the components largely overlap. On the other hand, a GMD (k) estimator has $3k-1$ unknown parameters to be estimated, which increases at k as we have made it clear in Table 1. The larger k is, the more estimation errors occur, and the larger the parameter uncertainty problem. Albeit these reasonable and probable doubts, the existence and magnitude of information loss due to these two concerns remain unclear.

To answer this question, we do not only focus on the variation of information loss within the specified GMD (k) estimators, but also compare them with potential competitors which alleviate the above two concerns and better capture the distribution of alphas. Nonparametric estimators are of special interests for these purposes: i) we can establish the uniform consistency results for kernel density estimators (see Hansen, 2008) and their performance is super robust due to their nonparametric nature and hence can easily beat their parametric counterparts in the presence of misspecification; ii) usually they suffer the least from the parameter uncertainty problem as they only require one parameter (bandwidth) to be estimated. There are two basic nonparametric kernel density estimators: global kernel estimator with fixed bandwidth, and adaptive kernel estimator with variable bandwidth. For a random sample $\alpha_1, \alpha_2, \cdots, \alpha_n$ drawn from a density $f(\alpha)$, the global kernel density estimator is as follows:

$$\hat{f}_{global}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{\alpha-\alpha_i}{h}\right), \tag{9}$$

where $K(\cdot)$ is the Gaussian kernel function without loss of generality and h is bandwidth satisfying that $h \to 0$ as $n \to \infty$. Observe that $\frac{1}{h}K\left(\frac{\alpha-\alpha_i}{h}\right)$ is actually the density function for the Gaussian distribution with mean α_i and variance h^2. Therefore, the above kernel density estimator can be regarded as a location-mixture of n Gaussian distributions and the Gaussian component densities have a common variance (denoted as h^2) and individual mean values. Unlike the traditional GMD (k) estimator with $3k-1$ unknown estimable parameters, there is only one parameter h to be estimated in the nonparametric global kernel density estimator.

Besides the above global kernel density estimator with fixed bandwidth, we also consider the sample-point adaptive kernel estimator with variable bandwidth, which was first introduced by Breiman et al. (1977) and nests the global kernel density estimator as a special case. The functional form of the sample-point adaptive kernel estimator is

$$\widehat{f}_{adaptive}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}K\left(\frac{\alpha-\alpha_i}{h_i}\right), \tag{10}$$

where $h_i = h(\alpha_i)$ is a function of α_i. Note that the adaptive kernel density estimator can also be regarded as a mixture of n Gaussian densities but each Gaussian density component has its own variance (denoted as h_i^2) and individual mean value. Without loss of generality, we simply let $h_i \propto f(\alpha_i)^{-1/2}$, and Abramson (1982) suggest that this adaptive choice outperforms the fixed bandwidth estimator $\widehat{f}_{global}(\alpha)$ in general.

Appendix B

Likelihood ratio test

The likelihood ratio statistic is commonly used to compare the goodness of fit of two statistical models, one of which (the null model) is a special case of the other (the alternative model). Let L_0 and L_1 denote the likelihood function evaluated at the model estimates for M_0 and M_1, respectively. Following Harvey and Liu (2018), we compute the likelihood-ratio statistic by

$$LR = -2(\log L_0 - \log L_1), \tag{11}$$

When M_1 significantly outperforms M_0, LR will be large and positive. Therefore, a large likelihood ratio statistic provides evidence against M_0. We present our results for the simulated data and real data in Table 9 and Table 10, respectively.

Similar to the fund subgroup selection results using BIC, in all three simulated scenarios the Likelihood ratio test suggests us to select GMD (2) as GMD (2) beats GMD (3) and GMD (4) in at least 87.4% (=437/500 in Scenario 2) cases. In other words, the Likelihood ratio test does not always select the right number of fund subgroups. In real data analysis, the Likelihood ratio test also does not provide a consistent suggestion on the number of fund subgroups, as its suggestion varies from one underlying asset pricing model to another, and from one sample filtering method to another.

Chi-squared test

Following Ferson and Chen (2017), we employ the Pearson Chi-squared statistic below as one of the criteria for density estimation of fund alphas.

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - M_i)^2}{O_i}, \tag{12}$$

where the sum is over K cells, O_i is the frequency of OLS estimated fund alphas that appear in cell i in the data sample, and M_i is the frequency of fund alphas that appear in cell i using a certain density estimation method. The null hypothesis is that the expected frequencies from the density estimation method match those of the data sample. In applications, we choose K cells, with the cell boundaries set so that an approximately equal number of fund alphas in the data sample appear in each cell. Since in our simulated data we only have 400 funds which is much less than the two thousand funds in Ferson and Chen (2017), we set $K = 10$ to make sure there are plenty of observations falling into each cell (i.e., 40 fund alphas in each cell). In real data analysis, we set $K = 101$ to roughly maintain consistency with Ferson and Chen (2017). We present our results for the simulated data and real data in Table 11 and Table 12, respectively. The results are consistent with our previous results using BIC and the Likelihood ratio test.

Table 9: **Likelihood ratio model selection results for simulated data.** This table reports the results of three likelihood ratio tests, (i) $H_0 : \alpha \sim GMD(2)$ against $H_1 : \alpha \sim GMD(3)$; (ii) $H_0 : \alpha \sim GMD(2)$ against $H_1 : \alpha \sim GMD(4)$; (iii) $H_0 : \alpha \sim GMD(3)$ against $H_1 : \alpha \sim GMD(4)$. Panel A and B respectively present the frequency of each density estimator being favored for each of the above test under 5% and 1% level of significance.

Scenario	GMD (2) VS GMD (3)		GMD (2) VS GMD (4)		GMD (3) VS GMD (4)	
	GMD (2)	GMD (3)	GMD (2)	GMD (4)	GMD (3)	GMD (4)
Panel A: 5% significance level						
Scenario 1	452	48	438	62	418	82
Scenario 2	437	63	448	52	407	93
Scenario 3	439	61	445	55	408	92
Panel B: 1% significance level						
Scenario 1	482	18	478	22	472	28
Scenario 2	473	27	480	20	465	35
Scenario 3	475	25	482	18	467	33

Table 10: **Likelihood ratio model selection results for real data.** Panel A, B and C report the LR test statistics to the real data set of U.S. mutual funds with a sample similar to Ferson and Chen (2017); Harvey and Liu (2018), while Panel D presents the counterpart results from the real data set of U.S. hedge funds with a sample similar to Ferson and Chen (2017). In all panels, for robustness we use two thresholds to filter out our final sample: at least 8 months of returns, and at least 30 months of returns. We use the one-factor market model, the Fama-French three-factor model, the Fama-French-Carhart four-factor model and the Fung-Hsieh seven-factor model to estimate fund alpha in Panel A, B, C and D, respectively. We present the LR test statistics when we test GMD (1) against GMD (2), GMD (2) against GMD (3), GMD (2) against GMD (4) and GMD (3) against GMD (4), respectively. The cases in which we cannot reject the null that the former is as good as the latter at 1% significance level are in bold.

Real data sample	n	(1) vs (2)	(2) vs (3)	(2) vs (4)	(3) vs (4)
Panel A: Mutual funds via the one-factor market model					
Mutual funds with at least 8 obs	3026	1275.44	112.08	130.46	18.37
Mutual funds with at least 30 obs	2557	643.36	80.76	101.90	21.13
Panel B: Mutual funds via the Fama-French three-factor model					
Mutual funds with at least 8 obs	3026	1063.81	161.24	205.44	44.20
Mutual funds with at least 30 obs	2557	678.57	75.78	84.27	**8.49**
Panel C: Mutual funds via the Fama-French-Carhart four-factor model					
Mutual funds with at least 8 obs	3026	2508.31	268.44	309.40	40.96
Mutual funds with at least 30 obs	2557	682.40	59.07	64.83	**5.76**
Panel D: Hedge funds via the Fung-Hsieh seven-factor model					
Hedge funds with at least 8 obs	3533	10067.23	793.88	939.82	145.94
Hedge funds with at least 30 obs	2072	605.37	35.20	39.80	**4.61**
1% critical value		11.34	11.34	16.81	11.34
5% critical value		7.81	7.81	12.59	7.81

Table 11: **Chi-squared model selection results for simulated data.** This table reports the results of three Chi-squared tests, (i) $H_0 : \alpha \sim GMD(2)$ against $H_1 : \alpha$ does not follow GMD (2); (ii) $H_0 : \alpha \sim GMD(3)$ against $H_1 : \alpha$ does not follow GMD (3); (iii) $H_0 : \alpha \sim GMD(4)$ against $H_1 : \alpha$ does not follow GMD (4). Panel A presents the averaged Chi-squared statistic for each of the above tests over 500 replications; Panel B and C respectively present the frequency of each null being supported for the above test under 5% and 1% level of significance.

Scenario	GMD (2)	GMD (3)	GMD (4)
Panel A: Average statistic			
Scenario 1	15.91	17.26	18.[illegible]2
Scenario 2	15.28	16.97	18.33
Scenario 3	16.01	17.73	19.06
Panel B: 5% critical value = 16.92			
Scenario 1	348	27[illegible]	254
Scenario 2	365	272	241
Scenario 3	331	253	206
Panel C: 1% critical value = 21.67			
Scenario 1	459	424	374
Scenario 2	46[illegible]	415	373
Scenario 3	452	405	344

Table 12: **Chi-squared model selection results for real data.** Panel A, B and C report the Chi-squared statistic to the real data set of U.S. mutual funds with a sample similar to Ferson and Chen (2017); Harvey and Liu (2018), while Panel D presents the counterpart results from the real data set of U.S. hedge funds with a sample similar to Ferson and Chen (2017). In all panels, for robustness we use two thresholds to filter out our final sample: at least 8 months of returns, and at least 30 months of returns. We use the one-factor market model, the Fama-French three-factor model, the Fama-French-Carhart four-factor model and the Fung-Hsieh seven-factor model to estimate fund alpha in Panel A, B, C and D, respectively. We present the Chi-squared statistic when we specify one subgroup, two subgroups, three subgroups and four subgroups using GMD (1), GMD (2), GMD (3) and GMD (4), respectively. The specifications which have passed the Chi-squared test at 1% level are in bold.

Real data sample	n	GMD (1)	GMD (2)	GMD (3)	GMD (4)
Panel A: Mutual funds via the one-factor market model					
Mutual funds with at least 8 obs	3026	1000.66	166.57	**93.40**	**91.13**
Mutual funds with at least 30 obs	2557	503.05	142.84	**105.53**	**87.92**
Panel B: Mutual funds via the Fama-French three-factor model					
Mutual funds with at least 8 obs	3026	1481.71	193.21	**115.67**	**79.91**
Mutual funds with at least 30 obs	2557	541.54	**126.54**	**91.22**	**92.19**
Panel C: Mutual funds via the Fama-French-Carhart four-factor model					
Mutual funds with at least 8 obs	3026	2136.06	254.30	**110.59**	**95.55**
Mutual funds with at least 30 obs	2557	567.83	**123.04**	**95.51**	**93.47**
Panel D: Hedge funds via the Fung-Hsieh seven-factor model					
Hedge funds with at least 8 obs	3583	19748.78	896.08	176.38	**119.63**
Hedge funds with at least 30 obs	2072	516.64	**121.16**	**90.56**	**88.27**
1% critical value is 135.8067					
5% critical value is 124.3421					