

# **Diseconomies of Scale in the Actively-Managed Mutual Fund Industry: What Do the Outliers in the Data Tell Us?\***

John Adams, Darren Hayunga, and Sattar Mansi

## **Abstract**

Recent research suggests that improper identification of outliers can lead to distorted inference. We investigate this issue by examining the role that multivariate outliers play in research outcomes using the Chen, Hong, Huang, and Kubik (2004) study. We find that the documented negative relation between scale and return performance in the actively managed mutual fund industry is an artifact of extreme observations. A manual examination of the most influential observations with verifications against outside sources shows that these outliers are largely bad data. Removing the errors reduces the point estimates on the effect of fund size, rendering it economically and statistically insignificant. Further analysis employing regressions that mitigate outlier-induced bias and extending the sample through 2014 confirm our findings. Our evidence contributes to the recent research on the importance of outlier identification in finance research.

*Key Words:* Diseconomies of scale, Mutual fund, liquidity, Influential Observations, Outliers

*JEL Specifications:* G10, G11, G12

May 23, 2018

---

\* Adams is at the University of Texas at Arlington (jcadams@uta.edu), Hayunga is at the University of Georgia (hayunga@uga.edu), and Mansi is at Virginia Tech (smansi@vt.edu). The authors would like to thank Joseph Chen for providing help with his programming code, Ambrus Kesckes, David Reeb, and seminar participants at Virginia Tech and the World Finance Conference in Sardinia for comments and suggestions that improved the paper. The remaining errors are ours.

## 1. Introduction

The potential for multivariate outlier-induced bias in empirical research is largely ignored.<sup>1</sup> Although almost all empirical studies in finance require multivariate analysis, the majority of the work to identify and treat outliers relies on univariate identification (e.g., winsorizing, trimming, or dropping) to solve a multivariate problem.<sup>2</sup> We address this shortcoming by examining the role that multivariate outliers play in research outcomes using one of the most controversial issues facing research on mutual fund performance, namely diseconomies of scale in active management.<sup>3</sup> We use the much-cited Chen, Hong, Huang, and Kubik (2004) as a laboratory for identifying and treating outliers. The choice of the Chen et al. study stems from its significant contribution to the extant literature on the scale-performance relation as well as the availability of their actual data and code (from the American Economic Review website).

Chen et al. (2004) provide evidence that the adverse effect of scale on performance is stronger among illiquid, or small-cap, funds. They argue that increased inflow to funds with illiquid holdings increases trading costs and price pressure on stocks and thus lowers fund return; although for large fund families economies of scale in trading commissions and securities' lending fees help improve performance. Subsequent research evidence in this area is mixed. For example, Yan (2008) confirm the findings of Chen et al. using stock transactions data and direct measures of liquidity. More recently, Harvey and Liu (2017) using scale at industry and fund levels and borrowing from the literature on cross-country economic growth find similar results. On the contrary, Pástor, Stambaugh, and Taylor (2015) refute the Chen et al. findings due to the endogeneity of fund size. They argue fund manager skill is an omitted variable that is correlated with fund size. Reuter and Zitzewitz (2015) provide similar results using a regression-discontinuity approach to investigate the endogeneity of skill and scale. Zhu (2017) reexamines the Pástor et al. study using an enhanced empirical strategy by adding an intercept to the first stage regression and find results consistent with the Chen et al. study.

---

<sup>1</sup> Multivariate outliers are observations that do not follow the pattern of the majority of the data (e.g., Rousseeuw and van Zomeren, 1990). In regression analysis, an outlier is influential when it has a large effect on the parameter estimates.

<sup>2</sup> For clarity, we present a simple illustration of the univariate identification problem in the presence of multivariate outliers in Appendix A. Bollinger and Chandra (2005) show that univariate outlier treatments can induce or exacerbate outlier bias. Adams et al. (2017) find that over 80% of published research in the top finance journals covering the period from 2008 to 2012 utilize univariate outlier mitigation techniques.

<sup>3</sup> Despite the fact that mutual fund managers underperform passive benchmarks, mutual fund investors chase performance and flows into and out of mutual funds are strongly related to lagged measures of excess returns (see Chevalier and Ellison 1997; Sirri and Tufano 1998).

The above studies employ different methodologies (e.g., OLS, OLS with fixed effects, regression discontinuity design, and recursive demeaning regressions) and datasets to address the scale-performance relation in active management with different outcomes. A plausible explanation for the inconsistent empirical results is simply bad data.<sup>4</sup> Unfortunately, due to the large sample sizes common in mutual fund research it is difficult to manually examine all the data at the observation level.

We propose a multivariate outlier identification approach to detect potential bad data. At first glance, our approach may seem unnecessary given improvements in data quality over time. Chen et al. (2004) construct their sample for the years 1963 to 1999 using data from the second generation CRSP Mutual Fund Database (available 1995 to 2007). Later work by Berk and van Binsbergen (2015) and Pástor, Stambaugh, and Taylor (2015) utilize refined data using both CRSP and Morningstar Databases. However, their efforts are largely restricted to reconciling errors in fund returns and fund asset sizes. Implicitly, while this approach corrects for errors in the main variables of interest, it does not address errors in other characteristics (e.g., investment objectives). In fact, even if all of the variable information is valid, an observation can still be an outlier. In this case, researchers will want to know whether their findings are robust to the exclusion of these influential observations (e.g., what percentage of the sample is responsible for the results). Also, since outliers can be the result of omitted variables they can provide new insights into the data generation process and help improve modeling.

We begin our analysis by testing whether the Chen et al. (2004) regression estimates are biased using regressions that down weigh the influence of outliers as a diagnostic tool. For confirmation, we follow Dehon, Gassner, and Verardi (2012) and conduct a formal outlier bias test that compares estimates from outlier-robust estimators with OLS. The results from both tests indicate significant multivariate outlier bias in the Chen et al. results. We then identify the multivariate outliers following Rousseeuw and van Zomeren (1990), who note that these outliers can be hard to detect especially when regressions have more than two variables and one cannot rely on visual perception. We measure outlyingness using outlier robust estimates of location and covariance since classical identification methods (e.g., studentized residuals or Cook's D) rely on the sample mean and covariance matrix, which are themselves influenced by outliers. Lastly, we manually check the most influential observations against outside data sources (e.g., fund annual reports) to determine whether they are caused by bad data. This allows us to make decisions on whether to correct, delete, or mitigate these

---

<sup>4</sup> Bad data includes recording and sampling errors.

bad observations. In the case of the Chen et al. where the dataset is extremely large, manually checking every outlier is not feasible.

We use economic theory to guide our mitigation approach. Berk and Green (2004) use a rational expectations equilibrium framework to make an argument that skilled managers can earn high returns but face increasing marginal costs on funds they actively manage. Thus, as the size of an active fund increases subsequent performance declines.<sup>5</sup> Since the hypothesized relation between size and return is a general effect and not an outlier effect (i.e., driven by rarely occurring events or circumstances), outlier bias can be mitigated by either dropping the most extreme outliers or employing outlier robust regressions that place less weight on extreme observations than OLS does. However, this mitigation approach is not appropriate when outliers, as tail risk events, are the most informative observations since removing the largest manifestation of an effect can make it appear insignificant when it is not. For example, when examining the impact of low-probability economic disasters on equity premiums as in Reitz (1988), Barro (2006), and Welch (2016), naively dropping the most influential outliers would lead to incorrect inferences.<sup>6</sup> In these cases, outlier mitigation should be limited to removing or correcting data errors and improving model specification to account for any omitted variables.

We document that Chen et al. (2004) main results are attributable to outliers that comprise less than two percent of the sample observations. A manual examination of the 20 most influential funds with verifications against outside sources shows that the origin of these outliers is largely sampling error (i.e., bad data). In further analysis, we find that about one quarter of the original Chen et al. dataset contains observations beyond their intended sample of domestic diversified equity mutual funds (e.g., utility funds, international funds, and precious metals funds).<sup>7</sup> Dropping the data errors reduces the point estimates on the effect of fund size and fund family size, rendering both economically and statistically insignificant. In further analysis, we employ robust regressions that mitigate outlier-induced bias and confirm the insignificant relation between fund size and fund return.

---

<sup>5</sup> Berk and Green (2004) implies there is no predictability in funds' excess returns. However, Pástor, Stambaugh, and Taylor (2015) argue there is a difference between the subjective distribution of next period's returns perceived by investors and the objective distribution of the econometrician using the full sample. That is, returns are unpredictable in real time but are predictable in historical data.

<sup>6</sup> Another example is Lee and Mykland (2008), where it would be incorrect to mitigate price jumps in daily stock returns since investors care more about trading days with price jumps than typical trading days.

<sup>7</sup> Chen et al. (2004) use a single domestic equity index to compute excess returns for all of the funds in their dataset. However, we find other non-domestic (e.g., gold, international) funds in the sample. Gruber (1996) shows that failing to include indices for types of securities held by mutual funds leads to incorrect inferences regarding performance.

Chen et al. (2004) results for liquidity are also sensitive to outliers. Controlling for the most influential observations and testing the liquidity hypothesis by comparing funds that invest in small market cap stocks to those that can invest in large market cap stocks, we find that fund size and the interaction of an indicator for not small cap stocks and fund size have economically and statistically insignificant coefficient estimates. Our analyses using Chen et al. data show that fund size is not a determinant of returns even after controlling for fund style.

To ensure that our results are not period or sample specific, we extend the sample using updated observations from the latest version of the CRSP mutual fund database through 2014 and rerun the performance models. After controlling for outliers, the results continue to demonstrate a lack of a negative economic or statistical relation between fund size and return. We do find, however, a consistent positive relation between family size and returns but only in the 2001-2014 period. Unlike Bhojraj, Cho, and Yehuda (2012), who find that the positive relation between family size and fund return is weaker after the enactment of Regulation Fair Disclosure in 2000, we do not observe a weakening relation using robust specifications.

Our research contributes to a growing literature on identification and estimation techniques in finance.<sup>8</sup> Though researchers often use univariate mitigation techniques to control for outliers, these methods cannot reliably correct for outlier influence in multivariate analysis. This issue is of importance given that the majority of empirical studies use OLS as the primary statistical inference technique. Our goal is to understand the underlying process that generates the influential observations and what this might mean for inferences.<sup>9</sup> Finally, the purpose of this research is not to diminish the contributions of the Chen et al. (2004) study. In fact, their seminal work has propelled many researchers to examine the then under-studied relation between scale and performance in the actively managed mutual fund industry. We hope that our findings will spur researchers to go the extra step in identifying and learning from multivariate outliers prior to conducting regression analysis.

---

<sup>8</sup> See for example the recent work of Jiang (2017), Adams et al. (2017), Bowen, Frésard, and Taillard (2015), Roberts and Whited (2013), and Angrist and Pischke (2010).

<sup>9</sup> Current published work that incorporates replications highlights the importance of carefully examining and treating outliers. In a study on board composition, for example, Chhaochharia and Grinstein (2009) (CG) find a negative and significant relation between CEO pay and board independence enhancements. Subsequently, Guthrie, Sokolowsky, and Wan (2012) (GSW) replicate CG's work and show the relation is driven by two CEOs out of a sample of 865 firms. Thereafter, a comment by CG and a rejoinder by GSW attempt to confirm their reported results. Altogether, three publications and a tremendous amount of time was spent attempting to validate each author's work.

## 2. Effect of Scale on Performance: Replication of the Chen et al. (2004) Results

### 2.1 Using OLS Estimators

Our primary data source in this research is the original Chen et al. (2004) files located on the American Economic Review website. Chen et al. construct their sample for the years 1963 to 1999 using data from the second generation CRSP Mutual Fund Database (available 1995 to 2007).<sup>10</sup> Historical data in the second generation was manually collected from printed sources.<sup>11</sup> We use their data and code to reproduce the main published findings, which were reported in Table 3 of their paper.

The main variables of interest are fund size, computed as the log of total net assets (LOGTNA), and family fund size, or the log of one plus the cumulative TNA of the other funds in the fund's family (LOGFAMSIZE). The control variables are additional fund characteristics. These are the: turnover ratio (TURNOVER), or the minimum of purchases and sales over average TNA for the calendar year; average fund age (AGE), expense ratio as a fraction of year-end TNA (EXPRATIO); average total load as a percentage of investments (TOTLOAD); fund's TNA in month  $t$  minus the product of the fund's TNA at month  $t - 12$  with the net fund return between months  $t - 12$  and  $t$ , all divided by the fund's TNA at month  $t - 12$  (FLOW); and past fund return (LAGFUNDRET). All variables are lagged one period.

Panels A and B of Table 1 report Fama-MacBeth (1973) regressions for gross returns and net returns (after deducting fees and expenses), respectively. For ease of comparison, we provide the published results in Chen et al. (2004, p. 1286) in Columns 1 through 4 and our replication findings in Columns 5 through 8. As in Chen et al. (2004), we report the slope coefficients and  $t$ -statistics of models where the dependent variable is market-adjusted returns, beta-adjusted returns, 3-factor returns, and 4-factor returns. The results show that the coefficient estimates in the replications match very closely to those found in Chen et al. (2004). However, we find consistent, albeit trivial, differences between our replicated and the published Newey-West standard errors. We attribute these small discrepancies to Chen et al. (2004) manual calculations rather than the use of a statistical software package when computing standard errors.

---

<sup>10</sup> The original (first generation) CRSP Mutual Fund Database contained open-end mutual fund data beginning December 1961 through December 1995 developed by Mark Carhart for his 1995 dissertation.

<sup>11</sup> These include the Fund Scope Monthly Investment Company Magazine, the Investment Dealers Digest Mutual Fund Guide, Investor's Mutual Fund Guide, the United and Babson Mutual Fund Selector, and the Wiesenberger Investment Companies Annual Volumes.

## 2.2 Using Outlier Robust Estimators

The robustness of an estimator is the level of resistance to change that an estimator has to outliers. Since the underlying error distribution is rarely known with certainty, the ideal robust estimation procedures have estimators that are only slightly less efficient than OLS when the errors are normally distributed but are considerably less biased in the presence of outliers in the data. The ideal estimator efficiently provides precise (i.e., unbiased) coefficient estimates. There are many robust estimators including regression quantiles, least median squares, least trimmed squares, S-estimation, and MM-estimation that vary in levels of robustness and efficiency. A regression quantile (e.g., median regression) minimizes a weighted sum of the absolute values of the errors, with positive errors weighted differently than negative errors. Unfortunately, while easy to compute quantile regressions only mitigate bias from outliers in the dependent variable. Least median squares (LMS) replaces the summing of the squared errors in OLS with the median of squared residuals and least trimmed sum squares (LTS) minimizes the sum of squared trimmed residuals. LMS and LTS suffer from low efficiency and irreproducible coefficient estimates.<sup>12</sup> We report results for median, LMS, and LTS regressions in Appendix B.

The S-estimators seek to minimize a measure of residual dispersion that is less sensitive to outliers than variance. While OLS minimizes the variance of the residuals; the smallest  $\sigma$  that satisfies the equality  $\frac{1}{n} \sum_{i=1}^n \left( \frac{e_i}{\sigma_e} \right)^2 = 1$ , S-estimation replaces the square in the variance with an estimator of scale which awards less importance to large residuals. The S-estimators are robust to outliers in both the dependent and independent variables, to outliers comprising up to 50% of a sample, and are more efficient than LMS and LTS. The *MM*-estimators combine the outlier robustness of S-estimation with relatively high efficiency e.g., up to 95% relative to OLS under the Gauss-Markov assumptions. Numerical optimization is required because there are no formulas for the S- and *MM*-estimators (see Yohai, 1987 and Saliban-Barrera and Yohai, 2006). The algorithm for computing the initial S-estimators begins by estimating regression parameters on randomly selected subsets. The intuition for multiple subsets is to obtain at least one subset without outliers and the final S-scale estimate is

---

<sup>12</sup> Each run of the LMS and LTS regressions can yield different coefficient estimates due to subsampling algorithms used in most computations.

from the subset with the smallest scale. After obtaining the S-estimator, *MM*-estimator is computed via an iteratively reweighted algorithm.

Since the primary interest is the effect of fund size on performance for the overwhelming majority of funds that are not outliers, we employ regression methods that control for outliers as a diagnostic tool. Knez and Ready (1997) note that the OLS loss function employed in almost all applications of Fama-Macbeth (1973) regressions is sensitive to extreme observations. They develop robust estimates by replacing the OLS loss function with least trimmed squares regressions. We apply a similar method using S-estimation, which is more efficient than least trimmed squares. Stromberg, Hössjer, and Hawkins (2000) show that least trimmed squares has a relative efficiency of only 7%. Croux, Rousseeuw, and Hössjer (1994) note the relative efficiency of the outlier robust S-estimator is about 30%. We also consider the more efficient *MM*-estimators which can have relative efficiency values comparable to OLS, but we find the frequency and magnitude of the outlier problem in the Chen et al. (2004) sample too severe to use *MM*-estimation (at higher efficiency levels *MM*-estimators are not reliably robust to outlier influence).

Columns 9 through 12 of Table 1 report the results from the outlier robust regressions. They show the two variables of interest; fund size (LOGTNA) and fund family size (LOGFAMSIZE), are statistically insignificant for either gross or net returns (Panels A and B). The remaining estimated coefficients are generally insignificant and similar to those obtained using OLS (Columns 5 through 8) for each of the four return measures. The one predictive determinant of contemporaneous returns is the lagged return (LAGFUNDRET). The slope coefficients of about 0.029 and six standard errors different than zero are similar in size, sign, and significance across the OLS and robust regression specifications.

For confirmation of outlier bias in the Chen et al. (2004) results, we follow Dehon, Gassner, and Verardi (2012), who provide a procedure that compares estimates from outlier-robust estimators with OLS, and conduct a formal outlier bias test.<sup>13</sup> If the test fails to reject the null hypothesis of no outlier effect OLS is the best estimator (i.e., OLS estimates are not biased by outliers). The *p*-values reported at the bottom of Columns 5 through 8 for both the gross and net returns (Panels A and B) reject the null and show that multivariate outliers bias the OLS coefficient estimates in the Chen et al. study.

---

<sup>13</sup> They follow the methodology of Hausman (1978).



To corroborate these results, we also replicate the series of robustness tests used by Chen et al. (2004). Specifically, we examine three unique specifications: (1) all funds including those in the smallest quintiles, which were excluded initially to mitigate any undue influence by the smallest funds, (2) 12-month non-overlapping returns, and (3) two sample sub-periods from 1963 to 1980 and from 1981 to 1999. Since the results of these robust tests confirm our main finding of a lack of support for the scale hypothesis, we report them in Appendix C. The findings show that the Chen et al. (2004) results are driven by outlier funds. In the next section, we determine how often these extreme funds appear in the sample, and how they are different from most other funds.

### 3. The Impact of Outliers on the Chen et al. (2004) Estimates

#### 3.1 Outlier Identification

We first identify the extreme observations to better understand whether the outliers in the Chen et al. (2004) study are due to bad data. We follow Rousseeuw and van Zomeren (1990) and plot the robust standardized residuals on the y-axis as a measure of vertical distance and the robust Mahalanobis distances on the horizontal x-axis.<sup>14</sup> Figure 1 displays the plot of the gross market adjusted return model for January, 1990 using the same specification that produces the results in Table 1. The two horizontal lines mark the  $y$  bounds at  $\pm 2.25$ , which are values from the standard normal distribution that separate the  $\pm 1.25\%$  most remote regions from the central mass of fund observations. It is standard practice to classify points above the upper and below the lower horizontal lines as having large vertical distances. Similarly, observations to the right of the vertical line located at  $(\chi^2_{p,0.975})$ , where  $p$  is the number of parameters in the model, are classified as having large horizontal distances.

Funds with large vertical and small horizontal distances are vertical outliers, which are found in the region labeled “V” in Figure 1. Vertical outliers are fund observations with very large or small returns, but they are not outlying in the size and other independent-variables space. Figure 1 reports several vertical outlier funds. Funds with large vertical and large horizontal distances are horizontal

---

<sup>14</sup> Mahalanobis distance is a measure of the multivariate outlyingness of an observation in terms of the explanatory variables, defined as  $d_i = \sqrt{(X_i - \mu)\Sigma^{-1}(X_i - \mu)'}$ , where  $\mu$  is the multivariate location vector,  $\Sigma$  is the covariance matrix of the explanatory variables, and  $X_i$  is the  $i^{\text{th}}$  row vector of matrix  $X$ , for  $1 \leq i \leq n$ .

outliers (labeled Regions “H” in Figure 1).<sup>15</sup> A horizontal outlier (also known as a bad leverage point) is an observation that is outlying in the independent variable space and located far from the true regression line. Horizontal outliers significantly affect the estimation of both the intercept and slope coefficients.

Figure 1 shows a number of funds in January 1990 that are horizontal outliers.<sup>16</sup> We label several of these outliers to provide some preliminary insights as to their cause. Most of the horizontal outlier funds are either international or specialized sector funds (e.g., gold and other commodities funds) and most have large vertical distances. The presence of funds other than diversified U.S. equity mutual funds in the sample suggest sampling error since Chen et al. (2004) expressly exclude international and specialized sector funds. We examine this issue in more depth later.

The main purpose of Figure 1 is to show that the Chen et al. (2004) sample contains unusual mutual funds that have the potential to significantly bias the OLS coefficient estimates. Note that while vertical outliers can be easily mitigated using a procedure such as median regression, horizontal outliers are more problematic. Figure 1 also proves the effectiveness of our approach to identifying bad data.

### *3.2 Incidence of Outliers*

Once the outlier funds are identified, the next step is to determine how often they occur over each of the 444 months in the original sample. This analysis helps us better understand whether the outliers are uniformly present throughout, randomly distributed, or concentrated in a specific period. Figure 2 shows the percentage of sample funds that are vertical outliers in each month. The number of mutual funds that are multivariate vertical outliers—funds whose returns are much larger or smaller than the regression model predicts—varies between 2% and 8% in most months. However, in several months around the beginning and end of the sample period (years 1963 and 1999) the percentage of funds that are vertical outliers exceeds 8%. Since the Fama-MacBeth (1973) coefficient estimates are simple

---

<sup>15</sup> Although not reported, Rousseeuw and van Zomeren (1990) plots for all of the other months also show large numbers of vertical and horizontal outlier funds

<sup>16</sup> Funds with large horizontal and small vertical distances only marginally affect parameter estimation but they can affect statistical inference by deflating standard error estimates. They are often referred to as ‘good outliers’ in the statistics literature (labeled Region G in Figure 1). The non-outlier funds, which constitute the vast majority of the observations, are located in Region N of Figure 1.

averages of the monthly coefficient estimates, Figure 2 provides a caution against using the outlier sensitive OLS loss function in the Fama-MacBeth (1973) regressions.

Figure 3 shows the monthly mean incidence of multivariate horizontal outliers. Across the 444 months, the magnitudes are in the range of 2% to 10%. However, as with vertical outliers, the percentage of funds that are horizontal outliers is especially great during the earlier 1960s (close to 18%). This may be the result of actual mutual fund economics during those years. The high incidence of outlier funds in the early periods could also be the result of errors in the second generation of the CRSP mutual fund database when historical fund information was hand collected from a variety of printed sources (see footnote 6). The data in these sources were in turn hand collected from the original source data (fund prospectuses and annual reports). The findings in Chen et al. (2004) appear to support the latter explanation as they find a negative relation between size and return performance that was greater in the earlier part of the sample (see their Table 4), where one might expect hand collection errors to be more prevalent.

### *3.3 How Different are the Outliers?*

Next, we investigate how outlier funds vary from typical mutual funds by examining differences in fund characteristics between the two outlier samples (vertical and horizontal outlier funds) and the non-outlier funds. Table 2 reports descriptive statistics of our results. We compute monthly mean and median magnitudes for both the dependent and independent variables using the market-adjusted gross returns specification to identify the multivariate outliers (Column 1 in panel A of Table 1) in Chen et al. (2004).

Table 2 shows that the differences in the returns between the non-outlier sample and the two outlier samples are remarkable. The mean and median monthly market-adjusted, beta-adjusted, 3-factor, and 4-factor returns for vertical and horizontal outlier funds are much larger than the non-outlier or typical mutual funds. In fact, the mean and median monthly returns are negative for typical non-outlier funds and positive for vertical and horizontal outlier funds. The difference in mean returns between vertical outlier and non-outlier funds is about 41 basis points per month. However, the difference in median monthly returns is much larger, about 190 basis points or almost 24% per year. In addition, there are large differences in the mean and median values for all fund types. For example, the mean market-adjusted return for vertical outliers is 0.257% per month while the median market-

adjusted return is 1.863% per month. These magnitudes equate to mean annual returns of 3.08% versus 22.36%, or more than seven times different. The magnitudes indicate that even within the vertical- and horizontal-outlier subsamples there are funds with more extreme returns compared to their cohorts. This finding suggests that some, but not all, of the outlier observations account for the Chen et al. (2004) results.

We also find statistically significant differences in the independent variables. Most importantly, non-outlier funds are significantly larger than outlier funds with the differences especially larger between horizontal and non-outlier funds. In terms of organizational size, vertical outlier funds have the largest mean and median LOGFAMSIZE values and horizontal outlier funds the smallest. The differences in TURNOVER and FLOW between the non-outlier and horizontal subsamples are especially notable. The horizontal funds have a mean turnover rate that is 60% greater than the non-outliers (90.56 vs 56.88). The mean difference in FLOW for the horizontal subsample is 15 times greater than that of the non-outlier subsample (89.86 vs 5.99). That horizontal outlier funds trade assets more often and experience extremely large cash inflows from investors suggests that the diseconomies of scale documented by Chen et al. (2004) can be influenced by funds with particular investment strategies as well as the popularity of those strategies with investors. We investigate this possibility later in the paper.

### *3.4 Removing Outliers from the Chen et al. (2004) Study*

#### *3.4.1 Regressions of Fund Size and Performance without outliers*

A potential criticism in using a robust estimator is that it is difficult to know whether the difference in results is due to outliers or the estimation procedure (since OLS places more weight to extreme observations than outlier-robust estimation). There is also the concern that outlier-robust estimation can be less efficient than OLS, which could result in less precision in the form of larger standard errors. One way to address both concerns is to simply drop the funds we classify as multivariate vertical and horizontal outliers in Table 2.<sup>17</sup> We can then examine the determinants of fund performance in a sample that contains no multivariate outliers using the same OLS based Fama-McBeth (1973) regressions as Chen et al. (2004).

---

<sup>17</sup> Note, in this test we do not simply drop univariate outliers (i.e., dropping observations that appear extreme in terms of a specific characteristic) because dropping univariate outliers can introduce selection bias (Heckman, 1979).

Table 3 reports the results using this mitigation method for gross and net returns. The results continue to exhibit an insignificant relation between fund size and either gross or net returns using the various return measures. Similarly, family size does not affect fund performance. And as in our prior tests, past fund return is the only statistically significant determinant of return and similar in sign, magnitude and significance to Chen et al. (2004) and our models in Columns 1 through 12 in Table 1.

### *3.4.2 Identifications of Outliers*

Based on the above analyses, we next examine whether a subset of the outliers is responsible for the Chen et al. (2004) diseconomies of scale results. We more precisely identify and quantify the funds that are driving their original findings. Using their regression model for the market adjusted gross returns (Column 1 in Table 1), we remove increasing percentages of funds with large vertical and horizontal distances from the sample and note the change in the estimated coefficient on LOGTNA. Our goal is to find the point where the extreme observations begin to affect the slope coefficients and standard errors of fund size.

Table 4 reports the parameter estimates and standard errors of a two-way sort. Recall that Figure 1 illustrates some funds have returns that are much larger (large positive vertical distances) or much smaller (large negative vertical distances) than the model fitted on the typical funds, which make up the overwhelming majority of the sample, would predict. Therefore, we present the results in two panels. In panel A, we remove funds with large positive vertical distances, and in panel B we remove funds with large negative vertical distances. In both panels we contrast these removals against eliminating funds with the most extreme horizontal distances.

Column 1 and Row 1 in Panel A reports the estimated coefficient of LOGTNA is -0.028 (t-stat of 3.15) when no funds are dropped. This is the same specification and result as originally reported in Chen et al. (2004) Table 3 and our replicated results (Column 5 of Table 1). As we move across Row 1 from left to right, funds with the most extreme horizontal distances are dropped in 1% increments, however the size and significance of the LOGTNA coefficients exhibit negligible changes. Moving down to the next row, where funds in the top one percentile in terms of vertical distances are dropped, Panel A shows the size and significance of the LOGTNA coefficients decrease substantially. This suggests that the statistically negative coefficients of LOGTNA are due to a slightly more than 1% of fund-observations. In the next row, where funds with the 2% most extreme vertical distances are

dropped, all of the LOGTNA coefficient estimates are small and statistically insignificant. The same result holds for the 3%, 4%, and 5% vertical distance rows. Across each row in Panel A, removing a greater percentage of horizontal distances does not monotonically reduce the size and significance of the slope coefficients.

In contrast, panel B reports that removing negative vertical outliers increases the size and significance of the LOGTNA coefficients substantially. In summary, the overall conclusion of the results in Table 4 is that between 1% and 2% of funds with the largest positive vertical distances are the source of Chen et al. (2004) diseconomies of scale finding. We will examine these influential funds below.

### *3.5 A Closer Look at the Funds*

#### *3.5.1 Most Influential Funds: Vertical Distances*

Table 5 lists the 20 most influential funds and their actual investment objective according to the second generation CRSP mutual fund database. We manually verify the actual fund investment objectives using fund annual reports and other outside sources. The table provides the fund names in Column 1, the number of months out of 444 the fund is in the top 2% of vertical distances in Column 2, the fund's actual investment objectives in Column 3, and the number of months the fund has the largest vertical distance among all funds in Column 4.

This listing provides insights into why the original published results are incorrect. The influential funds have the largest vertical distances more often than random chance can explain. For example, Van Eck's International Investors Gold Fund is a top 2% vertical distance fund 90 times and has the largest vertical distance 11 times in the sample. The Franklin and U.S. Global Investors Gold mutual funds are in the top 2% 79 and 70 times, respectively, and the Lexington Strategic Investments fund has the largest vertical distance 26 times.

#### *3.5.2 Sampling Errors: Investment Objective Misclassifications*

The fund names and actual investment objectives point to a sampling error in the original results. On page 1280 of their paper, Chen et al. (2004) write, "we restrict our analysis to diversified U.S. equity mutual funds by excluding from our sample bond, international and specialized sector funds."

Although they provide no explicit reasoning for this restriction; presumably Chen et al. (2004) attempt to avoid the necessity of finding additional indices to compute excess returns on the investments held by bond, international, and specialized sector funds. This is an important consideration as Gruber (1996) shows that failing to include indices that cover the types of securities held by mutual funds can lead to incorrect inferences regarding performance (i.e., fund alpha is incorrectly measured). Therefore, it is surprising to find that out of the 20 funds listed in Table 5 only one, the Hartwell Emerging Growth fund, conforms to the Chen et al. (2004) sample restriction.

The remaining funds listed in Table 5 are international and specialized sector funds and 13 are gold and other precious metal commodity funds.<sup>18</sup> For these latter funds, any relation between fund size and future returns is more likely influenced by the volatile and cyclical nature of commodities funds returns and flows than for diseconomies of scale (e.g., greater transaction costs, organizational diseconomies, or the need to add poorer performing securities).

The sampling error problem is not limited to the top 20 most influential funds. Panel B reports the distributions of fund misclassifications using fund names and by manually checking against outside sources for the overall sample (for quintiles 2 through 5 as in Chen et al. (2004)). We find that specialized sector and international funds account for most of the sampling errors, 9.78% and 7.55% of the overall sample, respectively. In total, approximately one quarter of the funds in the original sample are not diversified domestic equity funds. Figure 4 reports the distribution of investment objective misclassifications over the sample period. The incidence of misclassifications is relatively low in the early years and increases in the 1980s and 1990s. This corresponds to the dramatic increase in mutual fund offerings over the period.

The Chen et al. (2004) sample also includes fixed income, income, growth and income, total return, and balanced funds that have substantial investments in interest paying securities including corporate and U.S. government bonds whose returns are not properly captured by their benchmark portfolios. Some of the sampling errors are due to Chen et al. (2004) sample generation process that includes growth and income funds and that also includes a fund if it has a domestic diversified equity fund classification at any point during the sample period (see footnote 5, p. 1280 of the published paper). The latter criterion is problematic as the investment objective codes in the second generation CRSP database are incorrect in some years, a frequent occurrence in our examination.

---

<sup>18</sup> Funds that primarily invest in the securities of companies that mine, process, or deal in gold, platinum, silver and other precious metals.

For example, the CRSP database incorrectly classifies the Franklin Gold fund (which is the second most influential fund in Table 5 and has vertical distances in the top two percent in 79 months) as a domestic growth equity fund in 1985 through 1989 before identifying it, correctly, as specialized gold commodity fund in later years. Thus, the combination of sample generation process and errors in investment objective coding in CRSP leads to Franklin Gold fund's inclusion in Chen et al. (2004) diversified equity mutual fund sample. In another example, the CRSP database incorrectly classifies the Bartlett Capital Fixed Income fund (which has vertical distances in the top two percent in multiple months) as a domestic growth equity fund in 1986 and 1987 before identifying it, correctly, as a bond fund in later years.<sup>19</sup>

Our count of misclassified investment objectives is most likely conservative as some funds are not easily classified. For example, the Perkins Opportunity fund, a top influential fund for several months in 1994, was in the top 1% of all funds tracked in that year and in 1995 was up about 64% which ranked it in second place of all funds. The fund specializes in finding tiny, thinly traded companies in Minnesota. Much of its strong performance is the result of the fund's family owned private management company buying stakes in these thinly traded stocks after the fund made its purchases. Despite its performance, the fund did not attract large cash inflows, perhaps because of the controversy over its private management company allegedly backing up the fund's trades.<sup>20</sup> The CRSP mutual fund database classified Perkins Opportunity as a growth firm but it is not representative of a typical growth fund so including it introduces another type of sampling error. Furthermore, the Perkins Opportunity fund illustrates the advantage of multivariate outlier analysis in identifying potential sampling errors since simple keyword searches would not flag the fund as unusual.

### *3.5.3 Other Misclassifications*

We rerun the market adjusted gross return model (column 1 in Table 1 and all columns in Table 4) using OLS regression after removing the 58,815 misclassifications listed in Table 5. Table 6 presents the results in column 1. The estimated coefficients and t-statistics for the control variables are not reported. After removing the misclassified funds, the estimated coefficient on LOGTNA remains

---

<sup>19</sup> Additional examples of top two percent vertical distance funds that were temporarily misclassified in the original sample include Fidelity Electric Utilities Portfolio fund, Fidelity Energy Service Portfolio fund, Fidelity Industrial Materials Portfolio fund, Lehman Brothers International Equity Portfolio fund, Sentinel Balanced fund, Smith Barney Natural Resources fund, EuroPacific Growth fund, and Nomura Pacific Basin fund.

<sup>20</sup> What the Perkins Fund Doesn't Tell Its Shareholders, by Edward Wyatt, New York Times, November 26, 1995.



negative and statistically significant. In contrast, LOGFAMSIZE is insignificant. This indicates the published finding of a family scale effect on fund performance is an artifact of sampling errors caused by investment objective misclassifications.

We then remove increasing percentages of funds with large vertical distances and note the change in the estimated coefficients on LOGTNA and LOGFAMSIZE to identify any remaining influential observations. Column 2 reports results when dropping 0.10% of the observations in each month with the largest vertical distances (i.e., fund returns). In columns 3, 4, 5, and 6, we repeat the process by dropping 0.50% to 10% of monthly observations. The estimated coefficient is marginally significant in column 2, insignificant in columns 3 and 4, and positively significant in columns 5 and 6. In terms of LOGTNA, the most influential observations are those in the top 0.10% that drive the negatively significant coefficient in column 1. The results indicate no influential observations for LOGFAMSIZE since it remains insignificant for all dropping levels.

Figure 5 shows the distribution of influential observations, mean monthly market adjusted returns, and fund TNA deciles. The most number of influential observations occurs in the third decile (we follow Chen et al (2004) and drop quintile 1) and the least in the tenth decile. In fact, Figure 5 shows a near monotonic decrease in influential observations as fund size increases. However, in terms of mean monthly market adjusted returns there is little difference across the deciles with the exception of the fifth decile. The mean monthly market adjusted returns for the influential observations are around 7% for most deciles, values that are much larger than the 0.25% reported in Table 2. We next consider how the influential observations are distributed across time in Figure 6. The mean TNA deciles for the influential observations are generally between the fourth and sixth deciles with no apparent trend across time. Thus, it does not appear that extremely large or small funds are driving the overall sample estimated diseconomies of scale result. The mean monthly market adjusted returns on the other hand, are volatile and increasing over time. Thus, smaller funds have more observations that are influential, influential fund returns do not vary with size, and influential fund returns are larger in the later years. Overall, the influential observations, the ones that drive the fund level diseconomies of scale finding, do not seem supportive of a general scale effect.

We next manually examine each of the 482 influential observations using fund annual reports, Morningstar, Kiplinger, SEC EDGAR, and other online resources. Column 2 reports that by dropping 0.10% of the sample for each monthly regression, 482 observations or 0.26% of the 184,977 observations are actually removed. The difference in the intended dropping of 0.10% and the actual

removal of 0.26% is because dropping the top 0.10% of observations effectively removes one fund per month in most years (there are more funds in the years after 1996 so these years have two observations per month in the top 0.10%). We categorize each observation as either an additional misclassification, unusual, fraud, or normal. Additional misclassifications are funds that our initial name search criteria did not catch, unusual funds have unique characteristics that may influence potential scale effects, fraud funds are, or have been, under investigation for fraudulent activities, and normal funds are the remaining. We report the category results in Figure 7.

The unusual 44 funds include CRSP data errors such as the inclusion of closed end funds (Omega fund) and duplications, irrevocable trusts like the American Century Giftrust fund that have minimum investment horizon requirements, dual purpose funds with targeted maturity dates (Leverage fund of Boston), funds with abnormal NAV values like the Equity Strategies fund that was organized as a tax paying c-corporation rather than a tax free regulated investment company, and the above-mentioned Perkins Opportunity fund. The 32 fraud observations include the Mates Investment funds (settled in 1972), the Van Wagoner funds (settled in 2004), the Steadman Investment fund (settled in 1981), and the Gintel fund (settled in 2002). All of the fraud categorized observations cover activities during the original sample period but some were not settled until after the sample period ended so it is not clear how the fraud accusations affected flows. The 90 additional misclassifications include the PaineWebber Atlas fund (International), Dreyfus Capital Value fund (International), Texas fund (Debt), the Capstone Medical Research fund (Specialized Sector), and the Rydex URSA fund (Inverse).

Panel B of Table 6 shows the results after removing the unusual, fraud, and additional misclassification categorized influential funds. Removing the unusual and fraud funds has minimal effect on the estimated coefficients of LOGTNA in columns 1 and 2. However, removing the additional misclassifications results in an insignificant LOGTNA coefficient in column 3. Thus, the published empirical finding of significant diseconomies of scale effect is driven by sampling errors in the form of investment objective misclassifications.

## 4. Liquidity, Organization, and Out of Sample Tests

### 4.1 *Liquidity and Organization: A Test of Fund Size, Family Size, and Fund Styles*

The analysis thus far demonstrates that any diseconomies of scale in mutual fund performance findings are the result of a small number of influential fund observations with incorrect investment

objective classifications. It is interesting to note the OLS results after cleaning the data and the robust regressions are similar. For simplicity, and to focus on the relation between scale and returns and not outlier detection, we employ outlier robust regressions in the remaining analyses. We next consider the assertion in the literature that liquidity and organization affect fund performance.

Chen et al. (2004) examine liquidity as a possible explanation for the finding that size erodes fund performance. They compare funds that invest in stocks with small market caps to all other funds. The rationale is that funds that invest in small stocks are more likely to rely on new stock ideas with growth in assets under management while funds that invest in large cap firms can increase their position without liquidity concerns.

Table 7 replicates their results that restricts the sample to funds in size quintiles two through five (Table 6 in the published paper). Panels A and B report gross and net returns, respectively. The models also include a binary variable equal to one if the self-reported fund style is not Small Cap Growth,  $IND_{(not\ SCG)}$ , as well as the interaction of  $LOGTNA$  and  $IND_{(not\ SCG)}$ . Consistent with our prior findings, the estimated coefficient on  $LOGTNA$  is not economically or statistically significant for either the robust gross (Panel A) or net (Panel B) return models. This suggests that fund size is not a determinant of returns even after controlling for fund style. More importantly, the interaction term for all of the robust return models in Panels A and B is insignificant, which is in contrast to Chen et al. (2004). Overall, the results indicate that liquidity is not a determinant in the relation between fund size and performance.

The estimates for  $LOGFAMSIZE$  and the interaction of  $LOGFAMSIZE$  and  $IND_{(not\ SCG)}$  are insignificant in both Chen et al. (2004), our replications (columns 1-4), and our robust regressions. Chen et al. (2004) interpret this finding as evidence that family size is just as important for small cap growth funds, where liquidity is a concern, as it is for other funds. However, given that our robust regressions in Table 1 show no relation between family size and returns, our interpretation of the insignificant interaction term is different. Simply put, family size does not matter in this sample. As in the earlier tables, prior returns ( $LAGFUNDRET$ ) in both the published and robust results are positively associated with future returns. To confirm robustness, in Panels C and D we use the entire sample of funds (all size quintiles) and find similar results.

Beyond the fund and family size findings, the robust results in Table 7 show that expense ratios are negatively related to future mutual fund returns. This is an expected outcome for net fund returns (see Panel B) since net returns are computed as gross returns less expenses so, *ceteris paribus*, higher

expenses lead to lower net returns. This finding is consistent with Amihud and Goyenko (2013), Christoffersen and Musto (2002), and Carhart (1997), who find that expenses represent a large proportion of the return differential among funds.

Table 7 also reports that expense ratios are negatively related to future gross returns in Panel A, which might seem counter-intuitive since gross returns are computed prior to deducting management, administrative, marketing and other fees that makeup reported expenses. However, as Chen et al. (2014) note transactions costs and securities lending fees (and income) are not included in reported fund expense ratios but instead are added to or subtracted from fund returns. The finding of a negative relation between fund expenses and gross returns suggests a positive relation between reported expenses and unreported transactions costs which is consistent with Livingston and O'Neal (2014), who report that fund expense ratios are positively correlated with commissions per trade.

The significant negative relation between returns and expense ratios provides an important clue for interpreting the role of size in mutual fund returns. Chen et al. (2014) report a negative correlation between size and expense ratios. Similarly, Adams, Mansi, and Nishikawa (2010) find that after controlling for other fund characteristics that fund size is negatively related to expenses. Thus, fund size lowers expense ratios and lower expense ratios lead to higher returns. Instead of eroding fund performance, the results in Table 7 suggest possible positive economies of scale in the mutual fund industry through amortization of expenses.

#### *4.2 Out-of-Sample Tests*

The mutual fund industry has changed considerably since the end date of the Chen et al (2004) sample. These changes include enormous growth in assets under management (from around \$7 trillion in 1999 to more than \$18 trillion in 2014), in the number of investment companies (about 8,000 in 1999 to over 9,000 in 2014), in passively managed index and exchange traded funds, and a trend towards industry concentration at the parent level (the largest 10 fund families accounted for 55% of total net assets of mutual funds and ETFs in 2014, up from 32% at the end of 1999). In addition, regulatory changes since 1999, including the Regulation Fair Disclosure (FD) in 2000, Sarbanes-Oxley Act in 2002, and Dodd-Frank Act in 2010 have affected the governance and investment environments

for mutual funds.<sup>21</sup> Given the magnitude of these changes it is an open question whether our findings hold outside of the original sample period.

Specifically, we are interested in determining whether fund and family sizes have had any effect on performance in subsequent years. We conduct tests for the 1963 to 2014, 1981 to 1999, and 2001 to 2014 periods. Our tests not only employ different sample periods but also differently sourced data using the third generation of the CRSP Mutual Fund database. We follow the method used by Chen et al. (2004) to generate the sample with the exception that we use CRSP portfolio identification numbers to identify the largest share class of each fund rather than identifying funds and share classes via a manual review of fund names. The results of the out-of-sample tests are reported in Table 8 for gross and net returns using the same OLS specifications as in Chen et al. (2004) in columns 1 through 4 and robust Fama-MacBeth regressions in columns 5 through 8.

Panel A reports the gross returns and Panel B reports the net returns for the 1963 to 2014 sample period. In both cases we find results similar to Chen et al. (2004) in columns 1 through 4, i.e., LOGTNA is negative and LOGFAMSIZE is positive and both are significant. Similar to our prior findings, the robust results in columns 5 through 8 report no economically or statistically significant relations between size and returns. Panels C and D present the results for the 1981 to 1999 period. We select this period to be comparable to the Chen et al. (2004) sub-sample analysis (see Table 4 in their original paper) and because the third generation CRSP database coverage prior to 1980 is limited. Thus, Panels C and D provide insights into how errors in the second-generation CRSP mutual fund database impact the Chen et al. (2004) results.

Our replications in columns 1 through 4 for the estimated coefficients for LOGTNA generally have the same sign as in the Chen et al. (2004) paper (see our Appendix C Table 1, Panels G and H or Panel D of Table 4 in the published paper) - although there are differences in the magnitudes and statistical significance of the estimates. Most notably, columns 3 and 4 (3-factor and 4-factor returns) report that the estimated coefficients on LOGTNA are economically small and statistically insignificant for both gross (Panel C) and net (Panel D) returns. In contrast, the robust estimated coefficients on LOGTNA (columns 5 through 8) are positive and statistically significant in a few specifications.

---

<sup>21</sup> See the annual Investment Company Fact Books, available at [www.icifactbook.org](http://www.icifactbook.org), for more details on changes in mutual funds' regulatory, operating, and investment environments.

The estimated coefficients for LOGFAMSIZE in the out of sample regressions in these panels are remarkably similar in sign, size, and significance to those reported by Chen et al. (Panels C and D of Table 4 in the original paper). The robust results for LOGFAMSIZE using the third generation CRSP database are economically larger when compared to our previous results using the Chen et al. (2004) sample. However, robustly estimated coefficients for LOGFAMSIZE are mostly statistically insignificant.

We next conduct out of sample tests for the 2001 to 2014 period. This coincides with Bhojraj, Cho, and Yehuda (2012), who examine fund returns following the implementation of Regulation FD in 2000. Similar to our 1963 to 2014 results, the estimated coefficients for LOGTNA are significant for the OLS replications in columns 1 through 4 but insignificant for the robust results in columns 5 through 8.

Alternatively, in contrast to the 1981 to 1999 panels, for the 2001 to 2014 period we find consistently positive and significant relations between LOGFAMSIZE and returns in the robust regressions. Also, in contrast to Bhojraj, Cho, and Yehuda (2012), we do not find that the relation between LOGFAMSIZE and returns weakens post Regulation FD's implementation in 2000 in using either the Chen et al (2004) or robust methods. Using OLS regressions the LOGFAMSIZE parameter estimate has a positive and significant value of 0.010 for 1963 to 2014 (Panel A), 0.013 for 1981 to 1999 (Panel C), and 0.013 for 2001 to 2014 (Panel E). For the robust regressions, the relation between LOGFAMSIZE and returns is insignificant for the 1963-2014 (Panel A) and for 1981 to 1999 (Panel C) periods, but positive (value of 0.009) and statistically significant at the 1% level for the 2001-2014 period (Panel E).

Similar to our prior findings, the out-of-sample analysis in Table 8 fails to robustly identify diseconomies of scale in the returns of mutual funds. The out-of-sample results are also consistent with our in-sample analysis where a small numbers of atypical observations influence the negative relation between fund size and fund return. At the very least, fund size does not matter for the overwhelming majority of mutual funds. The out of sample analysis in Table 8 coupled with the in sample results (see Appendix C Table 1, Panel D) suggests that family size is positively related to fund returns but that the phenomenon is recent (post 2000). Whether the positive economies of family scale is due to Regulation FD (larger families can more afford the higher costs of obtaining valuable investment information), industry consolidation, or other factors is an open question.

## 5. How Widespread is the Outlier Problem in Finance Research?

An interesting question is whether the outlier problem discussed in this paper is an isolated incidence or a common problem in mutual fund scale studies and finance research in general. Research on diseconomies of scale relies on the correct classification of a fund's investment objective (either self-selected by the fund or assigned by a third party such as a CRSP data provider or Morningstar) to construct samples and to evaluate mutual fund performance. Investment objective misclassifications, as we demonstrate in this paper, can lead to outlier concerns. Recent published work that relies on investment objective classifications and therefore may be subject to outlier concerns include Zhu (2018), Berk and van Binsbergen (2015), Pastor, Stambaugh and Taylor (2015), and Cremers, Petajisto, and Zitzewitz (2013).

In general, the majority of the academic work in finance use univariate identification to solve the multivariate outlier problem. Adams et al. (2017) document that of the 1,196 studies published in the top four finance journals over the period 2008 to 2012, only 404 (34%) of the papers that utilize OLS regression mention outliers.<sup>22</sup> Of those 404 studies, 49% winsorize, 16% trim, and 15% drop the data, with the remainder using other identification methods (e.g., quantile regressions, least trimmed squares, least median squares).

To shed light on whether the outlier issue is a widespread problem in research, we have replicated numerous papers from two mainstream areas in finance using our methodology of identifying and treating multivariate outliers. First, in the asset pricing area we have replicated the works of Menkhoff, Sarno, Schmeling and Schrimpf (2011), Wahal and Yavuz (2013), Pastor, Stambaugh, and Taylor (2015), and Zhu (2017). Second, in the corporate finance area we have examined the works of Peterson (2009), Chhaochharia and Grinstein (2009), Golubov, Petmezas, and Travlos (2012), Panousi and Papanikolaou (2012), Guthrie, Sokolowsky, and Wan (2012), Becker and Stromberg (2012), and Greenwood and Hanson (2012). Collectively, in all but two instances we find the results differ from those initially reported and the changes range from moderate to severe in some papers. The only two papers where the results held up completely are those of Menkhoff, Sarno, Schmeling and Schrimpf (2011), and Greenwood and Hanson (2012). While this is a small sample, our replications suggest that empirical researchers should carefully consider identifying and treating multivariate outliers.

---

<sup>22</sup> The top four finance journals in the Adams et al. (2017) study include *Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*, and *Journal of Financial and Quantitative Analysis*.

## 6. Conclusion

Recent research suggests that multivariate rather than univariate identification of outliers leads to proper inference (e.g., Jiang, 2017; Adams et al., 2017; Bollinger and Chandra, 2005). We contribute to this line of research by replicating the work of Chen, Hong, Huang, and Kubik (2004) on diseconomies of scale in the actively managed mutual fund market, where the authors provide their data and code. We focus on the Chen et al. study because of its significant contribution to the work on the relation between fund size and return performance in the mutual fund industry. Using a large sample covering the period from 1963 to 1999, Chen et al. examine whether scale matter for performance and document that lagged fund size (fund family size) is negatively (positively) related to fund performance. They find support for their results using interactions between liquidity and organizational size.

After controlling for multivariate outliers, we find that their results are driven by extreme observations that account for no more than two percent of the sample observations. Our analysis shows the source of these outliers is largely investment objective misclassifications. After removing the data errors we show that (i) fund size is insignificantly related to fund performance, (ii) fund family size is insignificantly related to fund return, and (iii) the results are not attributed to liquidity. We draw similar conclusions based on robust regressions that down weigh the influence of outliers. Out of sample analyses using the most recent version of the CRSP mutual fund database for the 1963 to 2014 period confirms there is not an observable negative relation between fund size and returns. However, for the period from 2001 to 2014 we document positive economies of scales at the fund family level. Possible explanations for larger families outperforming smaller families in the post 2000 era include industry consolidation and regulatory changes including the Regulation Fair Disclosure (FD) in 2000, Sarbanes-Oxley Act in 2002, and Dodd-Frank Act in 2010.

Collectively, our evidence contributes to the recent research on the importance of outlier identification in empirical finance research. Our research also provides confirmation to the recent work that documents an insignificant relation between fund scale and return performance in the actively mutual fund market (e.g., Reuter and Zitzewitz 2015, Pastor et al. 2015).



## References

- Adams, John C., Darren K. Hayunga, Sattar Mansi, David M. Reeb, and Vincenzo Verardi. 2017. "Identifying outliers in finance." *Working Paper*. University of Georgia.
- Adams, John C., Sattar Mansi, and Takeshi Nishikawa. 2010. "Internal governance mechanisms and operational performance: Evidence from index mutual funds." *Review of Financial Studies* 23: 1261–1286.
- Amihud, Yakov and Ruslan Goyenko. 2013. "Mutual fund's R2 as predictor of performance." *Review of Financial Studies* 26 (3): 667–694.
- Andersen, Robert. 2008. *Modern Methods for Robust Regression*. Sage Publishing, No. 152.
- Angrist, Joshua D., and Jörn-Steffen Pischke, 2010. "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics." *Journal of Economic Perspectives* 24(2): 3-30.
- Barro, Robert J., 2006. "Rare disasters and asset markets in the twentieth century." *Quarterly Journal of Economics* 121 (3), 823–866.
- Becker, B. and P. Stromberg, 2012. "Fiduciary duties and equity-debtholder conflicts." *Review of Financial Studies* 25, 1931–1969.
- Berk, Jonathan B., and Richard C. Green. 2004. "Mutual fund flows and performance in rational markets." *Journal of Political Economy* 112(6): 1269–95.
- Berk, Jonathan B. and Jules H. van Binsbergen, 2015. "Measuring skill in the mutual fund industry." *Journal of Financial Economics* 118, 1–20.
- Bhojraj, Sanjeev, Young Jun Cho, and Nir Yehuda. 2012. "Mutual fund family size and mutual fund performance: The role of regulatory changes." *Journal of Accounting Research* 50(3): 647–684.
- Bollinger, Christopher R. and Amitabh Chandra. 2005. "Iatrogenic specification error: A cautionary tale of cleaning data." *Journal of Labor Economics* 23(2): 235–257
- Bowen, Donald, Laurent Frésard, and Jerome Taillard. 2015. "What's your identification strategy? Innovation in corporate finance research." *Management Science*, forthcoming.
- Carhart, Mark. M. 1997. "On persistence in mutual fund performance." *Journal of Finance* 52: 57–82.
- Chen, Joseph, Harrison Hong, Ming Huang, and Jeffrey D. Kubik. 2004. "Does fund size erode mutual fund performance? The role of liquidity and organization." *American Economic Review* 94(5): 1276–1302.
- Chhaochharia, V., Grinstein, Y., 2009. "CEO compensation and board structure." *Journal of Finance* 64, 231–261.

- Christoffersen, Susan E. K. and David Musto. 2002. "Demand curves and the pricing of money management." *Review of Financial Studies* 15 (5): 1499–1524.
- Golubov, A., D. Petmezas, and N. Travlos, 2012. "When it pays to pay your investment banker: New evidence on the role of financial advisors in M&As." *Journal of Finance* 67, 271–311.
- Cremers, M., Petajisto, A., Zitzewitz, 2013. "Should benchmark indices have alpha? Revisiting performance evaluation." *Critical Finance Review* 2, 1–48.
- Greenwood, Robin, and Samuel G. Hanson. 2012. "Share issuance and factor timing." *Journal of Finance* 67 (2): 761–798.
- Dehon, Catherine, Marjorie Gassner, Vincenzo Verardi. 2012. "Extending the Hausman test to check for the presence of outliers." *Advances in Econometrics* 29, 435–453.
- Edelen, Roger M., Richard B. Evans, and Gregory B. Kadlec. 2007. "Scale effects in mutual fund performance: The role of trading costs." *Working Paper*. University of California, Davis.
- Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake. 2012. "Does mutual fund size matter? The relationship between size and performance." *Review of Asset Pricing Studies* 2: 31–55.
- Fama, Eugene F., and James D. MacBeth. 1973. "Risk, return, and equilibrium: Empirical tests." *Journal of Political Economy* 81 (3): 607–636.
- Gruber, Marin J. 1996. "Another puzzle: The growth in actively managed mutual funds." *Journal of Finance* 51 (3): 783–810.
- Guthrie, K., Sokolowsky, J., Wan, K., 2012. "CEO compensation and board structure revisited." *Journal of Finance* 67, 1149–1168.
- Hausman, Jerry A. 1978. "Specification tests in econometrics." *Econometrica* 46, 1251–1271.
- Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica* 47, 153–161.
- Jiang, Wei. 2017. "Have instrumental variables brought us closer to truth?" *Review of Corporate Finance Studies*, Forthcoming.
- Knez, Peter J. and Mark J. Ready, 1997. "On the robustness of size and book-to-market in cross-sectional regressions." *Journal of Finance* 52, 1355–1382.
- Harvey, Campbell R. and Liu, Yan, 2017. "Decreasing returns to scale, fund flows, and performance." *Working Paper*. Duke I&E Research Paper No. 2017-13.
- Lee, Suzanne S. and Mykland, Per A., 2008. "Jumps in financial markets: A new nonparametric test and jump dynamics." *Review of Financial Studies* 21 (6), Pages 2535–2563,

- Livingston, Miles, and Edward S. O'Neal. 1996. "Mutual fund brokerage commissions." *Journal of Financial Research* 19(2): 273–92.
- McLemore, Ping. 2016. "Do mutual funds have decreasing returns to scale? Evidence from fund mergers." *Working Paper*. Federal Reserve Bank of Richmond.
- Menkhoff, Lukas and Sarno, Lucio and Schmeling, Maik and Schrimpf, Andreas. 2011. "Carry trades and global foreign exchange volatility." *Journal of Finance* 67, 168-718.
- Panousi, V. and D. Papnikolaou, 2012. "Investment, idiosyncratic risk, and ownership." *Journal of Finance* 67, 1113–1148.
- Pástor, Luboš, Robert F. Stambaugh, and Lucian A. Taylor, 2015. "Scale and skill in active management." *Journal of Financial Economics* 116: 23–45.
- Petersen, M., 2009. "Estimating standard errors in finance panel data sets: Comparing approaches." *Review of Financial Studies* 22, 435–480.
- Rietz, Thomas A., 1988. "The equity risk premium: A solution." *Journal of Monetary Economics* 22(1), 117-131.
- Reuter, Jonathan, and Eric Zitzewitz. 2015. "How much does size erode mutual fund performance? A regression discontinuity approach." *Working Paper*. Boston College.
- Roberts, Michael and Toni Whited. 2013. *Endogeneity in Empirical Corporate Finance*. In: Constantinides, G., R. Stulz, and M. Harris, Handbook of the Economics of Finance, Vol 2, Part A. Elsevier, Amsterdam, Netherlands, Ch. 7, p. 493-572.
- Rousseeuw, Peter J., and Bert C. van Zomeren. 1990. "Unmasking multivariate outliers and leverage points." *Journal of the American Statistical Association* 85, 633–639.
- Stromberg, Arnold, Ola Hössjer, and Douglas M. Hawkins. 2000. "The least trimmed difference regression estimator and alternatives." *Journal of the American Statistical Association* 95, 853–864.
- Welch, Ivo, 2016. "The (time-varying) importance of disaster risk." *Financial Analysts Journal* 72(5), 14-30.
- Wahal, S. and M. Yavuz, 2013. "Style investing, comovement and return predictability." *Journal of Financial Economics* 107, 136-154.
- Yan, Xuemin. 2008. "Liquidity, investment style, and the relation between fund size and fund performance." *Journal of Financial and Quantitative Analysis* 43(3): 741–768.
- Zhu, Min. 2017. "Information fund size, managerial skill, and investor rationality." *Journal of Financial Economics*. Forthcoming.

TABLE 1  
REGRESSION OF FUND PERFORMANCE ON LAGGED FUND SIZE

*Panel A: Gross fund returns*

	<u>Published</u>				<u>Replicated</u>				<u>Robust</u>			
	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
INTERCEPT	0.056 (0.96)	0.087 (1.63)	0.080 (1.28)	0.019 (0.30)	0.056 (1.01)	0.087 (1.68)	0.080 (1.27)	0.041 (0.63)	-0.036 (0.56)	-0.009 (0.14)	-0.012 (0.15)	-0.037 (0.46)
LOGTNA <sub>i,t-1</sub>	-0.028 (3.02)	-0.028 (3.04)	-0.023 (2.42)	-0.020 (2.15)	-0.028 (3.15)	-0.028 (3.17)	-0.023 (2.41)	-0.021 (2.22)	-0.002 (0.17)	-0.001 (0.08)	0.004 (0.32)	0.005 (0.47)
LOGFAMSIZE <sub>i,t-1</sub>	0.007 (2.26)	0.007 (2.27)	0.007 (2.22)	0.007 (2.21)	0.007 (2.30)	0.007 (2.31)	0.007 (2.27)	0.007 (2.26)	0.004 (0.95)	0.004 (0.86)	0.003 (0.65)	0.003 (0.56)
TURNOVER <sub>i,t-1</sub>	0.000 (0.85)	0.000 (0.86)	0.000 (0.85)	0.000 (0.83)	0.000 (0.90)	0.000 (0.92)	0.000 (0.91)	0.000 (0.90)	-0.000 (1.23)	-0.001 (1.41)	-0.000 (1.31)	-0.000 (1.19)
AGE <sub>i,t-1</sub>	-0.001 (0.62)	-0.001 (0.62)	-0.001 (0.64)	-0.001 (0.63)	-0.001 (0.58)	-0.001 (0.58)	-0.001 (0.60)	-0.001 (0.59)	-0.000 (0.45)	-0.001 (0.54)	-0.000 (0.23)	-0.000 (0.36)
EXPRATIO <sub>i,t-1</sub>	-0.004 (0.11)	-0.004 (0.09)	-0.007 (0.18)	-0.007 (0.18)	-0.004 (0.11)	-0.004 (0.09)	-0.007 (0.17)	-0.007 (0.18)	-0.040 (0.75)	-0.038 (0.70)	-0.021 (0.39)	-0.035 (0.67)
TOTLOAD <sub>i,t-1</sub>	-0.003 (1.19)	0.003 (1.25)	0.003 (1.26)	0.003 (1.29)	0.003 (1.19)	0.003 (1.18)	0.003 (1.20)	0.003 (1.22)	0.003 (1.12)	0.004 (1.19)	0.002 (0.79)	0.003 (0.94)
FLOW <sub>i,t-1</sub>	0.000 (0.50)	0.000 (0.50)	0.000 (0.51)	0.000 (0.49)	0.000 (0.56)	0.000 (0.55)	0.000 (0.56)	0.000 (0.56)	-0.004 (0.59)	-0.000 (0.45)	-0.000 (0.06)	-0.000 (0.43)
LAGFUNDRET <sub>i,t-1</sub>	0.029 (6.00)	0.028 (5.98)	0.028 (6.00)	0.029 (5.99)	0.029 (5.85)	0.028 (5.83)	0.028 (5.84)	0.029 (5.83)	0.030 (5.41)	0.030 (5.49)	0.030 (5.39)	0.030 (5.31)
Outlier Test $p$ -value					0.000	0.000	0.000	0.000				
No. of Months	444	444	444	444	444	444	444	444	444	444	444	444

Panel B: Net fund returns

	Published				Replicated				Robust			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
INTERCEPT	0.026 (0.44)	0.056 (1.05)	0.049 (0.79)	-0.011 (0.18)	0.026 (0.46)	0.056 (1.09)	0.049 (0.79)	0.010 (0.15)	-0.024 (0.35)	-0.023 (0.34)	-0.006 (0.07)	-0.039 (0.50)
LOGTNA <sub>i,t-1</sub>	-0.025 (2.75)	-0.025 (2.76)	-0.020 (2.16)	-0.018 (1.89)	-0.025 (2.87)	-0.025 (2.89)	-0.020 (2.15)	-0.019 (1.96)	0.000 (0.00)	-0.000 (0.01)	0.003 (0.23)	0.005 (0.49)
LOGFAMSIZE <sub>i,t-1</sub>	0.007 (2.33)	0.007 (2.34)	0.007 (2.29)	0.007 (2.28)	0.007 (2.37)	0.007 (2.38)	0.007 (2.33)	0.007 (2.32)	0.003 (0.68)	0.004 (1.36)	0.003 (0.68)	0.003 (0.69)
TURNOVER <sub>i,t-1</sub>	0.000 (0.77)	0.000 (0.78)	0.000 (0.77)	0.000 (0.75)	0.000 (0.82)	0.000 (0.83)	0.000 (0.82)	0.000 (0.81)	-0.000 (1.25)	-0.000 (1.36)	-0.000 (1.05)	-0.000 (1.31)
AGE <sub>i,t-1</sub>	0.000 (0.52)	0.000 (0.53)	-0.001 (0.55)	-0.001 (0.54)	-0.000 (0.49)	-0.000 (0.49)	-0.001 (0.52)	-0.001 (0.51)	-0.001 (0.84)	-0.001 (0.54)	-0.001 (0.49)	-0.000 (0.41)
EXPRATIO <sub>i,t-1</sub>	-0.039 (0.97)	-0.038 (0.95)	-0.041 (1.04)	-0.041 (1.05)	-0.039 (0.92)	-0.038 (0.90)	-0.041 (0.99)	-0.042 (1.01)	-0.099 (1.76)	-0.081 (1.50)	-0.076 (1.41)	-0.085 (1.64)
TOTLOAD <sub>i,t-1</sub>	0.003 (1.21)	0.003 (1.20)	0.003 (1.21)	0.003 (1.25)	0.003 (1.14)	0.003 (1.13)	0.003 (1.15)	0.003 (1.17)	0.003 (0.95)	0.003 (1.08)	0.003 (0.86)	0.002 (0.75)
FLOW <sub>i,t-1</sub>	0.000 (0.47)	0.000 (0.47)	0.000 (0.48)	0.000 (0.46)	0.000 (0.52)	0.000 (0.52)	0.000 (0.53)	0.000 (0.52)	-0.001 (0.87)	-0.000 (0.50)	-0.000 (0.24)	-0.000 (0.75)
LAGFUNDRET <sub>i,t-1</sub>	0.029 (6.03)	0.029 (6.01)	0.029 (6.03)	0.029 (6.02)	0.029 (5.88)	0.029 (5.86)	0.029 (5.87)	0.029 (5.86)	0.031 (5.54)	0.030 (5.45)	0.030 (5.35)	0.029 (5.26)
Outlier Test <i>p</i> -value					0.000	0.000	0.000	0.000				
No. of Months	444	444	444	444	444	444	444	444	444	444	444	444

*Note:* This tables replicates the results of Chen et al. (2004) using Fama-MacBeth (1973) estimates of monthly fund returns on fund characteristics lagged 1 month using OLS. The sample includes funds that fall within fund size quantiles two to five. Columns 1 through 4 provides actual results from Chen el al. Table 3. Using their data and code, columns 5 through 8 provides our replicated results, and columns 9 through 12 report outlier-robust regressions. Panel A reports gross returns before deducting fees and expenses, and Panel B reports net returns after deducting fees and expenses. Returns are adjusted using the market model (Market-Adj), CAPM (Beta-Adj), and Fama-French three and four factor models. LOGTNA and LOGFAMSIZE are the natural logarithms of fund TNA and one plus the family TNA (less the fund), respectively. TURNOVER is the fund turnover, AGE is the number of years since inception, EXPRATIO is the annual fees and expenses scaled by fund TNA, and TOTLOAD is the sum of front-end, rear-end, and deferred sales charges also scaled by fund TNA. FLOW is the percentage of new fund flow in the fund in the past year. LAGFUNDRET is the buy and hold return of the fund over the prior 12 months. The sample is from January 1963 to December 1999. The *t*-statistics in parentheses are adjusted for serial correlation using Newey-West (1987) lags of order three. The table also reports Dehon, Gassner, and Verardi (2012) outlier test *p*-values for each OLS replication.

TABLE 2  
DESCRIPTIVE STATISTICS BY OUTLIER TYPE

	Mean (Median) Values			Mean (Median) Differences	
	Non-Outlier (1)	Vertical (2)	Horizontal (3)	Vertical – Non-Outlier (2) – (1)	Horizontal – Non-Outlier (3) – (1)
Market-Adj <sub>i,t</sub>	-0.155 (-0.097)	0.257 (1.863)	0.277 (0.563)	0.412 <sup>a</sup> (1.960) <sup>a</sup>	0.431 <sup>a</sup> (0.423) <sup>a</sup>
Beta-Adj <sub>i,t</sub>	-0.079 (-0.031)	0.328 (1.990)	0.344 (0.710)	0.406 <sup>a</sup> (2.011) <sup>a</sup>	0.423 <sup>a</sup> (0.741) <sup>a</sup>
3-Factor <sub>i,t</sub>	-0.022 (0.022)	0.386 (1.835)	0.384 (0.657)	0.408 <sup>a</sup> (1.813) <sup>a</sup>	0.406 <sup>a</sup> (0.635) <sup>a</sup>
4-Factor <sub>i,t</sub>	-0.047 (-0.017)	0.356 (1.802)	0.354 (0.598)	0.403 <sup>a</sup> (1.819) <sup>a</sup>	0.401 <sup>a</sup> (0.615) <sup>a</sup>
LOGTNA <sub>i,t-1</sub>	5.009 (4.890)	4.787 (4.718)	4.532 (4.416)	-0.0221 <sup>a</sup> (-0.172) <sup>a</sup>	-0.477 <sup>a</sup> (-0.474) <sup>a</sup>
LOGFAMSIZE <sub>i,t-1</sub>	6.701 (7.141)	7.023 (7.368)	6.420 (7.046)	0.321 <sup>a</sup> (0.227) <sup>a</sup>	-0.200 <sup>a</sup> (-0.095) <sup>a</sup>
TURNOVER <sub>i,t-1</sub>	56.875 (42.800)	65.010 (53.100)	90.526 (50.000)	8.135 <sup>a</sup> (10.300) <sup>a</sup>	33.651 <sup>a</sup> (7.200) <sup>a</sup>
AGE <sub>i,t-1</sub>	14.562 (9.000)	11.527 (8.000)	13.272 (9.000)	-3.035 <sup>a</sup> (-1.000) <sup>a</sup>	-1.290 <sup>a</sup> (0.000) <sup>a</sup>
EXPRATIO <sub>i,t-1</sub>	1.012 (1.000)	1.192 (1.200)	1.328 (1.300)	0.180 <sup>a</sup> (0.200) <sup>a</sup>	0.316 <sup>a</sup> (0.300) <sup>a</sup>
TOTLOAD <sub>i,t-1</sub>	3.424 (3.500)	3.379 (3.750)	3.153 (3.000)	-0.045 (0.250) <sup>c</sup>	-0.271 <sup>a</sup> (-0.500) <sup>a</sup>
FLOW <sub>i,t-1</sub>	5.989 (-1.944)	3.004 (-3.573)	89.856 (7.081)	-2.985 <sup>a</sup> (-1.629) <sup>a</sup>	83.867 <sup>a</sup> (9.025) <sup>a</sup>
LAGFUNDRET <sub>i,t-1</sub>	13.637 (13.727)	12.322 (11.568)	16.238 (12.949)	-1.315 <sup>a</sup> (-2.069) <sup>a</sup>	2.601 <sup>a</sup> (-0.778) <sup>a</sup>

*Note:* This table reports comparative statistics by observation type using the original market-adjusted model in Table 1. The table reports mean (median) values and provides segmentations for each observation type as well as mean (median) differences in the non-outlier and outlier values. The returns (gross) are calculated before deducting fees and expenses. Returns are adjusted using the market model (Market-Adj), CAPM (Beta-Adj), and Fama-French three and four factor models. LOGTNA and LOGFAMSIZE are the natural logarithms of fund TNA and 1 plus the family TNA (less the fund), respectively. TURNOVER is the fund turnover, AGE is the number of years since inception, EXPRATIO is the annual fees and expenses scaled by fund TNA, and TOTLOAD is the sum of front-end, rear-end, and deferred sales charges also scaled by fund TNA. FLOW is the percentage of new fund flow in the fund in the past year. LAGFUNDRET is the buy and hold return of the fund over the prior 12 months. The sample is from January 1963 to December 1999. The notations <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> denotes statistical significance at the 1%, 5%, and 10% levels, respectively.

TABLE 3  
REGRESSIONS OF FUND PERFORMANCE AFTER REMOVING OUTLIERS

	Gross Returns				Net Returns			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
INTERCEPT	-0.050 (0.01)	-0.020 (0.40)	-0.007 (0.09)	-0.044 (0.64)	-0.063 (1.13)	-0.034 (0.64)	-0.019 (0.28)	-0.057 (0.83)
LOGTNA <sub>i,t-1</sub>	-0.005 (0.59)	-0.005 (0.59)	-0.002 (0.24)	-0.001 (0.06)	0.004 (0.48)	-0.004 (0.48)	-0.002 (0.73)	0.000 (0.02)
LOGFAMSIZE <sub>i,t-1</sub>	0.004 (1.28)	0.004 (1.29)	0.004 (1.26)	0.004 (1.25)	0.005 (1.48)	0.005 (1.49)	0.005 (1.45)	0.005 (1.44)
TURNOVER <sub>i,t-1</sub>	-0.000 (0.63)	-0.000 (0.61)	-0.000 (0.61)	-0.000 (0.62)	-0.000 (0.75)	-0.000 (0.74)	-0.000 (0.73)	-0.000 (0.74)
AGE <sub>i,t-1</sub>	-0.000 (0.36)	-0.000 (0.36)	-0.000 (0.39)	-0.000 (0.38)	-0.001 (0.84)	-0.001 (0.40)	-0.000 (0.42)	-0.000 (0.42)
EXPRATIO <sub>i,t-1</sub>	-0.014 (0.29)	-0.013 (0.28)	-0.015 (0.33)	-0.016 (0.33)	-0.099 (1.14)	-0.053 (1.13)	-0.055 (1.19)	-0.056 (1.20)
TOTLOAD <sub>i,t-1</sub>	0.003 (1.24)	0.003 (1.23)	0.003 (1.25)	0.003 (1.27)	0.003 (1.18)	0.003 (1.17)	0.003 (1.19)	0.003 (1.21)
FLOW <sub>i,t-1</sub>	-0.000 (0.33)	-0.000 (0.33)	-0.000 (0.32)	-0.000 (0.33)	-0.000 (0.04)	-0.000 (0.05)	-0.000 (0.03)	-0.000 (0.03)
LAGFUNDRET <sub>i,t-1</sub>	0.029 <sup>a</sup> (5.50)	0.029 <sup>a</sup> (5.57)	0.029 <sup>a</sup> (5.58)	0.029 <sup>a</sup> (5.57)	0.030 <sup>a</sup> (5.62)	0.030 <sup>a</sup> (5.61)	0.030 <sup>a</sup> (5.62)	0.030 <sup>a</sup> (5.61)
No. of Months	444	444	444	444	444	444	444	444

*Note:* This table reports Fama-MacBeth (1973) estimates of monthly fund returns calculated before deducting fees and expenses (gross returns) in columns 1 through 4 and after deducting fees and expenses (net returns) in columns 5 through 8. All variables are constructed using code provided by Chen et al. (2004). The sample is from January 1963 to December 1999. The *t*-statistics in parentheses are adjusted for serial correlation using Newey-West (1987) lags of order three.

TABLE 4  
EFFECT OF EXTREME (HORIZONTAL AND VERTICAL) OUTLIERS

<i>Panel A: Positive Vertical Distances: Gross Returns</i>						
% Most Extreme Positive Vertical Distance Dropped	% Most Extreme Horizontal Distance Dropped					
	0%	1%	2%	3%	4%	5%
0%	-0.028 (3.15)	-0.029 (3.40)	-0.028 (3.34)	-0.027 (3.19)	-0.026 (3.16)	-0.024 (3.03)
1%	-0.017 (1.96)	-0.020 (2.26)	-0.020 (2.32)	-0.017 (2.01)	-0.016 (1.85)	-0.013 (1.61)
2%	-0.008 (0.91)	-0.010 (1.17)	-0.011 (1.27)	-0.008 (0.94)	-0.007 (0.80)	-0.005 (0.57)
3%	-0.001 (0.08)	-0.003 (0.33)	-0.004 (0.47)	-0.002 (0.22)	-0.001 (0.10)	0.001 (0.14)
4%	0.006 (0.65)	0.003 (0.37)	0.002 (0.23)	0.004 (0.43)	0.005 (0.57)	0.007 (0.85)
5%	0.010 (1.15)	0.007 (0.85)	0.006 (0.71)	0.008 (0.93)	0.009 (1.08)	0.011 (1.37)
<i>Panel B: Negative Vertical Distances: Gross Returns</i>						
% Most Extreme Negative Vertical Distance Dropped						
	0%	1%	2%	3%	4%	5%
0%	-0.028 (3.15)	-0.029 (3.40)	-0.028 (3.34)	-0.027 (3.19)	-0.026 (3.16)	-0.025 (3.03)
1%	-0.028 (3.34)	-0.028 (3.41)	-0.027 (3.30)	-0.026 (3.28)	-0.027 (3.34)	-0.027 (3.41)
2%	-0.035 (4.26)	-0.035 (4.33)	-0.033 (4.19)	-0.033 (4.17)	-0.033 (4.26)	-0.033 (4.30)
3%	-0.039 (4.72)	-0.039 (4.80)	-0.038 (4.66)	-0.037 (4.67)	-0.038 (4.77)	-0.038 (4.87)
4%	-0.042 (5.00)	-0.042 (5.10)	-0.040 (4.93)	-0.039 (4.94)	-0.040 (5.05)	-0.041 (5.14)
5%	-0.047 (5.64)	-0.046 (5.66)	-0.044 (5.47)	-0.044 (5.45)	-0.045 (5.57)	-0.045 (5.64)

*Note:* This table reports the estimated slope coefficients and *t*-statistics for LOGTNA after removing a percentage of vertical and horizontal outliers. The model in each case is the same as column 1 in Chen et al.'s primary analysis in Table 3 using Fama-MacBeth (1973) regressions and the market model (Market-Adj). Panel A reports results when percentiles of positive vertical and horizontal distances are dropped. Panel B reports results when percentiles of negative vertical and horizontal outliers are dropped. The models include the covariates of LOGFAMSIZE, TURNOVER, AGE, EXPRATIO, TOTLOAD, FLOW, and LAGFUNDRET. The *t*-statistics in parentheses are adjusted for serial correlation using Newey-West (1987) lags of order three.



TABLE 5  
INVESTMENT OBJECTIVE MISCLASSIFICATIONS

*Panel A: Top 20 Most Extreme of Fund Observations and Their Investment Objectives*

Fund Name	No. of Months Fund in the Top 2%	Actual Investment Objective	No. of Months Fund has Largest Vertical Distance
Van Eck Funds: International Investors Gold/A	90	International	11
Franklin Gold Fund/A	79	Specialized Sector	9
US Global Investors Funds: Gold Shares	70	International	16
Lexington Strategic Investments Fund	56	Specialized Sector	26
Fidelity Select: Precious Metals & Minerals	46	Specialized Sector	1
Century Shares Trust	43	Specialized Sector	13
Fidelity Select: Gold Portfolio	38	Specialized Sector	1
Pilot International Equity Fund/Pilot	37	International	11
Evergreen Precious Metals Holdings Fund/B	34	Specialized Sector	0
US Global Investors Funds: World Gold Fund	34	Specialized Sector	2
INVESCO Sector Funds: Gold	34	Specialized Sector	1
USAA Invest Trust: Gold Fund	33	Specialized Sector	2
Vanguard Specialized Fund: Gold & Prec Metals	33	Specialized Sector	0
Alliance Technology Fund/A	32	Specialized Sector	3
Hartwell Emerging Growth Fund/A	32	Small Cap Growth	2
Lexington Gold Fund	31	Specialized Sector	0
Lexington Silver Fund	30	Specialized Sector	3
AXP Precious Metals Fund/A	30	Specialized Sector	1
Franklin Custodian Funds: Utilities Series/A	30	Specialized Sector	9
Merrill Lynch Pacific Fund/A	29	International	10

*Panel B: Investment Objective Misclassifications*

Actual Investment Objective	Number of Observations	Percentage of Sample
Specialized Sector	23,852	9.78%
International	18,395	7.55%
Debt (including Money Markets)	9,352	3.84%
Index Funds	4,902	2.03%
Other (including Long-Short, Real Estate, Inverse, etc.)	2314	0.95%
All Misclassifications	58,815	24.15%

*Note:* This table documents investments objective misclassifications. Panel A lists the 20 funds that most frequently have vertical distances (large returns in a multivariate sense) in the top 2%. Panel B provides the incidence of the misclassified observations. Data are provided by Chen et al. (2004).

**TABLE 6**  
REGRESSIONS OF FUND PERFORMANCE AFTER REMOVING INVESTMENT OBJECTIVE MISCLASSIFICATIONS  
AND PERCENTILES OF VERTICAL DISTANCES

% OBSERVATIONS DROPPED	0.00%	0.10%	0.5%	1.00%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)
LOGTNA <sub>i,t-1</sub>	-0.020 (2.43)	-0.013 (1.65)	-0.009 (1.19)	-0.006 (0.77)	0.017 (1.95)	0.028 (3.22)
LOGFAMSIZE <sub>i,t-1</sub>	0.003 (1.18)	0.004 (1.31)	0.003 (1.27)	0.004 (1.37)	0.002 (0.64)	0.000 (0.13)
Observations	184,977	184,495	186,793	182,907	175,524	166,325
Observations Removed	0	482	1,184	2,070	9,453	18,652
Actual Removed (%)	0.00%	0.26%	0.64%	1.12%	5.11%	10.08%
No. of Months	444	444	444	444	444	444

OBSERVATIONS DROPPED	UNUSUAL	FRAUD	ADDITIONAL MISCLASSIFICATIONS
	(1)	(2)	(3)
LOGTNA <sub>i,t-1</sub>	-0.020 (2.52)	-0.020 (2.53)	-0.012 (1.45)
LOGFAMSIZE <sub>i,t-1</sub>	0.003 (1.17)	0.004 (1.39)	0.003 (0.87)
Observations	183,190	183,251	180,182
Observations Removed	1,787	1,726	4,795
Removed (%)	0.97%	0.93%	2.59%
No. of Months	444	444	444

*Note:* This table reports the estimated slope coefficients and  $t$ -statistics for LOGTNA and LOGFAMSIZE after removing investment objective misclassifications identified by fund name. Panel A reports results when additional percentages of vertical outliers are dropped from the sample. Panel B reports results from dropping funds that are unusual, funds that are the subject of fraud examinations and enforcement action, and additional misclassified funds. Unusual, fraud, and additional misclassifications are identified via manual examinations at the fund level. The model in each case is the same as column 1 in Chen et al. (2014) primary analysis in Table 3 using Fama-MacBeth (1973) regressions and the market-adjusted model. The models include the covariates of TURNOVER, AGE, EXPRATIO, TOTLOAD, FLOW, and LAGFUNDRET.  $t$ -statistics in parentheses are adjusted for serial correlation using Newey-West (1987) lags of order three.

TABLE 7  
EFFECT OF FUND AND FAMILY SIZE ON PERFORMANCE BY FUND STYLE

*Panel A: Fund size quintiles two to five: Gross Returns*

	<u>Replicated</u>				<u>Robust</u>			
	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
INTERCEPT	0.200 (1.22)	0.252 (1.47)	0.316 (2.35)	0.283 (2.09)	0.039 (0.21)	0.016 (0.09)	0.064 (0.39)	-0.024 (0.15)
IND <sub>(not SCG)</sub>	-0.245 (1.48)	-0.247 (1.49)	-0.255 (1.56)	-0.257 (1.57)	-0.088 (0.46)	-0.024 (0.13)	-0.019 (0.10)	0.032 (0.14)
LOGTNA <sub>i,t-1</sub>	-0.051 (3.01)	-0.052 (3.03)	-0.056 (3.52)	-0.055 (3.45)	-0.023 (0.82)	-0.011 (0.39)	-0.022 (0.77)	-0.005 (0.17)
LOGTNA <sub>i,t1</sub> × IND <sub>(not SCG)</sub>	0.033 (1.99)	0.034 (2.02)	0.036 (2.18)	0.037 (2.20)	0.021 (0.69)	0.008 (0.26)	0.017 (0.57)	0.003 (0.08)
LOGFAMSIZE <sub>i,t-1</sub>	0.007 (0.65)	0.007 (0.65)	0.007 (0.64)	0.007 (0.65)	0.007 (0.44)	0.005 (0.28)	0.010 (0.65)	0.007 (0.41)
LOGFAMSIZE <sub>i,t</sub> × IND <sub>(not SCG)</sub>	0.004 (0.37)	0.004 (0.37)	0.004 (0.38)	0.004 (0.37)	-0.004 (0.23)	-0.001 (0.08)	-0.007 (0.43)	-0.004 (0.23)
TURNOVER <sub>i,t-1</sub>	0.000 (1.72)	0.000 (1.71)	0.000 (1.72)	0.000 (1.72)	0.000 (1.41)	0.000 (1.19)	0.000 (1.26)	0.000 (1.25)
AGE <sub>i,t-1</sub>	-0.008 (0.96)	-0.001 (0.96)	-0.001 (0.97)	-0.001 (0.96)	-0.000 (0.36)	-0.000 (0.41)	-0.000 (0.49)	-0.000 (0.50)
EXPRATIO <sub>i,t-1</sub>	-0.019 (0.43)	-0.018 (0.42)	-0.020 (0.45)	-0.019 (0.45)	-0.073 (2.18)	-0.061 (1.73)	-0.063 (1.82)	-0.064 (1.87)
TOTLOAD <sub>i,t-1</sub>	0.002 (0.55)	0.002 (0.54)	0.001 (0.53)	0.001 (0.53)	0.003 (1.09)	0.003 (1.08)	0.003 (0.93)	0.003 (0.95)
FLOW <sub>i,t-1</sub>	-0.000 (1.09)	-0.000 (1.08)	-0.000 (1.01)	-0.000 (1.03)	0.000 (0.53)	0.000 (0.66)	0.000 (0.59)	0.000 (0.66)
LAGFUNDRET <sub>i,t-1</sub>	0.022 (3.46)	0.022 (3.45)	0.021 (3.43)	0.021 (3.42)	0.025 (3.69)	0.023 (3.54)	0.022 (3.42)	0.022 (3.40)
No. of Months	228	228	228	228	228	228	228	228

Panel B: Fund size quintiles two to five: Net Returns

	<u>Replicated</u>				<u>Robust</u>			
	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
INTERCEPT	0.170 (1.04)	0.222 (1.29)	0.285 (2.12)	0.252 (1.86)	-0.023 (0.13)	-0.041 (0.23)	-0.024 (0.15)	-0.006 (0.04)
IND <sub>(not SCG)</sub>	-0.244 (1.47)	-0.247 (1.48)	-0.255 (1.55)	-0.256 (1.56)	-0.050 (0.27)	0.009 (0.05)	0.050 (0.28)	0.001 (0.01)
LOGTNA <sub>i,t-1</sub>	-0.048 (2.86)	-0.049 (2.89)	-0.054 (3.38)	-0.052 (3.31)	-0.016 (0.56)	-0.010 (0.32)	-0.008 (0.26)	-0.005 (0.18)
LOGTNA <sub>i,t</sub> × IND <sub>(not SCG)</sub>	0.033 (1.97)	0.033 (2.00)	0.036 (2.16)	0.036 (2.18)	0.016 (0.52)	0.008 (0.25)	0.004 (0.12)	0.003 (0.09)
LOGFAMSIZE <sub>i,t-1</sub>	0.007 (0.65)	0.007 (0.65)	0.007 (0.64)	0.007 (0.65)	0.009 (0.55)	0.010 (0.62)	0.013 (0.81)	0.008 (0.49)
LOGFAMSIZE <sub>i,t</sub> × IND <sub>(not SCG)</sub>	0.004 (0.38)	0.004 (0.38)	0.004 (0.64)	0.004 (0.37)	-0.006 (0.34)	-0.006 (0.38)	-0.010 (0.59)	-0.005 (0.49)
TURNOVER <sub>i,t-1</sub>	0.000 (1.67)	0.000 (1.67)	0.000 (1.67)	0.000 (1.67)	0.000 (1.37)	0.000 (1.19)	0.000 (1.26)	0.000 (1.45)
AGE <sub>i,t-1</sub>	-0.001 (0.87)	-0.001 (0.86)	-0.001 (0.87)	-0.001 (0.86)	-0.000 (0.30)	-0.000 (0.33)	-0.000 (0.39)	-0.000 (0.22)
EXPRATIO <sub>i,t-1</sub>	-0.063 (1.45)	-0.063 (1.44)	-0.064 (1.48)	-0.064 (1.47)	-0.112 (3.28)	-0.109 (3.05)	-0.111 (3.13)	-0.122 (3.60)
TOTLOAD <sub>i,t-1</sub>	0.001 (0.48)	0.001 (0.48)	0.001 (0.46)	0.001 (0.46)	0.002 (0.78)	0.003 (1.01)	0.003 (0.91)	0.003 (0.94)
FLOW <sub>i,t-1</sub>	-0.000 (1.17)	-0.000 (1.16)	-0.000 (1.09)	-0.000 (1.10)	0.000 (0.82)	0.000 (0.89)	0.000 (0.63)	0.000 (0.68)
LAGFUNDRET <sub>i,t-1</sub>	0.022 (3.48)	0.022 (3.48)	0.021 (3.45)	0.022 (3.45)	0.024 (3.58)	0.023 (3.49)	0.023 (3.45)	0.023 (3.52)
No. of Months	228	228	228	228	228	228	228	228

Panel C: Fund size quintiles one to five: Gross Returns

	<u>Replicated</u>				<u>Robust</u>			
	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
INTERCEPT	0.274 (1.72)	0.338 (1.99)	0.385 (2.93)	0.358 (2.73)	0.079 (0.39)	0.084 (0.40)	0.119 (0.59)	0.072 (0.38)
IND <sub>(not SCG)</sub>	-0.353 (2.20)	-0.352 (2.20)	-0.361 (2.29)	-0.356 (2.26)	-0.141 (0.69)	-0.093 (0.46)	-0.089 (0.43)	-0.064 (0.32)
LOGTNA <sub>i,t-1</sub>	-0.067 (3.60)	-0.070 (3.76)	-0.071 (3.93)	-0.071 (3.91)	-0.030 (0.84)	-0.029 (0.83)	-0.034 (0.92)	-0.026 (0.73)
LOGTNA <sub>i,t</sub> × IND <sub>(not SCG)</sub>	0.065 (3.50)	0.065 (3.50)	0.067 (3.61)	0.066 (3.57)	0.030 (0.81)	0.027 (0.76)	0.033 (0.87)	0.024 (0.68)
LOGFAMSIZE <sub>i,t-1</sub>	0.012 (1.06)	0.012 (1.05)	0.012 (1.04)	0.012 (1.04)	0.001 (0.06)	0.006 (0.37)	0.012 (0.71)	0.010 (0.58)
LOGFAMSIZE <sub>i,t</sub> × IND <sub>(not SCG)</sub>	-0.003 (0.22)	-0.002 (0.20)	-0.002 (0.21)	-0.002 (0.20)	0.002 (0.10)	-0.004 (0.22)	-0.009 (0.52)	-0.007 (0.40)
TURNOVER <sub>i,t-1</sub>	0.000 (1.37)	0.000 (1.35)	0.000 (1.37)	0.000 (1.34)	0.000 (0.10)	0.000 (0.16)	0.000 (0.13)	0.000 (0.04)
AGE <sub>i,t-1</sub>	-0.002 (1.88)	-0.002 (1.82)	-0.002 (1.90)	-0.002 (1.86)	-0.001 (0.85)	-0.000 (0.75)	-0.001 (0.83)	-0.000 (0.77)
EXPRATIO <sub>i,t-1</sub>	-0.031 (1.16)	-0.030 (1.12)	-0.031 (1.15)	-0.030 (1.12)	-0.033 (1.44)	-0.038 (1.65)	-0.040 (1.78)	-0.041 (1.82)
TOTLOAD <sub>i,t-1</sub>	0.004 (1.48)	0.004 (1.44)	0.004 (1.45)	0.004 (1.43)	0.005 (1.74)	0.005 (1.79)	0.005 (1.60)	0.004 (1.56)
FLOW <sub>i,t-1</sub>	-0.000 (1.61)	-0.000 (1.59)	-0.000 (1.59)	-0.000 (1.59)	0.000 (0.04)	0.000 (0.21)	0.000 (0.20)	0.000 (0.33)
LAGFUNDRET <sub>i,t-1</sub>	0.022 (3.58)	0.022 (3.58)	0.022 (3.57)	0.022 (3.56)	0.024 (3.64)	0.024 (3.64)	0.024 (3.68)	0.024 (3.63)
No. of Months	228	228	228	228	228	228	228	228

Panel D: Fund size quintiles one to five: Net Returns

	<u>Replicated</u>				<u>Robust</u>			
	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
INTERCEPT	0.235 (1.48)	0.300 (1.76)	0.345 (2.64)	0.319 (2.43)	0.060 (0.31)	0.104 (0.48)	0.117 (0.61)	0.011 (0.05)
IND <sub>(not SCG)</sub>	-0.354 (2.22)	-0.353 (2.21)	-0.363 (2.30)	-0.358 (2.27)	-0.155 (0.78)	-0.148 (0.72)	-0.123 (0.62)	-0.039 (0.19)
LOGTNA <sub>i,t-1</sub>	-0.064 (3.45)	-0.067 (3.61)	-0.068 (3.78)	-0.068 (3.76)	-0.034 (0.95)	-0.037 (0.99)	-0.034 (1.00)	-0.014 (0.38)
LOGTNA <sub>i,t</sub> × IND <sub>(not SCG)</sub>	0.065 (3.51)	0.065 (3.51)	0.067 (3.62)	0.066 (3.58)	0.038 (1.04)	0.039 (1.02)	0.036 (1.04)	0.016 (0.42)
LOGFAMSIZE <sub>i,t-1</sub>	0.012 (1.07)	0.012 (1.07)	0.012 (1.06)	0.012 (1.06)	0.007 (0.40)	0.005 (0.31)	0.009 (0.52)	0.009 (0.50)
LOGFAMSIZE <sub>i,t</sub> × IND <sub>(not SCG)</sub>	-0.002 (0.21)	-0.002 (0.19)	-0.002 (0.20)	-0.002 (0.20)	-0.005 (0.26)	-0.003 (0.17)	-0.006 (0.34)	-0.006 (0.32)
TURNOVER <sub>i,t-1</sub>	0.000 (1.29)	0.000 (1.27)	0.000 (1.29)	0.000 (1.26)	-0.000 (0.05)	0.000 (0.12)	0.000 (0.12)	0.000 (0.10)
AGE <sub>i,t-1</sub>	-0.002 (1.80)	-0.002 (1.74)	-0.002 (1.82)	-0.002 (1.78)	-0.001 (0.87)	-0.000 (0.75)	-0.001 (0.82)	-0.000 (0.67)
EXPRATIO <sub>i,t-1</sub>	-0.072 (2.62)	-0.070 (2.59)	-0.072 (2.61)	-0.071 (2.58)	-0.077 (3.17)	-0.076 (3.17)	-0.080 (3.37)	-0.080 (3.39)
TOTLOAD <sub>i,t-1</sub>	0.004 (1.47)	0.004 (1.42)	0.004 (1.43)	0.004 (1.41)	0.005 (1.67)	0.005 (1.63)	0.004 (1.44)	0.004 (1.37)
FLOW <sub>i,t-1</sub>	-0.000 (1.70)	-0.000 (1.68)	-0.000 (1.67)	-0.000 (1.67)	0.000 (0.29)	0.000 (0.96)	0.000 (0.01)	0.000 (0.05)
LAGFUNDRET <sub>i,t-1</sub>	0.023 (3.60)	0.022 (3.60)	0.022 (3.60)	0.022 (3.58)	0.025 (3.65)	0.024 (3.64)	0.024 (3.68)	0.024 (3.65)
No. of Months	228	228	228	228	228	228	228	228

*Note:* The table reports the Chen et al. and robust results for fund and family size measures along with new variables that interact fund and family size with a binary variable, IND<sub>(not SCG)</sub>, that equals one if the self-reported fund style is Not Small Cap Growth (zero otherwise). The regression specifications are the same as in Chen et al.'s Table 6. Chen et al.'s Table 6 only reports estimated coefficients on style and size variables whereas this version of Table 6 reports the estimated coefficients for all covariates. The sample is funds in size quintiles two through five using gross and returns in panels A and B and funds in size quintiles one through five using gross and net returns in panels C and D. The *t*-statistics in parentheses are adjusted for serial correlation using Newey-West (1987) lags of order three.

TABLE 8  
OUT OF SAMPLE REGRESSIONS OF FUND RETURNS

<i>Panel A: Sample period is 1963-2014: Gross Returns</i>								
	OLS				Robust Regression			
	Market-Adj	Beta-Adj	3-Factor	4-Factor	Market-Adj	Beta-Adj	3-Factor	4-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.037 (4.58)	-0.036 (4.43)	-0.027 (2.89)	-0.021 (2.26)	-0.000 (0.05)	-0.001 (0.11)	0.006 (0.42)	0.003 (0.23)
LOGFAMSIZE <sub>i,t-1</sub>	0.010 (3.71)	0.010 (3.70)	0.010 (3.74)	0.010 (3.73)	0.004 (1.04)	0.002 (0.50)	0.001 (0.24)	0.002 (0.57)
No. of Months	612	612	612	612	612	612	612	612
<i>Panel B: Sample period is 1963-2014: Net Returns</i>								
	OLS				Robust Regression			
	Market-Adj	Beta-Adj	3-Factor	4-Factor	Market-Adj	Beta-Adj	3-Factor	4-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.037 (4.48)	-0.035 (4.34)	-0.027 (2.78)	-0.020 (2.15)	-0.002 (0.17)	0.001 (0.07)	0.006 (0.43)	0.005 (0.39)
LOGFAMSIZE <sub>i,t-1</sub>	0.010 (3.71)	0.010 (3.69)	0.010 (3.73)	0.010 (3.73)	0.003 (0.75)	0.003 (0.69)	0.003 (0.78)	0.002 (0.60)
No. of Months	612	612	612	612	612	612	612	612
<i>Panel C: Sample period is 1981-1999: Gross Returns</i>								
	OLS				Robust Regression			
	Market-Adj	Beta-Adj	3-Factor	4-Factor	Market-Adj	Beta-Adj	3-Factor	4-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.028 (2.03)	-0.026 (1.87)	-0.011 (0.68)	-0.000 (0.01)	0.009 (0.65)	0.012 (0.83)	0.035 (1.37)	0.045 (1.83)
LOGFAMSIZE <sub>i,t-1</sub>	0.013 (2.58)	0.013 (2.58)	0.013 (2.60)	0.013 (2.58)	0.010 (1.62)	0.011 (1.91)	0.009 (1.46)	0.009 (1.40)
No. of Months	228	228	228	228	228	228	228	228

Panel D: Sample period is 1981-1999: Net Returns

	OLS				Robust Regression			
	Market-Adj	Beta-Adj	3-Factor	4-Factor	Market-Adj	Beta-Adj	3-Factor	4-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.028 (2.03)	-0.026 (1.88)	-0.010 (0.65)	0.000 (0.01)	0.012 (0.80)	0.012 (0.79)	0.043 (1.66)	0.043 (1.74)
LOGFAMSIZ E <sub>i,t-1</sub>	0.013 (2.58)	0.013 (2.58)	0.013 (2.59)	0.013 (2.57)	0.010 (1.62)	0.009 (1.45)	0.008 (1.39)	0.010 (1.66)
No. of Months	228	228	228	228	228	228	228	228

Panel E: Sample period is 2001-2014: Gross Returns

	OLS				Robust Regression			
	Market-Adj	Beta-Adj	3-Factor	4-Factor	Market-Adj	Beta-Adj	3-Factor	4-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.031 (3.83)	-0.030 (3.68)	-0.021 (2.54)	-0.020 (2.66)	-0.006 (0.90)	-0.006 (0.74)	0.005 (0.51)	0.005 (0.60)
LOGFAMSIZ E <sub>i,t-1</sub>	0.013 (3.65)	0.013 (3.65)	0.013 (3.61)	0.013 (3.61)	0.009 (3.19)	0.009 (3.18)	0.008 (3.03)	0.009 (3.07)
No. of Months	168	168	168	168	168	168	168	168

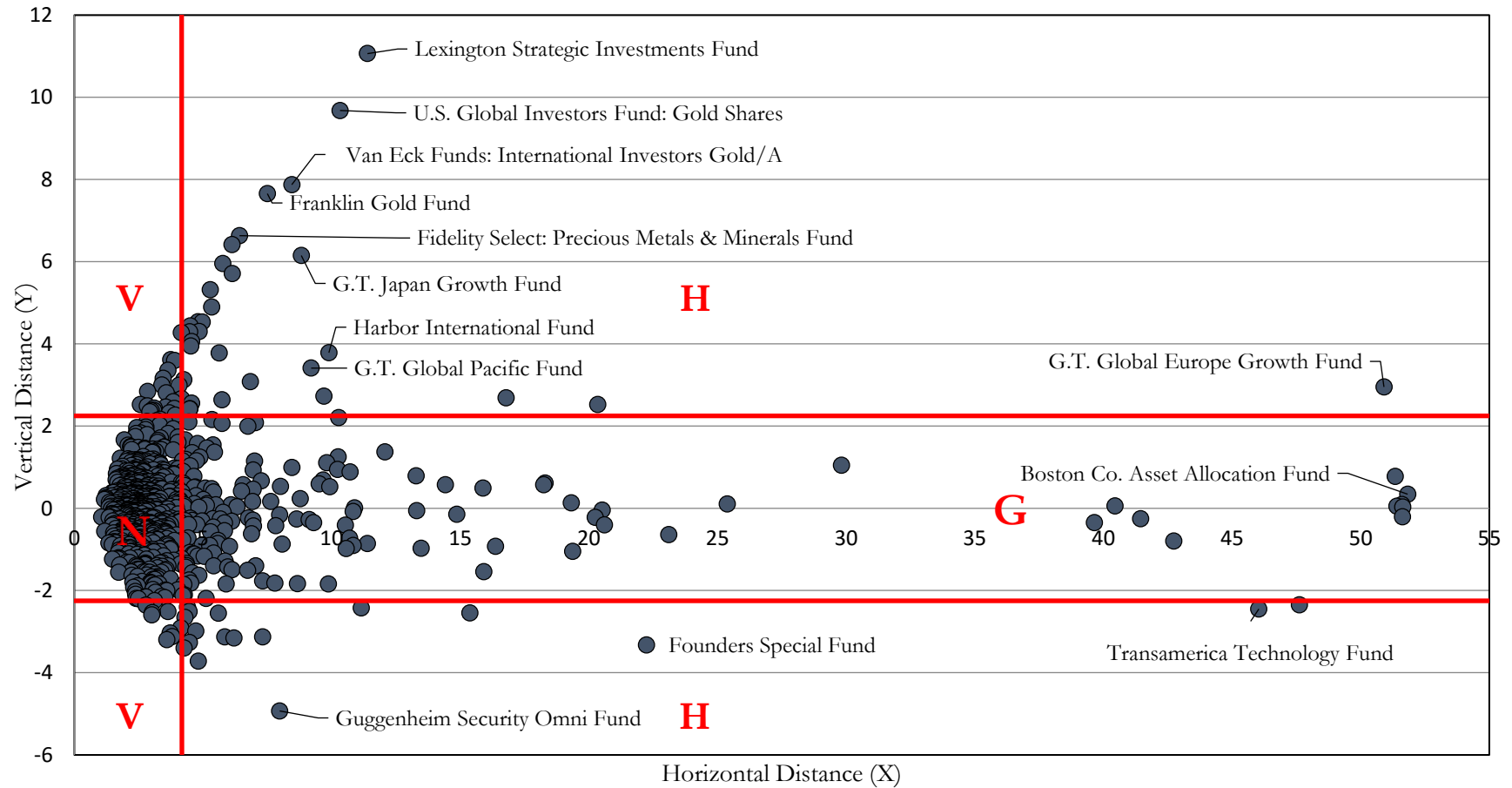
Panel F: Sample period is 2001-2014: Net Returns

	OLS				Robust Regression			
	Market-Adj	Beta-Adj	3-Factor	4-Factor	Market-Adj	Beta-Adj	3-Factor	4-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.031 (3.85)	-0.031 (3.69)	-0.021 (2.56)	-0.020 (2.69)	-0.006 (0.96)	-0.006 (0.90)	0.004 (0.41)	0.004 (0.49)
LOGFAMSIZ E <sub>i,t-1</sub>	0.013 (3.68)	0.013 (3.68)	0.013 (3.64)	0.013 (3.63)	0.009 (3.23)	0.009 (3.27)	0.009 (3.08)	0.009 (3.08)
No. of Months	168	168	168	168	168	168	168	168

*Note:* The sample period in Chen et al. ends at 1999 and employs the second generation CRSP mutual fund database. The table reports updated results for the sample period from 1963 to 2014. The results use the Chen et al. methods with fund data from the third generation CRSP Mutual Fund Database. As before, the table presents Fama-MacBeth (1973) OLS estimates in columns 1–4 and outlier-robust estimates in columns 5–8. Panels A, C, and E report gross returns whereas panels B, D, and F present net returns. The remaining panels present results for various subsamples. Additional covariates are the same as in previous models: TURNOVER, AGE, EXPRATIO, TOTLOAD, FLOW, and LAGFUNDRET. The *t*-statistics in parentheses are adjusted for serial correlation using Newey-West (1987) lags of order three.

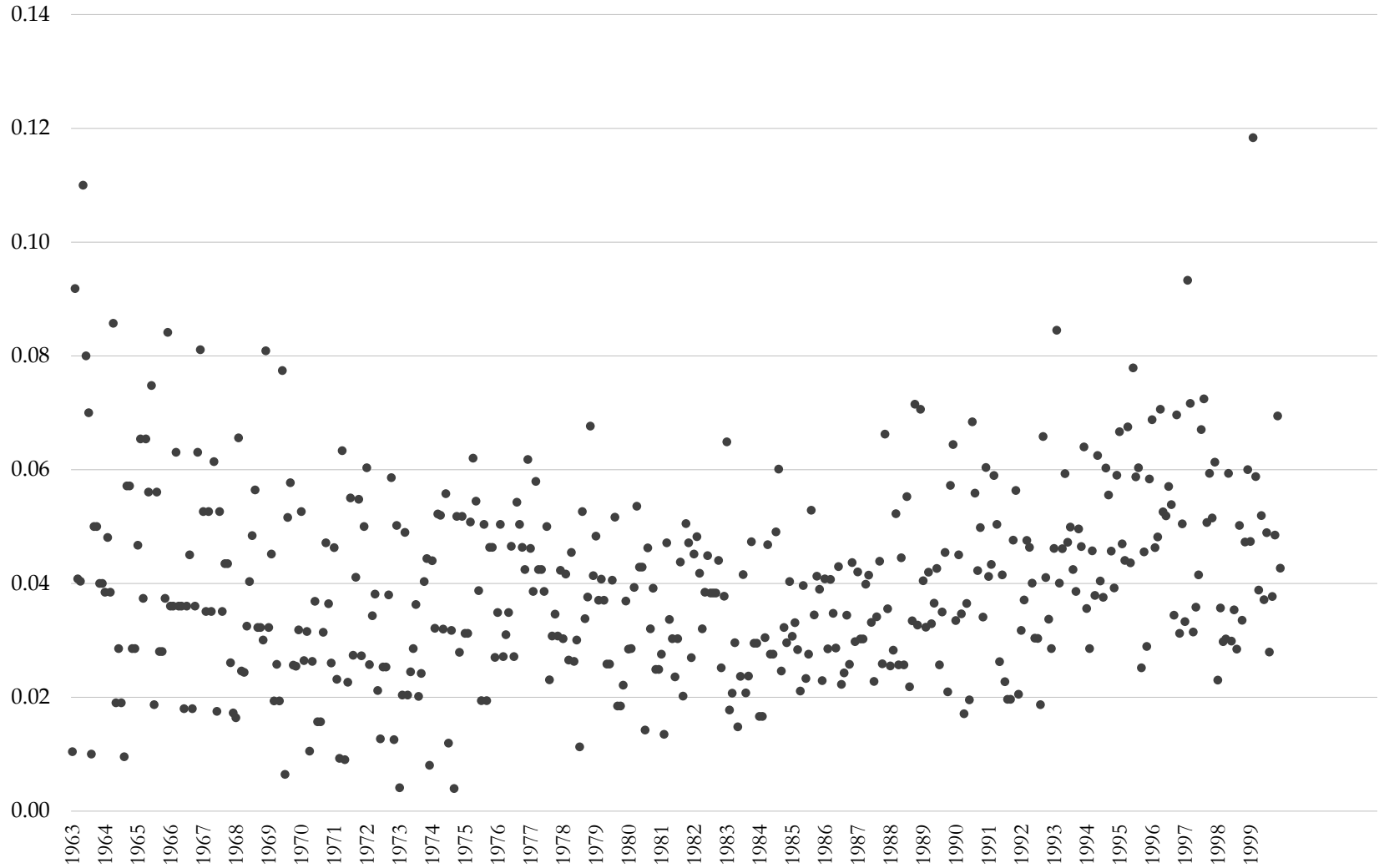


FIGURE 1  
VERTICAL AND HORIZONTAL DISTANCES: JANUARY 1990



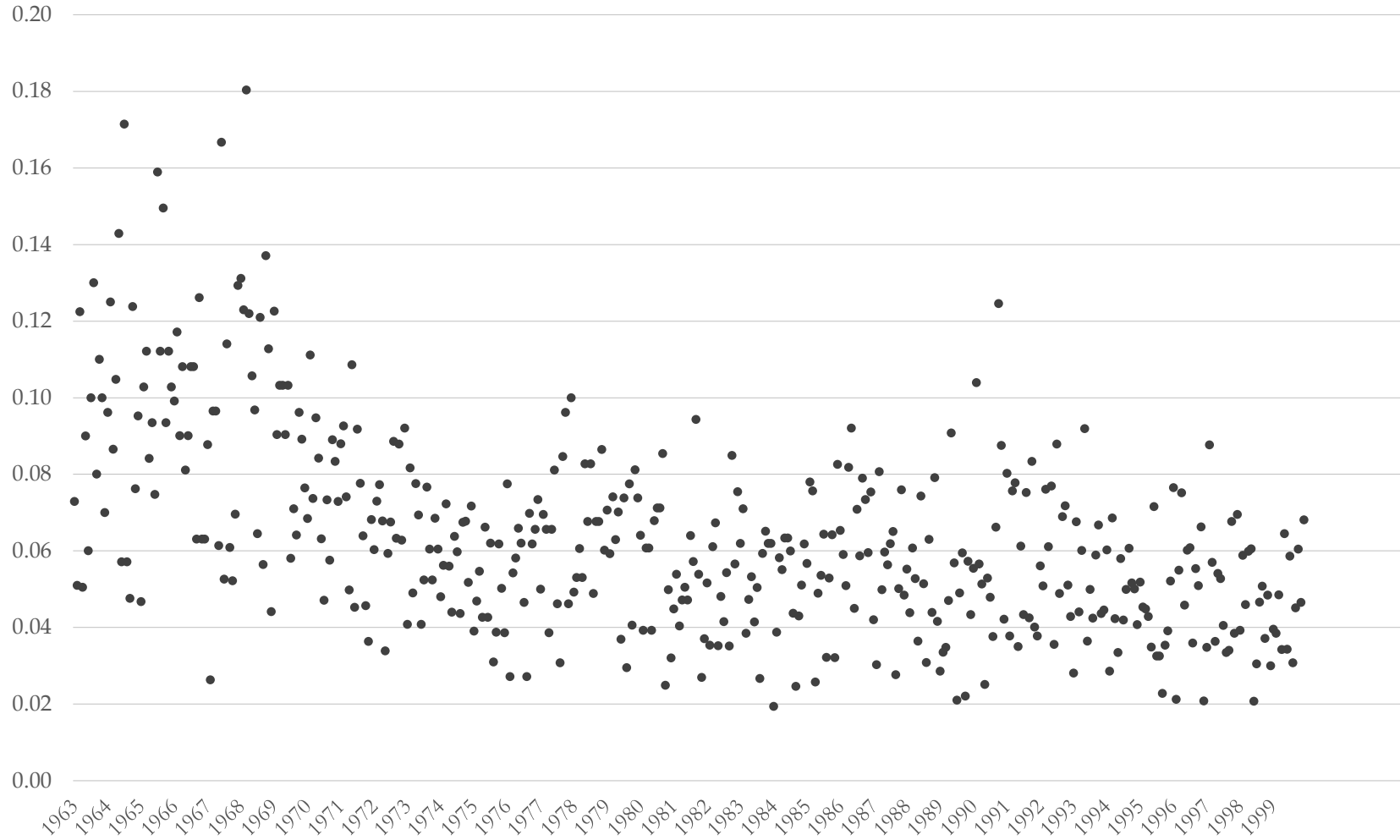
*Note:* Outlier detection plot of Chen et al. (2004) Model 1 in Table 1 for the month of January, 1990. Vertical distance are robust standardized residuals measuring each observation's outlyingness in the (Y) dependent variable. Observations with vertical distances outside the region identified by the two horizontal boundaries located at  $\pm 2.25$  (values from the standard normal distribution that separate the  $\pm 1.25\%$  most remote regions from the central mass of observations) are vertical outliers (labeled "V"). Observations with horizontal distances to the right of the vertical boundary located at  $(\chi^2_{p,0.975})$ , where  $p$  is the number of parameters in the model and outside the horizontal bands are horizontal outliers (labeled "H"). Observations with horizontal distances to the right of the vertical boundary and inside the horizontal bands are good outliers (labeled "G"). Non-outliers are in the region labeled N.

FIGURE 2  
MONTHLY PERCENTAGE OF VERTICAL OUTLIERS



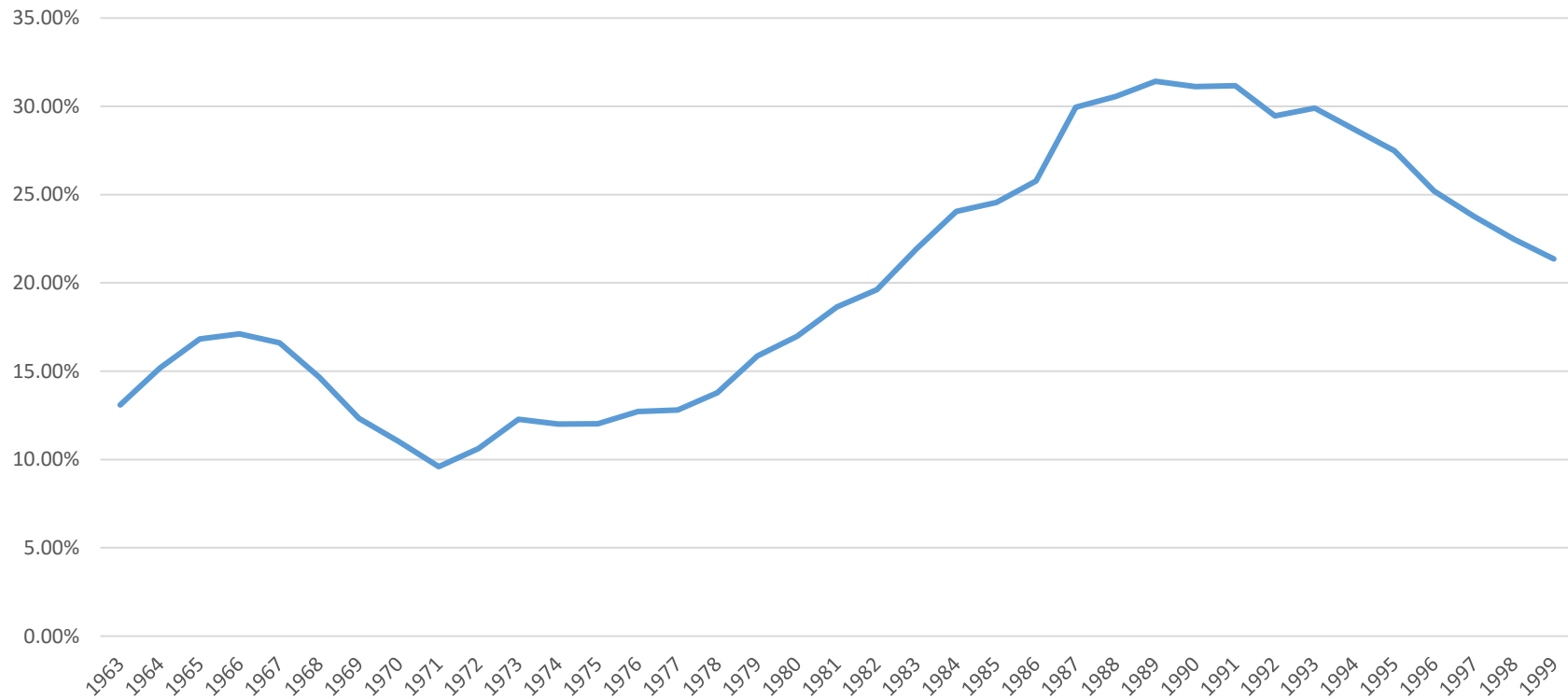
*Note:* This figure shows the monthly average percentage of vertical outliers in the Chen et al. (2004) data for Model 1 in Table 1 over the 1932-1999 sample period. Vertical outliers are outlying in the dependent variable dimension.

FIGURE 3  
MONTHLY PERCENTAGE OF HORIZONTAL OUTLIERS



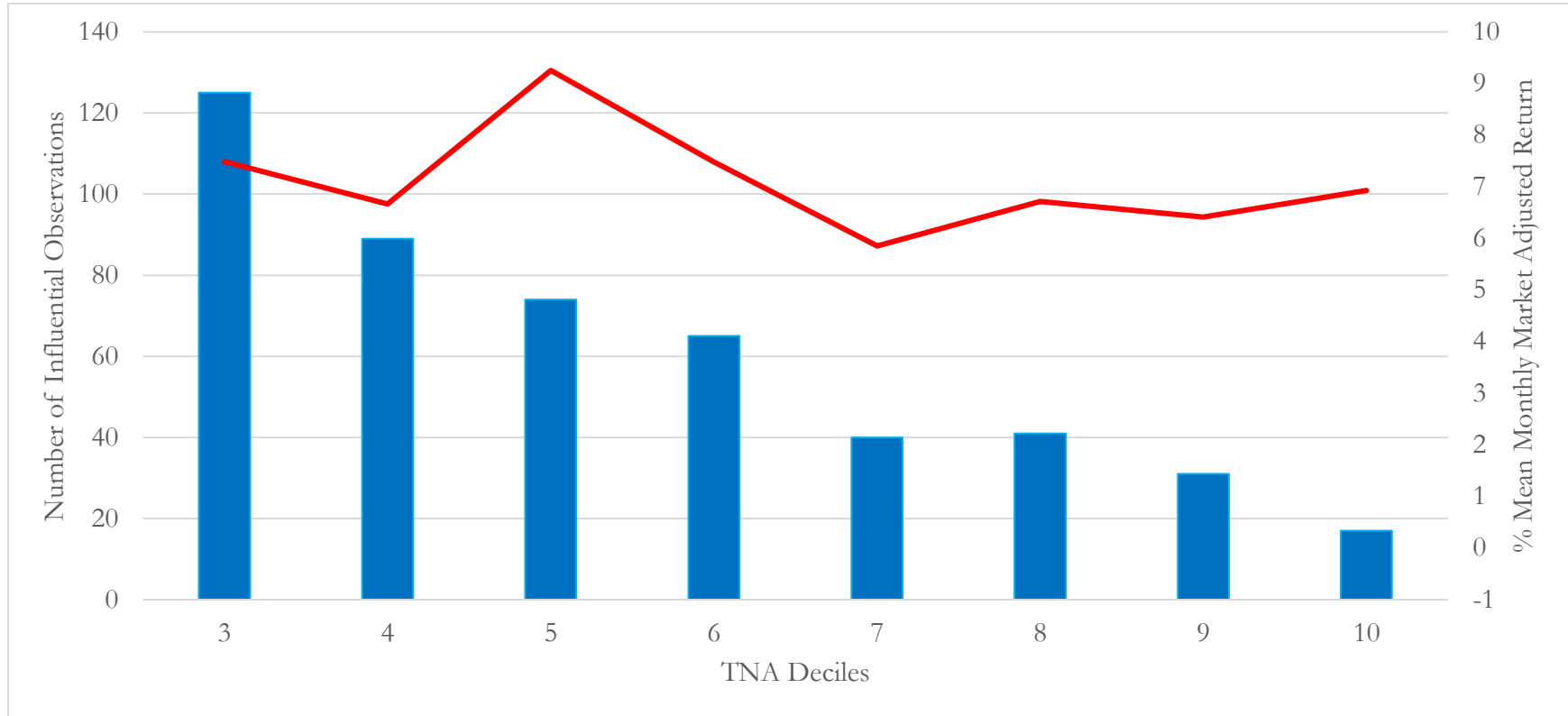
*Note:* This figure shows the monthly average percentage of horizontal outliers in the Chen et al. (2004) data for Model 1 in Table 1 over the 1963-1999 sample period. Horizontal outliers are outlying in the independent variable space.

FIGURE 4  
ANNUAL PERCENTAGE OF INVESTMENT OBJECTIVE MISCLASSIFICATIONS



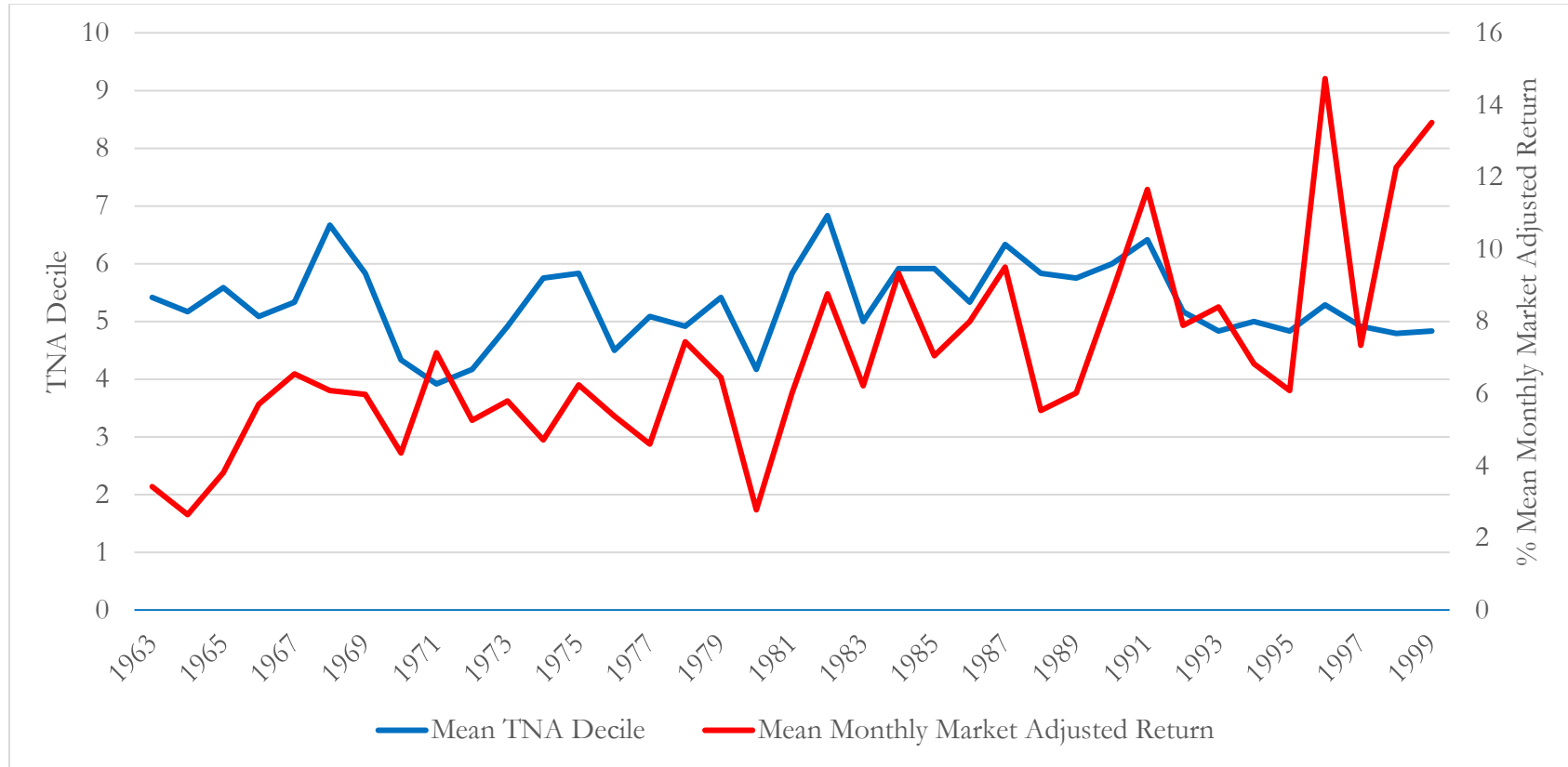
*Note:* This figure shows the annual average percentage of sampling errors caused by investment objective misclassifications in the Chen et al. (2004) data for Model 1 in Table 1 over the 1963-1999 sample period. This figure only includes verifiable investment objective misclassifications.

FIGURE 5  
INFLUENTIAL OBSERVATIONS BY TNA DECILE



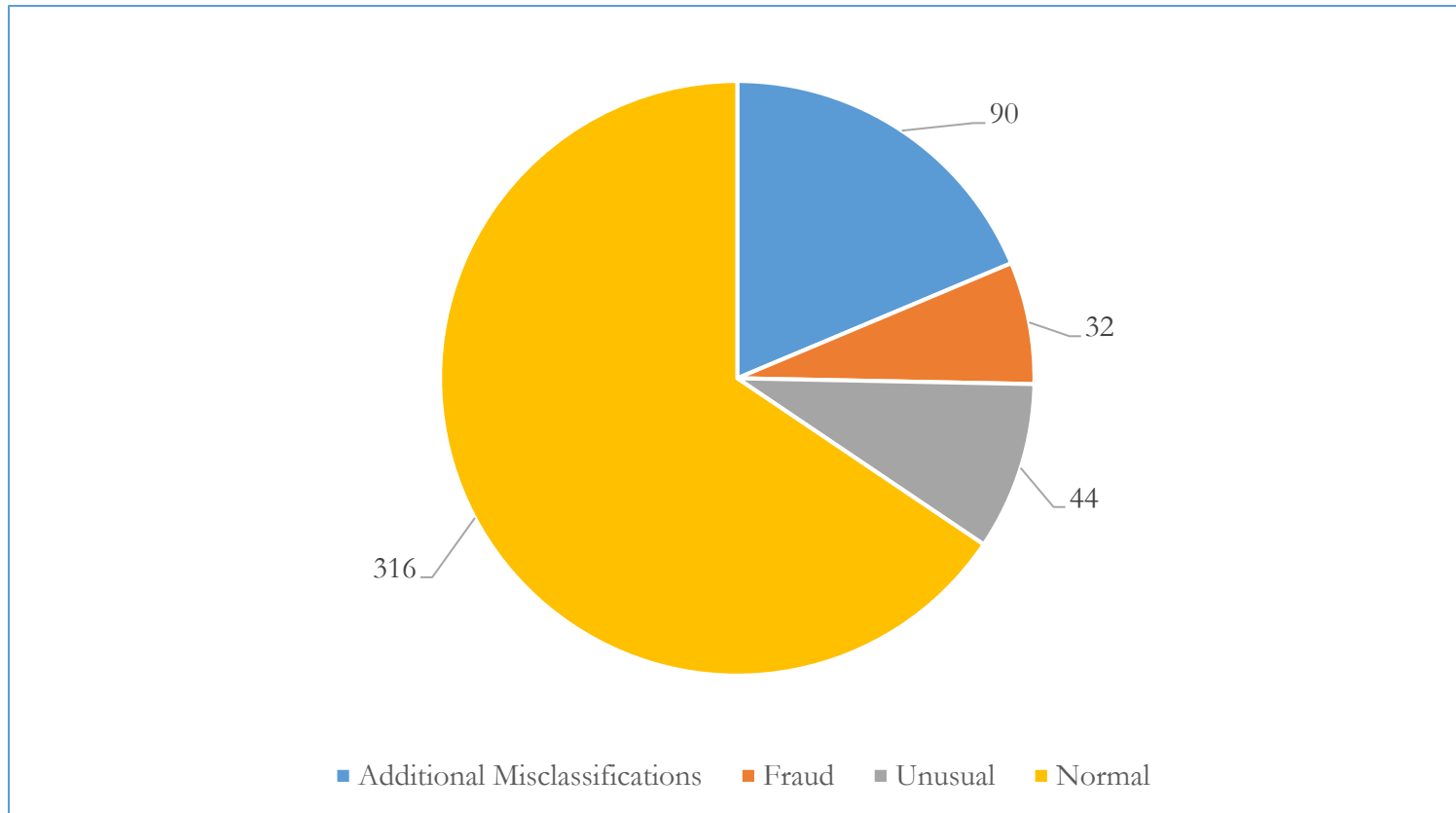
*Note:* This figure shows the number of influential observations by TNA Decile (with the exception of Deciles 1 and 2 similar to that of Chen et al. (2004)) over the 1963-1999 sample period. The line in red represents the influential observations mean monthly market adjusted return in percentages.

FIGURE 6  
INFLUENTIAL OBSERVATIONS BY YEAR



*Note:* This figure shows the number of influential observations by year covering the period from 1963-1999. The line in blue represents mean TNA decile and the line in red represents the mean monthly market adjusted return in percentage.

FIGURE 7  
ORIGINS OF INFLUENTIAL OBSERVATIONS



*Note:* This figure shows the origins of the influential observations in the sample. The classifications include: normal, unusual (e.g., dual purpose fixed maturity and irrevocable trust funds), fraud, and misclassifications.

APPENDIX A  
UNIVARIATE IDENTIFICATION AND TREATMENT OF MULTIVARIATE OUTLIERS: AN ILLUSTRATION

Appendix A displays three small datasets and demonstrates that neither winsorizing nor trimming mitigates the influence of multivariate outliers. Instead, these commonly used univariate outlier mitigation strategies actually exacerbate multivariate outlier induced bias (our example is consistent with Bollinger and Chandra, 2005). Intuitively, if multivariate outliers arise in a non-random fashion, trimming and dropping potentially introduces sample selection problems and biased coefficient estimates (Heckman, 1979).

Panel A presents the three data sets. Columns 1 and 2 list Y and X values for the “No Outliers” sample, columns 3 and 4 for the “Univariate Outlier” sample, and columns 5 and 6 for the “Multivariate Outlier” sample. In the “No Outliers” sample, there are no outliers and it is obvious there is a one-to-one relation between Y and X. In the “Univariate Outlier” sample, there is a single univariate outlier with very large Y and X values (equal to 10) when compared to the other observations. However, even with the large univariate outlier the relation between Y and X remains one-to-one. In the “Multivariate Outlier” sample, none of the Y or X values are extremely large or small but one observation, where  $Y=4$  and  $X=2$ , is a multivariate outlier. It is a multivariate outlier because its relation between Y and X does not fit the pattern formed by the other observations. Another type of multivariate outlier occurs when an observation has such extremely small or large values in either the Y or X dimensions that it affects the multivariate relation between Y and X.

Panel B reports coefficient estimates and  $R^2$  values from regressing Y on X for each of the three samples. The estimated coefficients and  $R^2$  values are, of course, equal to one for the “No Outlier” sample. Likewise, they are equal to one for the “Univariate Outlier” sample. This highlights an important but often overlooked characteristic of univariate outliers, univariate outliers are not a problem in terms of biasing coefficient estimates when they fit the pattern formed by the rest of the sample. In fact, large univariate outliers that fit the overall data pattern are good in that they decrease standard error estimates. In the “Multivariate Outlier” sample regression, the multivariate outlier biases the coefficient estimate and lowers the  $R^2$ . Winsorizing and trimming do not affect the regression estimates for the No Outlier and Univariate Outlier samples. For the latter, winsorizing and trimming have no impact because the univariate outlier is a ‘good’ outlier. If the univariate outlier had extreme Y or X values that did not fit the pattern of the rest of the data winsorizing could reduce, but not eliminate, the outlier’s effect on regression estimates.



Most importantly, Panel B shows winsorizing and trimming actually increases bias in the “Multivariate Outlier” sample. Our point is that the decision to winsorize or trim a sample should not be taken lightly. If a sample includes multivariate outliers winsorizing and trimming can lead to incorrect inferences. Our example is admittedly stylized but serves to illustrate an often-overlooked adverse effects of univariate outlier mitigation techniques in multivariate regressions. In unreported analyses, we conduct Monte Carlo simulations on large datasets and models that include several independent variables and reach the same conclusions.

Our approach to multivariate outliers in this paper is straightforward. We first conduct tests to find out if outliers are driving the diseconomies of scale findings in Chen et al. (2004). If so, we identify the most influential outliers and determine whether they are primarily due to bad data (e.g., data entry errors, sampling errors, variable construction, etc.), reflect some unusual fact that will lead to an improvement in economic theory or model specification (Zellner 1981), or simply occur by chance. Once this determination is made, we can then decide whether to keep, correct, delete, or mitigate them is most appropriate. Ideally, all genuinely bad data is corrected or dropped from a sample. Unfortunately, the data set in Chen et al. (2004), as in most asset pricing studies, is very large. This makes it impractical to identify and correct all bad data so outlier mitigation is necessary. We do this in two ways, by simply dropping the most extreme multivariate outliers and by using an outlier robust estimator that does not place more weight to extreme observations as does OLS.

Appendix A: TABLE 1  
A SIMPLE ILLUSTRATION OF THE MULTIVARIATE OUTLIER ISSUE

*Panel A: Data Scenarios*

<u>No Outliers</u>		<u>Univariate Outlier</u>		<u>Multivariate Outlier</u>	
Y-Value	X-Value	Y-Value	X-Value	Y-Value	X-Value
(1)	(2)	(3)	(4)	(5)	(6)
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	2
5	5	10	10	5	5

*Panel B: OLS Regression Estimates*

<u>No Outliers</u>		<u>Univariate Outlier</u>		<u>Multivariate Outlier</u>	
Y-Value	X-Value	Y-Value	X-Value	Y-Value	X-Value
All Observations					
$b$	1	$b$	1	$b$	.8
$R^2$	1	$R^2$	1	$R^2$	.7
Winsorize Smallest and Largest Observations					
$b$	1	$b$	1	$b$	.3
$R^2$	1	$R^2$	1	$R^2$	.2
Trim Smallest and Largest Observations					
$b$	1	$b$	1	$b$	0
$R^2$	1	$R^2$	1	$R^2$	0

*Note:* Table 1 provides three data scenarios and regression estimates to illustrate the multivariate outlier problem. Panel A presents the dependent variable Y and independent variable X for the no outlier scenario in columns 1 and 2, univariate outlier scenario in columns 3 and 4, and the multivariate outlier scenario in columns 5 and 6 where the multivariate outlier is the Y=4 and X=2 observation. Panel B presents regression estimates for each data scenario for all observations, winsorizing the smallest and largest observations, and trimming or dropping the smallest and largest observations.

## APPENDIX B

### RESULTS FOR ALTERNATIVE ROBUST ESTIMATORS

Appendix B reports results for three alternative robust estimators using the gross returns data and models in Panel A of Table 1. Least median squares (LMS) results are presented in Columns 1-4, least trimmed squares (LTS) in columns 5-8, and median quantiles regression results are presented in columns 9-12. In panel A, fund size (LOGTNA) is statistically insignificant for all return measures and estimators, results that are consistent with the outlier robust S-estimation results presented in Table 1. The results for family size (LOGFAMSIZE) are mixed, positively significant for LMS in column 2 (beta adjusted returns) and median regressions in columns 9-12 while insignificant for the remaining columns.

Panel B presents results for the second regressions where the models and data are the same as in Panel A. That is, we run the regressions a second time. The differing results for LMS and LTS regressions in Panel A and B demonstrate the reproducibility problem of each method. The LMS and LTS reproducibility problems are caused by large samples. As sample sizes increase the number of possible subsets needed to find global minima increases rapidly, thus random subsampling algorithms are used to estimate approximate, but not always adequate, solutions. A consequence of these random subsampling algorithms is that coefficient estimates change each time the regression is performed. For example, fund size (LOGTNA) is negative and significant for the market adjusted and beta adjusted gross return models in columns 5 and 6 of Panel B but insignificant in Panel A. In contrast, the median regression results are the same in each panel. Although not reported, the S-estimator results in all specifications are also reproducible (i.e., the estimated coefficients are the same each time the regressions are run). Overall, the results in Appendix B confirm the appropriateness of our decision to use S-estimation.

APPENDIX B: TABLE 1  
REGRESSION OF FUND PERFORMANCE (GROSS RETURNS) ON LAGGED FUND SIZE  
WITH ALTERNATIVE ESTIMATORS

*Panel A: First Regressions*

	<u>Least Median Squares</u>				<u>Least Trimmed Squares</u>				<u>Median Regression</u>			
	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor	Market- Adj.	Beta- Adj.	Three- Factor	Four- Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
INTERCEPT	0.553 (0.37)	0.114 (0.95)	-0.020 (0.15)	-0.030 (0.23)	0.130 (1.11)	-0.190 (1.58)	-0.121 (0.94)	-0.116 (0.77)	-0.023 (0.44)	0.003 (0.07)	0.016 (0.25)	-0.021 (0.32)
LOGTNA <sub>i,t-1</sub>	0.017 (0.80)	-0.026 (-1.36)	-0.003 (0.13)	0.016 (0.84)	-0.026 (1.41)	0.002 (0.08)	0.022 (1.20)	0.011 (0.53)	-0.010 (1.17)	-0.010 (1.18)	-0.008 (0.83)	-0.005 (0.60)
LOGFAMSIZE <sub>i,t-1</sub>	-0.005 (0.52)	0.017 (1.83)	-0.007 (0.81)	-0.014 (1.54)	-0.012 (1.29)	0.004 (0.54)	0.008 (0.78)	0.005 (0.62)	0.006 (1.86)	0.006 (1.90)	0.006 (1.86)	0.005 (1.74)
TURNOVER <sub>i,t-1</sub>	0.000 (0.45)	-0.000 (-0.20)	-0.000 (0.46)	-0.000 (0.03)	0.000 (0.83)	-0.000 (0.22)	-0.000 (0.39)	0.000 (0.78)	-0.000 (0.00)	-0.000 (0.00)	0.000 (0.03)	0.000 (0.02)
AGE <sub>i,t-1</sub>	-0.002 (1.25)	-0.001 (0.87)	0.000 (0.12)	-0.000 (0.13)	0.000 (0.28)	0.001 (0.75)	-0.001 (0.88)	-0.000 (0.22)	-0.000 (0.31)	-0.000 (0.33)	-0.000 (0.34)	-0.000 (0.37)
EXPRATIO <sub>i,t-1</sub>	-0.235 (2.23)	-0.101 (1.18)	-0.012 (0.13)	0.015 (0.17)	-0.109 (1.16)	0.056 (0.68)	0.040 (0.48)	-0.072 (0.83)	-0.013 (0.34)	-0.011 (0.29)	-0.020 (0.51)	-0.019 (0.48)
TOTLOAD <sub>i,t-1</sub>	-0.017 (2.16)	0.003 (0.52)	0.010 (1.57)	0.011 (1.50)	0.013 (1.85)	0.006 (0.88)	-0.008 (1.03)	-0.003 (0.42)	0.003 (1.05)	0.003 (1.11)	0.003 (1.25)	0.003 (1.31)
FLOW <sub>i,t-1</sub>	-0.000 (0.26)	0.000 (0.45)	0.000 (0.50)	-0.000 (0.41)	-0.000 (0.00)	-0.000 (0.38)	0.000 (0.23)	-0.000 (0.05)	-0.000 (0.04)	-0.000 (0.02)	-0.000 (0.06)	-0.000 (0.12)
LAGFUNDRET <sub>i,t-1</sub>	0.031 (5.05)	0.030 (4.69)	0.027 (4.25)	0.031 (5.29)	0.035 (5.81)	0.029 (4.69)	0.033 (5.53)	0.028 (4.31)	0.030 (6.03)	0.030 (5.98)	0.030 (5.99)	0.030 (5.99)
No. of Months	444	444	444	444	444	444	444	444	444	444	444	444

Panel B: Second Regressions

	Least Median Squares				Least Trimmed Squares				Median Regression			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
INTERCEPT	0.036 (0.28)	-0.008 (0.06)	0.075 (0.75)	0.030 (0.22)	0.173 (1.45)	0.236 (1.95)	0.095 (0.73)	0.055 (0.41)	-0.023 (0.44)	0.004 (0.07)	0.016 (0.25)	-0.021 (0.32)
LOGTNA <sub>i,t-1</sub>	-0.019 (1.00)	0.007 (0.36)	0.009 (0.47)	0.020 (0.94)	-0.036 (1.93)	-0.034 (1.90)	0.011 (0.57)	0.012 (0.58)	-0.010 (1.17)	-0.010 (1.18)	-0.008 (0.83)	-0.005 (0.60)
LOGFAMSIZE <sub>i,t-1</sub>	0.007 (0.77)	-0.003 (0.34)	-0.010 (0.12)	0.001 (0.07)	-0.004 (0.38)	0.010 (0.99)	0.001 (0.07)	0.005 (0.59)	0.006 (1.86)	0.006 (1.90)	0.006 (1.86)	0.005 (1.74)
TURNOVER <sub>i,t-1</sub>	-0.000 (0.02)	0.006 (1.02)	-0.001 (0.93)	-0.001 (1.28)	-0.000 (1.13)	-0.000 (0.42)	-0.001 (0.86)	0.000 (0.43)	-0.000 (0.00)	-0.000 (0.00)	0.000 (0.03)	0.000 (0.02)
AGE <sub>i,t-1</sub>	-0.002 (1.40)	-0.000 (0.02)	-0.003 (1.60)	-0.000 (0.07)	-0.000 (0.16)	-0.003 (1.57)	-0.002 (1.31)	-0.004 (2.49)	-0.000 (0.31)	-0.000 (0.33)	-0.000 (0.34)	-0.000 (0.37)
EXPRATIO <sub>i,t-1</sub>	0.013 (0.17)	-0.058 (0.64)	-0.065 (0.81)	-0.160 (1.95)	-0.101 (1.31)	-0.162 (1.86)	-0.141 (1.77)	-0.116 (1.36)	-0.013 (0.34)	-0.011 (0.29)	-0.019 (0.51)	-0.019 (0.48)
TOTLOAD <sub>i,t-1</sub>	0.003 (0.43)	-0.003 (0.45)	0.003 (0.49)	0.001 (0.09)	0.017 (2.66)	0.002 (0.37)	0.004 (0.67)	-0.011 (1.51)	0.003 (1.05)	0.003 (1.11)	0.003 (1.25)	0.004 (1.31)
FLOW <sub>i,t-1</sub>	0.000 (0.17)	0.000 (0.08)	-0.001 (0.85)	0.001 (1.42)	0.001 (1.04)	-0.001 (0.79)	0.001 (1.26)	-0.002 (1.98)	-0.000 (0.04)	-0.000 (0.02)	-0.000 (0.06)	-0.000 (0.12)
LAGFUNDRET <sub>i,t-1</sub>	0.033 (5.94)	0.027 (4.10)	0.026 (4.29)	0.033 (5.57)	0.032 (5.53)	0.029 (4.89)	0.034 (5.56)	0.034 (5.51)	0.030 (6.03)	0.030 (5.98)	0.030 (5.99)	0.030 (5.99)
No. of Months	444	444	444	444	444	444	444	444	444	444	444	444

*Note:* This tables replicates the results of Chen et al. (2004) Chen el al. Table 3 using Fama-MacBeth (1973) estimates of gross monthly fund returns on fund characteristics lagged 1 month (compare to Table 1 of this paper). The sample includes funds that fall within fund size quantiles two to five. Columns 1 through 4 provides results from least median sum of squares (LMS), columns 5 through 8 provides least trimmed sum of squares (LTS), and columns 9 through 12 report median quantile regressions. Returns are adjusted using the market model (Market-Adj), CAPM (Beta-Adj), and Fama-French three and four factor models. LOGTNA and LOGFAMSIZE are the natural logarithms of fund TNA and one plus the family TNA (less the fund), respectively. TURNOVER is the fund turnover, AGE is the number of years since inception, EXPRATIO is the annual fees and expenses scaled by fund TNA, and TOTLOAD is the sum of front-end, rear-end, and deferred sales charges also scaled by fund TNA. FLOW is the percentage of new fund flow in the fund in the past year. LAGFUNDRET is the buy and hold return of the fund over the prior 12 months. The sample is from January 1963 to December 1999. The *t*-statistics in parentheses are adjusted for serial correlation using Newey-West (1987) lags of order three. Panel A reports results for the first set of regressions and Panel B reports results for the second set.

## APPENDIX C

### ROBUSTNESS CHECKS

Chen et al. (2004) change their sample in three ways to demonstrate robustness in their findings. We replicate their tests and report their results along with our findings in Table A1. For their main regression model reported in our Table 1, Chen et al. (2004) use mutual funds that have sizes that places them in quintiles 2 through 5. This is meant to protect against the smallest mutual funds biasing the results. For both gross and net returns, Chen et al. (2004) confirm in the first robustness test that the results hold for all funds.

Panels A and B in Table A1 reproduce their results in Columns 1 through 4 and the robust regressions in Columns 5 through 8. After controlling for outliers, there is no significant correlation between fund or family sizes for either gross returns in Panel A or net returns in Panel B. In Panels C and D, we report their test that uses annual returns instead of monthly. Whereas Chen et al. (2004) find a negative relation, we do not. In Panels E, F, G and H, we follow Chen et al. (2004) and split the sample in half based on the sample time period. Again, fund size does not erode performance and there is no positive relation between family size and returns.

TABLE C1—REGRESSIONS OF FUND RETURNS ON LAGGED FUND SIZE, ROBUSTNESS CHECKS

<i>Panel A: Sample includes all funds: Gross Returns</i>								
	Published Results				Robust Regressions			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.023 (2.59)	-0.026 (2.94)	-0.019 (2.13)	-0.020 (2.27)	-0.006 (0.75)	-0.010 (1.18)	-0.002 (0.18)	-0.001 (0.09)
LOGFAMSIZE <sub>i,t-1</sub>	0.006 (1.99)	0.006 (2.03)	0.006 (1.98)	0.006 (2.03)	0.006 (1.73)	0.006 (1.65)	0.006 (1.68)	0.006 (1.83)
No. of Months	444	444	444	444	444	444	444	444
<i>Panel B: Sample includes all funds: Net Returns</i>								
	Published Results				Robust Regressions			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.020 (2.16)	-0.022 (2.49)	-0.015 (1.76)	-0.017 (2.02)	-0.005 (0.55)	-0.005 (0.55)	0.000 (0.03)	0.007 (0.65)
LOGFAMSIZE <sub>i,t-1</sub>	0.006 (2.09)	0.006 (2.15)	0.006 (2.11)	0.006 (2.17)	0.006 (1.67)	0.006 (1.76)	0.006 (1.65)	0.003 (0.69)
No. of Months	444	444	444	444	444	444	444	444
<i>Panel C: Dependent variable is 12-month (non-overlapping) fund returns: Gross Returns</i>								
	Published Results				Robust Regressions			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.436 (2.41)	-0.440 (3.27)	-0.345 (2.40)	-0.312 (2.19)	-0.183 (1.29)	-0.172 (1.20)	-0.103 (0.60)	-0.079 (0.46)
LOGFAMSIZE <sub>i,t-1</sub>	0.088 (2.05)	0.089 (2.06)	0.088 (2.05)	0.088 (2.07)	0.032 (0.87)	0.038 (0.98)	0.014 (0.41)	0.012 (0.37)
No. of Months	37	37	37	37	37	37	37	37
<i>Panel D: Dependent variable is 12-month (non-overlapping) fund returns: Net Returns</i>								
	Published Results				Robust Regressions			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.402 (3.07)	-0.406 (3.07)	-0.312 (2.20)	-0.280 (1.99)	-0.097 (0.95)	-0.156 (1.13)	-0.084 (0.50)	-0.076 (0.45)
LOGFAMSIZE <sub>i,t-1</sub>	0.090 (2.10)	0.090 (2.10)	0.090 (2.11)	0.090 (2.12)	0.039 (1.11)	0.036 (1.02)	0.016 (0.52)	0.030 (0.92)
No. of Months	37	37	37	37	37	37	37	37
<i>Panel E: Sample period is from 1963-1980: Gross Returns</i>								
	Published Results				Robust Regressions			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)

LOGTNA <sub>i,t-1</sub>	-0.035 (2.41)	-0.034 (2.42)	-0.021 (1.51)	-0.019 (1.33)	-0.009 (0.51)	-0.002 (0.13)	0.009 (0.52)	0.011 (0.64)
LOGFAMSIZE <sub>i,t</sub> -1	0.002 (0.49)	0.002 (0.50)	0.002 (0.45)	0.002 (0.44)	0.006 (0.69)	0.005 (0.81)	0.002 (0.31)	0.003 (0.42)
No. of Months	216	216	216	216	216	216	216	216

Panel F: Sample period is from 1963-1980: Net Returns

	Published Results				Robust Regressions			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.032 (2.22)	-0.032 (2.23)	-0.019 (1.32)	-0.016 (1.15)	-0.001 (0.03)	-0.001 (0.08)	0.006 (0.33)	0.017 (0.99)
LOGFAMSIZE <sub>i,t</sub> -1	0.002 (0.56)	0.002 (0.56)	0.002 (0.53)	0.002 (0.52)	0.002 (0.31)	0.006 (0.69)	0.005 (0.59)	0.003 (0.43)
No. of Months	216	216	216	216	216	216	216	216

Panel G: Sample period is from 1981-1999: Gross Returns

	Published Results				Robust Regressions			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.021 (1.84)	-0.022 (1.86)	-0.024 (1.93)	-0.022 (1.71)	-0.000 (0.02)	-0.000 (0.03)	-0.002 (0.17)	-0.001 (0.04)
LOGFAMSIZE <sub>i,t</sub> -1	0.012 (2.53)	0.012 (2.53)	0.012 (2.50)	0.012 (2.50)	0.003 (0.61)	0.003 (0.57)	0.003 (0.54)	0.002 (0.49)
No. of Months	228	228	228	228	228	228	228	228

Panel H: Sample period is from 1981-1999: Net Returns

	Published Results				Robust Regressions			
	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor	Market-Adj.	Beta-Adj.	Three-Factor	Four-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA <sub>i,t-1</sub>	-0.019 (1.64)	-0.019 (1.65)	-0.022 (1.74)	-0.019 (1.53)	0.000 (0.09)	0.001 (0.12)	-0.001 (0.08)	0.001 (0.04)
LOGFAMSIZE <sub>i,t</sub> -1	0.012 (2.55)	0.012 (2.55)	0.012 (2.52)	0.012 (2.52)	0.003 (0.67)	0.003 (0.62)	0.003 (0.62)	0.003 (0.60)
No. of Months	228	228	228	228	228	228	228	228

*Note:* The table reports robustness tests. Panels A and B report results for returns calculated before and after (gross and net) deducting fees and expenses for funds in the one to five size quintiles. Panels C and D report results for 12-month non-overlapping returns calculated before and after (gross and net) deducting fees and expenses. Panels E and F (G and H) report results for returns calculated before and after (gross and net) deducting fees and expenses for funds in the sample period 1963 to 1980 (1981-1999). Additional covariates are: TURNOVER, AGE, EXPRATIO, TOTLOAD, FLOW, and LAGFUNDRET. The *t*-statistics in parentheses are adjusted for serial correlation using Newey-West (1987) lags of order three.